The Genetics of the Pig, 2nd Edition
This page intentionally left blank
The Genetics of the Pig, 2nd Edition

Edited by

Max F. Rothschild

Department of Animal Science
Center for Integrated Animal Genomics
Iowa State University
USA

and

Anatoly Ruvinsky

The Institute for Genetics and Bioinformatics
University of New England
Australia
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contributors</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>Preface</td>
<td>Max F. Rothschild and Anatoly Ruvinsky</td>
<td>ix</td>
</tr>
<tr>
<td>1</td>
<td>Systematics and Evolution of the Pig</td>
<td>Anatoly Ruvinsky, Max F. Rothschild, Greger Larson and Jaime Gongora</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Genetic Aspects of Pig Domestication</td>
<td>Greger Larson, Thomas Cucchi and Keith Dobney</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>Molecular Genetics of Coat Colour Variation</td>
<td>Leif Andersson and Graham Plastow</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>Genetics of Morphological Traits and Inherited Disorders</td>
<td>Frank W. Nicholas</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>Molecular Genetics</td>
<td>Chris Moran</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>Immunogenetics</td>
<td>Joan K. Lunney, Tomoko Eguchi-Ogawa, Hirohide Uenishi, Nancy Wertz and John E. Butler</td>
<td>101</td>
</tr>
<tr>
<td>7</td>
<td>Cytogenetics and Chromosome Maps</td>
<td>Terje Raudsepp and Bhanu P. Chowdhary</td>
<td>134</td>
</tr>
<tr>
<td>8</td>
<td>Pig Genomics</td>
<td>Martien A.M. Groenen, Lawrence B. Schook and Alan L. Archibald</td>
<td>179</td>
</tr>
<tr>
<td>9</td>
<td>Behaviour Genetics of the Domestic Pig</td>
<td>Anna K. Johnson and John J. McGlone</td>
<td>200</td>
</tr>
<tr>
<td>10</td>
<td>Biology and Genetics of Reproduction</td>
<td>Jean-Pierre Bidanel</td>
<td>218</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Author(s)</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>11</td>
<td>Transgenics and Modern Reproductive Technologies</td>
<td>Jason W. Ross and Randall S. Prather</td>
<td>242</td>
</tr>
<tr>
<td>12</td>
<td>Developmental Genetics</td>
<td>Michael Dyck and Anatoly Ruvinsky</td>
<td>263</td>
</tr>
<tr>
<td>13</td>
<td>Pig Genetic Resources</td>
<td>Louis Ollivier and Jean-Louis Foulley</td>
<td>306</td>
</tr>
<tr>
<td>14</td>
<td>Genetics of Performance Traits</td>
<td>Archie C. Clutter</td>
<td>325</td>
</tr>
<tr>
<td>15</td>
<td>Genetics of Meat Quality and Carcass Traits</td>
<td>Daniel C. Ciobanu, Steven M. Lonergan and Elisabeth J. Huff-Lonergan</td>
<td>355</td>
</tr>
<tr>
<td>16</td>
<td>Genetic Improvement of the Pig</td>
<td>Jack C.M. Dekkers, Pramod K. Mathur and Egbert F. Knol</td>
<td>390</td>
</tr>
<tr>
<td>17</td>
<td>Pigs as a Model for Biomedical Sciences</td>
<td>Kristy N. Kuzmuk and Lawrence B. Schook</td>
<td>426</td>
</tr>
<tr>
<td>18</td>
<td>Breeds of Pigs</td>
<td>David S. Buchanan and Ken Stalder</td>
<td>445</td>
</tr>
<tr>
<td>19</td>
<td>Standard Genetic Nomenclature of the Pig, with Glossaries</td>
<td>Zhi-Liang Hu, Carissa A. Park and James M. Reecy</td>
<td>473</td>
</tr>
</tbody>
</table>

Index                                                                 | 495  |

The colour plate section can be found following p. 262.
Contributors

Leif Andersson, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-75124 Uppsala, Sweden
Alan L. Archibald, BBSRC Roslin Institute, Roslin, Midlothian EH25 9PS, UK
Jean-Pierre Bidanel, Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy, 78352, Jouy-en-Josas cedex, France
David S. Buchanan, Department of Animal Sciences, North Dakota State University, Dept #7630, PO Box 6050, Fargo, ND 58108-6050, USA
John E. Butler, Department of Microbiology, University of Iowa, Iowa City, IA 52242, USA
Bhanu P. Chowdhary, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, TX 77843-4458, USA
Daniel C. Ciobanu, A218k Animal Science Building, Department of Animal Science, University of Nebraska-Lincoln, Lincoln, NE 68583-0908, USA
Archie C. Clutter, Newsham Choice Genetics Research Center, 701 Crown Industrial Court, Saint Louis, MO 63005, USA
Thomas Cucchi, UMR 7209 ‘Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements’ MNHN-CP56, 55, rue Buffon, F-75231 Paris cedex 05, France; Department of Archaeology, University of Aberdeen, St Mary’s, Elphinstone Road, Aberdeen AB24 3UF, UK
Jack C.M. Dekkers, 239D Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Keith Dobney, Department of Archaeology, University of Aberdeen, St Mary’s, Elphinstone Road, Aberdeen AB24 3UF, UK
Michael Dyck, Department of Agricultural, Food and Nutritional Science, 3-10C Agriculture/Forestry Centre, University of Alberta, Edmonton, Alberta T6G 2P5, Canada
Tomoko Eguchi-Ogawa, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan
Jean-Louis Foulley, INRA, Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy, Domaine de Vilvert 78352, Jouy-en-Josas cedex, France
Jaime Gongora, Faculty of Veterinary Science, RMC Gunn Building B-19, The University of Sydney, NSW 2006, Australia
Martien A.M. Groenen, Animal Breeding and Genetics, Wageningen University, Animal Sciences, 6700AH, Wageningen, The Netherlands
Zhi-Liang Hu, 2255 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Elisabeth J. Huff-Lonergan, 2278 Kildee Hall, Department of Animal Science, Iowa State University, IA 50011-3150, USA
Anna K. Johnson, 2356 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011, USA
Egbert F. Knol, IPG-Institute for Pig Genetics B.V., PO Box 43, 6640AA Beuningen, The Netherlands
Kristy N. Kuzmuk, Department of Animal Sciences, Institute for Genomic Biology, 1201 West Gregory Drive, 352 ERML, University of Illinois, Urbana, IL 62801, USA
Greger Larson, Department of Archaeology, Durham University, South Road, Durham DH1 3HP, UK
Steven M. Lonergan, 2275 Kildee Hall, Department of Animal Science, Iowa State University, IA 50011-3150, USA
Joan K. Lunney, Animal Parasitic Diseases Laboratory, ANRI, ARS, USDA, Building 1040, Room 103, BARC-East Beltsville, MD 20705-0000, USA
Pramod K. Mathur, IPG-Institute for Pig Genetics B.V., PO Box 43, 6640AA Beuningen, The Netherlands
John J. McGlone, Pork Industry Institute, Department of Animal Science and Food Technology Texas Tech University, Lubbock, TX 79409-2141, USA
Chris Moran, Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, University of Sydney, NSW 2006, Australia
Frank W. Nicholas, Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, University of Sydney, NSW 2006, Australia
Louis Ollivier, INRA, Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy, Domaine de Vilvert 78352, Jouy-en-Josas cedex, France
Carissa A. Park, 2255 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Graham Plastow, Department of Agricultural, Food and Nutritional Science, Agriculture/Forestry Centre, University of Alberta, Edmonton, Alberta T6G 2P5, Canada
Randall S. Prather, Division of Animal Science, University of Missouri, E125 Animal Science Research Center, Columbia, MO 65211, USA
Terje Raudsepp, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, TX 77843-4458, USA
James M. Reecy, 2255 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Jason W. Ross, 2356 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Max F. Rothschild, 2255 Kildee Hall, Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, Ames, IA 50011-3150, USA
Anatoly Ruvinsky, The Institute for Genetics and Bioinformatics, University of New England, Armidale, NSW 2351, Australia
Lawrence B. Schook, Department of Animal Sciences, Institute for Genomic Biology, 1201 West Gregory Drive, 352 ERML, University of Illinois, Urbana, IL 62801, USA
Ken Stalder, 109 Kildee Hall, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA
Hirohide Uenishi, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan
Nancy Wertz, Department of Microbiology and Iowa Interdisciplinary Immunology Program, University of Iowa, 3-550 Bowen Science Building, Iowa City, IA 52242-1109, USA
The first edition of The Genetics of the Pig was published more than 12 years ago. During the past several years a deep transformation has occurred in all of the biological sciences. Beginning in 2001, several mammalian genomes were sequenced. The latest addition to this list is the pig genome, whose first draft sequence was obtained at the end of 2009. While the annotation of the pig genome is still an ongoing process one can confidently expect that in the near future this process will be significantly advanced. These developments in the biological sciences have led to the practical implementation of powerful laboratory methods which, together with the emergence of advanced bioinformatics, have caused dramatic accumulations of new data in diverse databases.

Genetic theory and practice have also evolved considerably over the last 12 years. The previously separate fields of classical and quantitative genetics have now been joined with genomics, molecular genetics and immunogenetics, creating new methods and insights into understanding numerous biological processes. Today, more than in the past, genetics serves as a key pillar of modern agricultural sciences, medicine and the biotechnological industry. Animal breeders and geneticists, more widely than ever before, are using knowledge from the different fields of genetics for development and improvement of their livestock. Given these dramatic changes in current genetic applications and theory it seems an appropriate time to publish the second edition of books devoted to genetics of domestic livestock species. The Genetics of the Pig opens this series of books.

Domestication of the pig occurred some 9000 years ago, and the consequences of this process have been tremendously important for the food supply in different civilizations. Today, the pig continues to be a valued source of food worldwide. Modern biological discoveries and technological improvements in management practices have revolutionized pork production. Approximately one billion pigs are raised annually worldwide and pork remains the dominant meat source, representing over 40% of all the red meat eaten. Owing to its physiological and genetic similarities with man, the pig serves as an excellent animal model for biomedical research, and as an important source for xenotransplantation and other potential medical applications. Numerous animal scientists, geneticists, veterinarians, livestock producers, medical researchers and students are interested in the biology and genetics of the pig. This new edition of the book brings a wealth of knowledge that we hope will be useful for this diverse group of scientists and practitioners around the world.

The purpose of this book is to present in one location a complete, comprehensive and updated description of the modern genetics of the pig. It is our intention to combine essential knowledge from the various fields of genetics and biology of the pig, integrated with livestock management aspects, in order to provide an updated and informative reference book. The genetic improvements in the pig industry over the past couple of decades have been very impressive, with growth rates increasing, feed efficiency improving and a continued rise in overall
leanliness. These successes have in great part been due to the incredible progress in the understanding and application of genetics to pig production. As recently as 1990, only about 50 genes and markers were mapped or assigned to individual porcine chromosomes. Now the genome draft sequence has been obtained and the completion of very detailed genetic and genomic maps has been accomplished. Furthermore, genetic improvement within the pig industry that rests on the introduction of gene tests and on genomic selection based on thousands of genes is coming.

This book is addressed to a diverse audience, including students, researchers, veterinarians and pig breeders. The initial two chapters are devoted to the taxonomy and domestication of the pig. This area has advanced significantly over the last decade. Chapters 3 and 4 extensively cover the genetics of coat colour, morphological characteristics and inherited disorders. Molecular genetics and immunogenetics are described in Chapters 5 and 6. Cytogenetics, chromosome maps and genomics are presented in Chapters 7 and 8. Chapter 9 concentrates on the genetics of behaviour, while the next three chapters are relevant to the biology and genetics of reproduction, modern reproductive technologies and the genetics of development.

Chapter 13 addresses genetic diversity and concerns for maintaining exotic and rare local breeds. The genetics of performance traits and carcass and meat quality traits are discussed in Chapters 14 and 15. Chapter 16 is devoted to overall genetic improvement. Chapter 17 examines the pig’s contribution and future potential as an important model for biomedical sciences and a key species for possible organ donation. Chapters 18 and 19 cover pig breeds and genetic nomenclature.

A significant effort was made to consistently implement the current genetic nomenclature. Unfortunately, there are a few shortcomings that seem to be unavoidable. The genetic nomenclature requires that not only genes and alleles, but also traits, should be spelled according to the American style. Essentially, just a few words like colour (color), behaviour (behavior) and flavour (flavor) cause some difficulties. It is not always absolutely clear whether a word like the above describes a trait from the formal point of view and as such the American style should be used. As this book is published in the UK the British spelling rules should apply in all other cases. More on this nomenclature matter can be found in Chapter 19.

The considerable and never-ending progress in genetics research makes it impossible to cover all new and relevant literature. Therefore, it is inevitable that some publications will not be cited. We hope that any errors or omissions will be noted and brought to our attention. Also, during the time when this book was being written there were incremental changes in genetic nomenclature. Despite the consistent attempts to introduce all these changes into the book, we probably cannot claim to have been completely successful. Finally, in no way is this book meant to replace the many fine textbooks devoted to the theory of animal breeding.

This book is the result of international efforts. These efforts are dedicated to our families, to supportive colleagues and to the pig industry that employs and feeds hundreds of millions of people worldwide. The editors offer a special thank you to each of the authors who contributed chapters so graciously and without reservation. Previous authors are also thanked for providing materials useful to this second edition. Publication of colour plates was generously supported by PIC (part of Genus plc), Hendersonville, Tennessee, USA, and the USDA/CSREES Pig Genome Coordination program. Finally, the editors also acknowledge with appreciation all the efforts of CABI to help in producing the book. It is our hope that this text will serve as a useful resource for all those people who study or work with pigs.

Max F. Rothschild
Anatoly Ruvinisky
November 2010
Introduction

According to the current classification, pigs belong to order Cetartiodactyla, which includes even-toed ungulates (the former order Artiodactyla), and whales and dolphins, which are representatives of the former order Cetacea (Murphy et al., 2001; Novacek, 2001). Cetartiodactyla diverged from placentals approximately 87.2 million years ago (mya) (Murphy and Eizirik, 2009).

There are three well-established suborders: (i) Tylopoda – camels and llamas; (ii) Suiformes (also known as Suina) – pigs and peccaries (and formerly hippos); and (iii) Ruminantia. The fourth and more recently created suborder – called Cetancodonta – includes hippos, dolphins and whales (Price et al., 2005; O’Leary and Gatesy, 2008; Huffman, 2009). Phylogenetic analyses of the cytochrome b sequences from 264 of the 290 extant Cetartiodactyla show that the Suiformes and Ruminantia cluster together as a sister clade of Cetancodonta, while the Tylopoda cluster in a separate clade (Agnarsson and May-Collado, 2008). Molecular genetic studies of the Cetartiodactyla have shown good agreement with the basic structure of their phylogeny as supported by morphological data (Novacek, 1992), but the position of hippos and camels
remains contentious (Thewissen et al., 2007; O’Leary and Gatesy, 2008; Spaulding et al., 2009).

Suborder Suiformes

The Suiformes consist of two living families: Tayassuidae (peccaries) and Suidae (pigs). Currently Tayassuidae has three genera, each with a single species, and all these species live in the Americas. The modern Suidae family is more diverse and consists of six genera with 18 or 19 recognized species depending on the source of information. Many of these species live in South-east Asia, others broadly in Eurasia and several species are found only in Africa.

Family Tayassuidae

The family Tayassuidae (peccaries) diverged from pigs in South-east Asia somewhere in the late Eocene, or possibly later, and then migrated to Eurasia, Africa and North America, finally colonizing South America (Ducrocq, 1994). However, the modern species live only in the Americas. The peccaries, like pigs, have a snout disc, but the differences between the two families are significant. The stomach in peccaries is subdivided into three compartments. In several features of the digestive system, peccaries resemble ruminants; it is not known whether these features developed independently. Peccaries have three hooves on their back legs. In addition, their upper canine teeth are pointed down and the total number of teeth is 38. A scent-producing gland located on their backs is another specific feature. Peccaries are significantly smaller than pigs, with an average body size of approximately 30kg.

There are three recognized extant peccary species that are distributed on the American continent: *Catagonus wagneri* (Chacoan peccary), *Tayassu pecari* (white-lipped peccary) and *Pecari tajacu* (collared peccary) (Grubb, 1993a). The collared peccary spread across South and Central America and the southern part of North America. It is a very common species, reproducing well and widely hunted because of its good meat and leather quality. The white-lipped peccary is bigger than the collared peccary. This species is spread from southern Mexico to the south of Central America. Hybridization between the white-lipped peccary and the collared peccary has been observed in captivity (Sowls, 1997) and in the wild (Andrea et al., 2001), with the wild hybrid being sterile. The Chacoan peccary was known only through fossil records until it was discovered on the Paraguay–Bolivia–Argentina border about three decades ago (Wetzel, 1977a). This species has been shown to differ significantly from the two other peccaries in chromosome number (2n = 20) (Benirschke et al., 1985; Benirschke and Kumamoto, 1989).

There are conflicting hypotheses about the evolution and relationships of modern peccary species. Based on osteological and dental traits, collared and Chacoan peccaries were considered to be more closely related, while the white-lipped peccary was considered to be a member of a separate clade, along with other extinct species (Wright, 1998). However, other morphological studies (Wetzel, 1977b) suggested that collared and white-lipped peccaries are more closely related to each other than to the Chacoan peccary. In contrast, phylogenetic studies using mitochondrial and nuclear DNA sequences show that white-lipped and Chacoan peccary species are more closely related to each other than to the collared peccary (see Fig. 1.1) (Theimer and Keim, 1998; Gongora and Moran, 2005). Additional DNA studies have suggested that the geographically widespread and phenotypically diverse collared peccary may consist of at least two separate lineages deserving specific status, which are as genetically distinct as white-lipped and Chacoan peccaries (Gongora et al., 2006).

Taxonomy and Phylogeny of the Suidae Family

Systematics

The Suidae family includes the most widely spread species of non-ruminant even-toed ungulates, commonly known as pigs and hogs. All of them have an elongated muzzle with a snout disc and four-toe extremities with well-developed...
side toes. The canine teeth are large and the upper ones are curved. The stomach is simple with an additional sac. The Suidae are omnivorous. This family traces back to the upper Eocene of Thailand (∼35–40 mya), or possibly later (Ducrocq et al., 1998; Liu, 2003). During the Neogene, suids greatly diversified into over 30 genera, and colonized different parts of Eurasia and Africa, where they radiated further (Pickford, 1993, 2006). The extant family Suidae comprises 15 species grouped into several genera: Sus (domestic and wild pigs) from Eurasia; Porcula (pygmy hogs) from northern India (the separation of Porcula from Sus is not finally resolved); Babyrousa (babirusa) from the island of Sulawesi and its satellite islands; and Potamochoerus (bush pig and red river hog), Phacochoerus (common and desert warthogs) and Hylochoerus (forest hog) from sub-Saharan Africa (Grubb, 1993a,b). The origin, evolutionary relationships and dispersal patterns of Suidae remain contentious. For instance, it has been suggested that some modern suids from sub-Saharan Africa are more closely related to species from Eurasia than to their African congeners on the basis of cranial and dental similarities (Thenius 1970; Cooke, 1978) and cytochrome b sequence (Agnarsson and May-Collado, 2008). In contrast, other morphological analyses suggest that the modern African Suidae could be sister lineages (Harris and White, 1979; Bender, 1992; Geraads, 2004). This lack of consensus extends to the Asian congeners of the Suidae. For instance, it has been suggested that some modern suids from sub-Saharan Africa are more closely related to species from Eurasia than to their African congeners on the basis of cranial and dental similarities (Thenius 1970; Cooke, 1978) and cytochrome b sequence (Agnarsson and May-Collado, 2008). In contrast, other morphological analyses suggest that the modern African Suidae could be sister lineages (Harris and White, 1979; Bender, 1992; Geraads, 2004). This lack of consensus extends to the Asian congeners of the Suidae. For instance, it has been suggested (on the basis of mitochondrial DNA) that Sus salvanius (pygmy hog) deserves a separate status from the genus Sus, and no one has conclusively demonstrated whether Babyrousa from South-east Asia or Phacochoerus from Africa occupies a basal position within the Suidae (Pickford, 1993; Funk et al., 2007). Some DNA studies have contributed to understanding the relationships within Phacochoerus and Sus (Randi et al., 1996, 2002). Recently, a concatenated mitochondrial and nuclear DNA study by Gongora et al. (submitted) has provided new insights into the evolutionary relationships of Suidae as a way to resolve conflicting and unresolved Suidae topologies generated by individual sequences. This study shows that all sub-Saharan African genera cluster in a monophyletic clade separate from the Eurasian Sus species, and confirms that Babyrousa is the sister taxon to the other extant species of Suidae, when Tayassuidae is used as the out-group (Fig. 1.1). Accordingly, Gongora et al. (submitted) propose that the five extant genera of Suidae should be grouped into the subfamilies Babyrousinae and Suinae, with the latter consisting of three tribes, the so-called ‘true’ pigs (Suini), warthogs and the forest hog (Phacochoerini) and the bush pig and red river hog (Potamochoerini).

### Subfamily Babyrousinae

- **Genus Babyrousa**
  - Species *Babyrousa babyrussa* (babirussas on the Sula Islands and Buru Island)
  - Species *Babyrousa celebensis* (babirussas from the northern arm of Sulawesi)
  - Species *Babyrousa togeanensis* (babirussas restricted to the Sulawesi Togean islands)

### Subfamily Suinae

#### Tribe Phacochoerini

- **Genus Phacochoerus**
  - Species *Phacochoerus africanus* (common warthog)
  - Species *Phacochoerus aethiopicus* (Cape and Somali warthog)
- **Genus Hylochoerus**
  - Species *Hylochoerus meinertzhageni* (forest hog)

#### Tribe Potamochoerini

- **Genus Potamochoerus**
  - Species *Potamochoerus porcus* (red river hog)
  - Species *Potamochoerus larvatus* (bush pig)

#### Tribe Suini

- **Genus Sus**
  - Species *Sus scrofa* (Eurasian wild boar)
  - Species *Sus verrucosus* (Javan warty pig)
  - Species *Sus barbatus* (bearded pig)
  - Species *Sus celebensis* (Sulawesi warty pig)
  - Species *Sus philippensis* (Philippine warty pig)
  - Species *Sus cebifrons* (Visayan warty pig)
  - Species *Sus salvanius* or *Porcula salvanius*? (pygmy hog)
The general taxonomy of the family seems reasonably justified, although the taxonomy of some genera, *Sus* and *Babyrousa* in particular, is likely to be the subject of future reconsiderations. There are indications concerning the existence of at least two more species: *Sus bucculentus* and *Sus heuneni*, which have been purported to inhabit South Vietnam and the Flores Islands (Indonesia), respectively. A recent paper based upon mitochondrial signatures, however, demonstrated that *Sus bucculentus* is not sufficiently different from *Sus scrofa* to deserve full species recognition (Robins *et al.*, 2006).

**Subfamily Babyrousinae**

**Genus Babyrousa**

Until recently, there was only one species of babirusa (‘babi-rusa’, Indonesian, babi = pig, rusa = deer): *B. babyrussa*. Based upon a reconsideration of morphological characters, Meijaard and Groves (2002) proposed that this single species be split into three extant species, *B. babyrussa*, *B. celebensis*, *B. togeanensis*, and a fourth extinct species from the southwest arm of Sulawesi which they call *Babyrousa bolabatuensis*. These designations are likely to be altered again when DNA evidence is brought to bear on the question.

The members of this genus differ significantly from Suinae in that they have long legs, relatively small heads and bodies mostly free of hair. The morphology of the stomach is complex. Grass comprises the main source of food and the typical digging or rooting behaviour of pigs is not known. The canine teeth in males are considerably more developed and are very large and curved, sometimes creating a spiral. The most unusual feature is that the mandibular canine and the maxillary canine teeth
are both pointed upwards, which is uncommon for mammals. Females usually deliver two offspring and the adults are large in size. Babirousas are spread across Sulawesi and nearby islands, and, because they can swim well, whether they arrived on nearby islands by themselves or were ferried by humans remains an open question.

Eleven of the autosome pairs and the X chromosome in this genus look very similar to the chromosomes of the domestic pig (Bosma et al., 1991a; see also Table 7.4). Future molecular phylogenetic investigations may give a clearer answer concerning the origin of the babirusa. A morphological study of the placenta and heart anatomy indicates that the babirusa has a significant number of traits in common with pigs (MacDonald, 1994).

Subfamily Suinae

Tribe Phacochoerini

GENUS PHACOCHOERUS. The common warthog (P. africanus), is widespread in many African countries. The name is derived from the large warts that are located on the muzzle, whose function is not known. The shape of the skull of the warthog differs essentially from that of other pigs and the number of teeth is greatly reduced. The canines are large, sharp and represented in both sexes. Adults are approximately 145–190 cm in length, 65–85 cm in height and weigh 50–150 kg. Average litter size is about three to four piglets. The offspring are susceptible to the cold immediately after birth and therefore do not leave their burrow, where the temperature is around 30°C constantly. Their main food is grass and the animals graze on their knees, which causes callus development. Adults, in contrast, to their offspring, enter the burrow backwards. The species is spread widely in sub-Saharan Africa, but not in the rainforests of Western Africa. In Southern Africa, this species has been reintroduced for hunting. The diploid chromosome number of the common warthog is 34 (Melander and Hansen-Melander, 1980).

The Cape and Somali warthog, P. aethiopicus, was recognized by zoologists as a separate species relatively recently (Grubb, 1993b). However, palaeontologists came to a similar conclusion many years ago, mainly because of the lack of functional incisors in P. aethiopicus. The common warthog has two incisors in the upper jaw and usually six in the lower jaw.

GENUS HYLOCHOERUS. The forest hog (H. meiertzhageni) is one of the largest wild pigs (length 155–190 cm, height up to 110 cm and weight up to 250 kg). The forest hogs of East Africa are particularly large (Grubb, 1993b). The head and muzzle are very broad, and the snout is big and well developed for extensive digging. This species and genus is a relatively recent discovery. It was found in Kenya during the early part of the 20th century and is now known to occur throughout the tropical forest region of Africa. The animals are covered with long black hair. The biology of this species is still under investigation.

Tribe Potamochoerini

GENUS POTAMOCHOERUS. P. porcus (red river hog) is one of the smallest African pigs. This species is widely spread in the central and southern parts of the continent and shows a significant variability in colour and size. The animals that live in West Equatorial Africa, for instance, are usually very bright and red in colour with a white bar on the back, white hairs on the muzzle and brushes of long hairs on the ears. Males from the majority of habitats are characterized by the canine apophyses (located between the ears and the nose), which in older males look like two small horns directed backwards. Length of the body varies from 100 to 150 cm. The height varies from 55 to 80 cm and the weight may reach 80 kg (Bannikov and Flint, 1989). The number of offspring is three to four. The skull of P. porcus is very much like that of Sus. Groves (1981) considered that the skull structure was an indication that these genera may be more closely related, or both have changed little from their more distant common ancestor. However, an absence of comparative cytogenetic (Bosma et al., 1991a) and molecular data makes it difficult to estimate phylogenetic distance from other genera.

The range of the bush pig (P. larvatus) is mainly to the east and south of that of the river hog and it occurs not only in East and South Africa, but even in Madagascar. It is also bristly, but bristly pelage extends from the head over
the whole body and gives the live animal a shaggy, crested appearance, which is different from that of the river hog. The two species are for the most part separated territorially, but in some places their areas may overlap (Grubb, 1993b). An introgression between the species is assumed (Kingdon, 1979). The limited amount of information about the biology and evolution of both of these species needs to be rectified.

**Taxonomy of the Genus Sus**

**Introductory remarks**

Sus most likely originated in Island Southeast Asia and then migrated into East Asia before heading west towards the Near East, North Africa and Europe (Pickford, 1993; Larson et al., 2005). The earliest fossil evidence of Sus (Sus arvernensis) is from the Late Miocene (∼6–5.3 mya) of Europe, but, as there is no direct known ancestor from this continent, it is considered that this lineage must have arrived by dispersal from Asia (van der Made and Moya-Sola, 1989; van der Made et al., 2006). Groves (1981) presented a comprehensive analysis of the taxonomy and phylogeny of the extant genus Sus based on morphological, palaeontological and biogeographical data. Groves and Grubb (1993) revised the systematics of the genus. This basic analysis and newly published data are summarized in the following section. DNA and morphological analyses suggest the existence of three evolutionary groups within Sus that could have diverged during the Pliocene – S. cebifrons/S. celebensis, S. barbatus and S. verrucosus/S. scrofa, with Sus ahoenobarbus as a new species within the latter group (Lucchini et al., 2005). This contrasts with the three groups (S. verrucosus, S. philippensis and S. scrofa) proposed by Groves (1981), as well as partially with other DNA studies that cluster S. verrucosus/S. barbatus as a sister clade of S. scrofa (Randi et al., 2002). Morphological studies of extant and extinct forms have divided the genus Sus into two groups, the primitive ‘scrofic’ and the derived ‘verrucosi’ (van der Made and Moya-Sola, 1989); other dental studies have also indicated a separation of S. scrofa from the rest of Sus (Genov, 2004). Several studies (e.g. those of Larson et al., 2005 and Mona et al., 2007) have pointed out that certain DNA sequences widely used to discover the relationships between taxa were not sufficient to discriminate between some species of Sus. At present, the genus Sus comprises seven species. Investigations of this genus conducted over the past 170 years have taken different approaches, and the number of discriminated species has varied from very few to a total of 37. Future classifications, based on classical and molecular phylogenetic data, are likely to alter the current knowledge concerning Sus taxonomy and evolution.

**Sus scrofa (Eurasian wild boar)**

The earliest known fossil of S. scrofa in Europe is from the Early Pleistocene (∼780,000 years ago) (van der Made, 1999; Franzen et al., 2000); that from Asia is also from the Early Pleistocene (J. van der Made, personal communication). The modern species exists in four forms: domestic livestock, domestic semi-wild, feral and wild (Genov, 2004). S. scrofa spreads naturally through vast territories, and covers most of Europe and Asia. The species was introduced into North and South America, Australia and Oceania. Domestic pigs are very common in the majority of countries worldwide, except for those that have religious restrictions. Several features, including tooth and skull morphology, external proportions, hair and colour patterns, biochemical and molecular polymorphisms, ecology and behaviour, reproductive isolation and natural areas, are used for discrimination of the many species in the genus.

Pigs are one of the most widespread mammalian species and S. scrofa is the primary ancestor of domesticated pigs, though other species may also have been involved (see Chapter 2). S. scrofa is extremely variable in the majority of traits studied. The number of subspecies is uncertain and depends upon the definition of the traits studied. However, it is possible to discriminate at least 16 more or less
distinct subspecies (Groves, 1981; Groves and Grubb, 1993):

<table>
<thead>
<tr>
<th>Subspecies</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. s. scrofa</td>
<td>Western, Central and parts of Southern Europe</td>
</tr>
<tr>
<td>S. s. attila</td>
<td>East Europe, northern slopes of Caucasus, parts of Western Siberia, Central and Western Asia</td>
</tr>
<tr>
<td>S. s. meridionalis</td>
<td>South Spain, Corsica and Sardinia</td>
</tr>
<tr>
<td>S. s. algira</td>
<td>North-west Africa</td>
</tr>
<tr>
<td>S. s. libica</td>
<td>Asia Minor, Middle East, southern part of Eastern Europe</td>
</tr>
<tr>
<td>S. s. nigripes</td>
<td>Southern Siberia, Central Asia</td>
</tr>
<tr>
<td>S. s. sibiricus</td>
<td>Eastern Siberia, Mongolia</td>
</tr>
<tr>
<td>S. s. ussuralicus</td>
<td>Russian Far East, Korea</td>
</tr>
<tr>
<td>S. s. moupinensis</td>
<td>Eastern China, South-east Asia</td>
</tr>
<tr>
<td>S. s. leucomystax</td>
<td>Japan</td>
</tr>
<tr>
<td>S. s. riukiuanus</td>
<td>Ryukyu Islands</td>
</tr>
<tr>
<td>S. s. taivanus</td>
<td>Taiwan</td>
</tr>
<tr>
<td>S. s. davidii</td>
<td>Western India</td>
</tr>
<tr>
<td>S. s. cristatus</td>
<td>Eastern India, western part of Indochina</td>
</tr>
<tr>
<td>S. s. affinis</td>
<td>Southern India, Sri Lanka</td>
</tr>
<tr>
<td>S. s. vittatus</td>
<td>Malaysia, Southern Indonesian Islands</td>
</tr>
</tbody>
</table>

The areas of these subspecies are close and the level of discriminating differences may be quite small, involving size, colour, proportions, skull characters and, in several cases, chromosome numbers. The variation in chromosome number is a result of two distinct Robertsonian translocations, which were found in the different geographical areas of the species (Tikhonov and Troshina, 1974; Bosma, 1976). The usual number of chromosomes in S. scrofa is 38 (Bosma et al., 1995). However, translocation I involving chromosomes 16 and 17 and translocation II involving chromosomes 15 and 17 were found in Kyrgyzstani and European boars (Tikhonov and Troshina, 1978), and reduce the number of chromosomes to 37 in crosses (heterozygotes) and to 36 in homozygotes.

Adaptations of these animals to different food and climatic conditions are dramatic. The flexible behaviour of the wild boar is perhaps one of the important features providing this adaptability. S. scrofa is well adapted to Siberian winters, tropical conditions, mountains and semi-deserts. Pigs can tolerate temperatures from −50°C to +50°C due to well-developed thermoregulatory and nest-building behaviour. Despite being under significant human and predator pressure, populations of wild boar and feral pigs are very numerous in many parts of the world (Choquenot et al., 1996).

Variations in body size are significant among subspecies. The largest subspecies are S. s. ussuralicus (males up to 300 kg) and S. s. attila (males up to 275 kg). Generally, mature weight is quite variable depending on age, sex, food availability, season and habitat. Body and head length is about 130–175 cm and height ranges up to 100 cm. The smallest forms of wild boar are from South-east Asia. Detailed descriptions of skulls, and osteometrical studies, have been published (Groves, 1981; Endo et al., 1994).

Current palaeontological knowledge regarding the evolution of S. scrofa is still limited, and fragmented data do not create a full-scale picture of the origin and phylogeny of the species. It is known that a fully evolved S. scrofa lived in the Biharian fauna in Europe and replaced the previously existing lineages of Sus strozzii, a possible descendant of S. arvernensis (Hünernann, 1969; van der Made, 1999). The facial shortening that occurred in S. celebensis and in S. scrofa has been used as a possible argument in favour of their common origin from perhaps the southern or the south-eastern regions of Asia (Groves, 1981). The spread of the two above-mentioned Robertsonian translocations does not contradict this possibility. However, it is obvious that an extensive molecular genetics study of the problem is necessary before any clear conclusions can be drawn. Thus far, an analysis of mitochondrial DNA from an extensive Old World sampling of Sus samples generally supported the subspecies listed above, though the resolution of these DNA studies was not sufficient to conclusively demonstrate the existence of all the named subspecies (Larson et al., 2005).
**Sus verrucosus** (Javan warty pig)

This species now lives mainly in Java. Two subspecies have been described (Groves and Grubb, 1993). The most typical common feature is three warts on a specific location of the muzzle, which are strongly developed in the adult males. Colour varies from overall black to a pale red. Size also varies from relatively large to small. Sexual dimorphism in size is greater than in other species. Despite some differences from *S. s. vittatus* and *S. celebensis* in G-banding and the structure of the Y chromosome, similarity is significant (Bosma et al., 1991b). The closeness of *S. verrucosus* and *S. s. vittatus* is supported by the observation and precise description of interspecies hybrids in nature (Blouch and Groves, 1990). Several morphological features make *S. verrucosus* close to the other South-east Asian species, *S. barbatus*.

**Sus barbatus** (bearded pig)

The common name of *S. barbatus* – the bearded pig – is due to the elongated whiskers around the muzzle from the mouth to the ears. A few warts on the muzzle are very typical. Mature size varies significantly between several subspecies and is on average close to that of *S. scrofa*. The length ranges from 100 to 160 cm and the weight is approximately 100 kg. Some males are much bigger. The bearded pig inhabits the Malaysian peninsula, Sumatra, Java, Borneo, Palau, Bangka, Palawan and some other islands. *S. barbatus* sometimes migrates, and these migrations involve thousands of animals. Fertile hybrids with *S. scrofa* obtained in captivity are known (Blouch and Groves, 1990).

**Sus celebensis** (Sulawesi warty pig)

This wild pig from Sulawesi and several other islands, including possibly Timor, has been recognized as a separate species from *S. verrucosus* by Groves (1981). Cytogenetic analysis strongly supported this point, though the structure of the Y chromosome has some differences from that of *S. verrucosus* (Bosma et al., 1991b). Animals are usually black with a few white or yellowish hairs intermixed, and they have crown tufts of hair. Other colour types have been described. The muzzle is short, like the Eurasian wild pig, and is of small size. Legs are also short, and the pig has small short ears with a relatively large head. This species is found on Sulawesi and other offshore islands. There are indications that *S. celebensis* was domesticated during the early Holocene and spread as far as Roti, a medium-sized island south-west of New Guinea, where the pigs are living now (Groves, 1981). In other places, they have probably been replaced by domestic *S. s. vittatus*. Based upon morphological characters, Groves originally claimed that the several forms of wild pigs in New Guinea could be a result of hybridization between *S. s. vittatus* and *S. celebensis* (Groves, 1981).

Recent papers based upon genetic evidence have suggested a different alternative. First, mitochondrial signatures obtained from *S. celebensis* samples on the island of Sulawesi are not monophyletic. Instead, two clades, one made up of samples from the northern arm of the island and one made up of southern samples, cluster within other *S. scrofa*, *S. barbatus* and *S. verrucosus* samples, suggesting not just two species, but that each may have arrived on to the island independently (Larson et al., 2005, 2007). Secondly, pigs from New Guinea possess a completely separate signature known as the Pacific Clade, and, though this does not rule out hybridization with *S. celebensis*, it strongly suggests that the maternal heritage of pigs on New Guinea and throughout the Pacific was derived from an Asian wild boar endemic to Peninsular South-east Asia (Larson et al., 2007).

**Sus philippensis** (Philippine warty pig)

According to the latest information, there are sufficient arguments to discriminate *S. philippensis* from *S. celebensis* and *S. barbatus* (Groves, 1981; Groves and Grubb, 1993). The species occurs on several islands of the eastern Philippines. The colour is black, sometimes with a pale snout band and red-brown patches in the mane. This pig is smaller than *S. barbatus*. Further investigations of the species are desirable.
Sus cebifrons (Visayan warty pig)

*S. cebifrons* is a small pig that occurs allopatrically to *S. philippensis* on the west-central islands of the Philippines (Groves and Grubb, 1993). Data on the biology of this species are very limited.

The pygmy hog: *Sus salvanius* or *Porcula salvanius?*

This is the smallest pig. The taxonomy is still controversial as follows from the discussion below. The pygmy hog does not have warts. Body and head length is 66–71 cm in males and 55–62 cm in females. The corresponding shoulder height is 23–30 cm and 20–22 cm, and weight is 9–10 kg and 6–7 kg in males and females, respectively (Mallinson, 1977). The basic colour is dark brown. Structure of the skull differs significantly from that of *S. scrofa*. The number of pairs of teats is three, instead of the six pairs typical of other *Sus* species, and the number of piglets born is usually three to four. Ears are large and rounded. The tail is very short and the inner toes are short compared with other pig species (Groves, 1981). The pygmy hog is currently distributed in quite a narrow part of northern Assam (India) in the long-grass belt. The number of animals in the area is very small. This species is, therefore, considered to be endangered. The diploid chromosome number is 38. Comparative analysis of G-bands shows that the chromosomes of the pygmy hog are very similar to those of the domestic pig and those of wild *S. scrofa*, which possess $2n = 38$ chromosomes (Bosma et al., 1983). Except for the small size of the body there are relatively few characters that may serve as diagnostic for discrimination of this species from *S. scrofa*.

The original description of this species by Hodgson in 1847 placed it as the sole species within the genus *Porcula*. This was overturned by Groves (1981), who used a series of morphological markers to assign the species as a member of the genus *Sus*. A recent molecular study of modern and museum specimens of this species, however, demonstrated that pygmy hogs possess a degree of genetic difference and that phylogenetic inference places them outside the *Sus* genus on a phylogenetic tree (Funk et al., 2007). Even though the tree did not possess enough resolution to confidently ascertain the relationship of *Porcula* to the other genera, its difference from *Sus* was robustly demonstrated. These observations led Funk et al. (2007) to suggest that the genus *Porcula* be revived and that Hodgson’s original classification was correct, although future reconsiderations using additional data may overturn this designation again.

Interrelationships of the species in genus *Sus*

A high level of morphological similarities between all species of the genus is an argument in favour of their relatively recent origin from a common ancestor. Their same chromosome number and their high level of homology support this conclusion. However, these close relationships complicate phylogenetic reconstruction. A possible phylogeny of the genus *Sus* is presented in Fig. 1.1 based upon an amalgamation of trees presented in a series of genetics papers.

It follows from the previous species descriptions that, in several cases, different *Sus* species coexist in the same area, yet have maintained significant differences in morphology, ecology and behaviour. This may be reasonably explained by a reproductive isolation that appears to exist between the species which may have contact. This is applicable to *S. scrofa* and the pygmy hog in northern India, and to *S. scrofa* and the ‘Indonesian’ species: *S. barbatus*, *S. verrucosus* and *S. celebensis*. None the less, interspecies hybridization with *S. scrofa* can occur, and this probably indicates a limited reproductive isolation (Groves, 1983). The production of fertile hybrids between a European wild boar and *S. barbatus* sows has been reported (Lotsy, 1922). Hybridization between *S. scrofa* and *S. verrucosus* has been recorded fairly recently in Java (Blouch and Groves, 1990). According to Groves (1996, personal communication) in some parts of the Philippines the indigenous wild pigs (especially *S. cebifrons*) are in danger of being hybridized
out of existence by crossing with feral domestic pigs. Groves (1981) assumed that the ancestors of these species were separated at least in the Middle Pleistocene, but this may have happened much longer ago.

A hypothetical scenario of Sus evolution has been suggested by Groves (1981). He proposed that the S. verrucosus–S. barbatus lineage, which has been present from the beginning of the Pliocene in Europe, entered Indonesia about 2 mya. It appears that these animals cohabited with the older S. celebensis lineage. S. scrofa may have possibly evolved out of the S. celebensis lineages and entered Europe about 700,000 years ago, where it replaced S. verrucosus-like pigs. The high level of similarity in chromosome structure of S. scrofa and S. celebensis does not contradict this hypothesis. Several independent sets of data support a Far East origin of S. scrofa and a steady spreading in a westerly direction. The previously mentioned Robertsonian translocations, which are typical for some Siberian, Central Asian and European populations, probably appeared and became fixed in the populations after or during their western movement. Numerous investigations have been devoted to comparing the geographical distribution of alleles for blood group antigens, isoenzymes and other proteins, but have not been directly used as arguments in resolution of the problem of the species origins (Gorelov, 1994). This story, generated from morphological and biological data, has recently been supported by DNA studies. The most likely scenario is that S. scrofa originated in Island South-east Asia, and migrated first across the Kra Isthmus into Peninsular East Asia, whereupon Sus radiated and diversified across East Asia. From there, the species spread west into Central Asia, the Near East and finally into Europe and North Africa (Larson et al., 2005).

Conclusions

The information presented in this chapter gives a general overview of the systematic position and phylogeny of the wild ancestors of the domestic pig. The family Suidae appeared during the evolution of the early Oligocene, some time after separation of the suborder Suiformes from other Artiodactyla. From a morphological point of view, the Suiformes are more primitive and less specialized.

Members of the Suidae have spread widely across Africa, Europe and Asia. Sus itself appeared in the Lower Pliocene at least 3–5 mya in Europe and Indonesia. A distribution of these ancestor species through the Indonesian Islands was possibly essential for speciation within the genus.

One of these species, S. scrofa, was tremendously successful and spread through Asia and Europe, replacing previous species. A number of more or less distinctive subspecies emerged, and some of them were independently involved in the domestication process that began over 10,000 years ago.

References


Introduction

Members of the Suidae family include five genera and as many as 15 different species, although the taxonomy of this group remains uncertain (see Chapter 1). Within this family, the genus Sus is represented by at least six species, the most geographically diverse of which is Sus scrofa. This species occupies a large number of ecological niches across the entirety of the Old World, and, though claims for the domestication of several species from this genus have been made (e.g. Groves, 1981), it is still generally accepted that only S. scrofa has been fully domesticated. It is certainly the case from morphological, behavioural and (most recently) genetic studies that almost every domesticated pig in the world today derives from a wild S. scrofa ancestor. Having said that, it is important to note that wild boar have diversified into numerous regionally distinct populations (sometimes classified as subspecies), and that several of these are likely to either have been independently domesticated, or have contributed genes to modern domestic pigs through hybridization.

The purpose of this chapter is to describe each of these populations in turn, the genetic evidence for domestication and the subsequent migration of populations of domestic pigs with people along migratory routes across the Old World. Though modern pig breeds are direct descendants of the earliest domestic pigs, the history of the most widespread commercial breeds and their differentiation from one another extends only 200 years into the past, a small fraction of the ~10,000 years since the first appearance of domestic pigs in the archaeological record. The historical development of modern breeds is interesting and has resulted in a tremendous degree of variation (see Chapter 18). This chapter, however, focuses instead on the deeper time of early
domestication in order to establish the origins of modern pigs.

Before doing this it is worth discussing the process of domestication itself. One of the main questions people ask about animal domestication is: ‘Why and when did humans first domesticate them?’ This question is based upon the premise that domestication was a goal oriented, intentional act carried out by people who saw in wild populations the potential of having smaller, more docile versions under their direct care. All of the available evidence suggests that this was the least likely scenario, and that the process was neither intentional nor rapid. Thus, before proceeding to the primary purpose of this chapter, we will explore what the terms ‘domestication’ and ‘domesticated animal’ mean, and how we can possibly hope to recognize the various stages of the process both in the archaeological record and in the genomes of modern domestic pigs.

The Process of Domestication

In his review of the role of animal behaviour in domestication, Price (1984, p. 3) rightly states that ‘it is difficult to formulate a definition of domestication that is general enough to account for the wide variation observed in different species, in different captive environments, yet specific enough to be meaningful in terms of the biological processes involved’. Nevertheless, he proceeds to attempt to define domestication as ‘that process by which a population of animals becomes adapted to man and to the captive environment by some combination of genetic changes occurring over generations, and environmentally induced developmental events reoccurring during each generation’. In a nutshell, Price’s definition highlights ‘captivity’ (i.e. direct human control) as a basic catalyst for the process.

Though the term ‘domestic animal’ carries universal meaning, the terminology typically employed by those studying these animals is often confusing and poorly defined. The primary reason for this stems from the inherent difficulty of assigning static terms to what is clearly a process likely to involve long-term and continuous change (Dobney and Larson, 2006). The terms ‘wild’ and ‘domestic’ have long been (and still largely are) terms that have been used to describe simple states of being, rather than what they actually represent – the extremes of a principally biological process driven by selection pressures, some of which are intimately linked with various and diverse aspects of human culture. As a result of this somewhat more complex view of human–animal relationships, a number of ‘intermediate’ stages of domestication have been proposed, e.g. ‘cultural control’ (Hecker, 1982; Hongo and Meadow, 1998, 2000), ‘pre-domestic’ (Vigne and Buitenhuis, 1999) and ‘intermediary stage’ (Ervynck et al., 2002).

A vast range of human–animal relationships has existed throughout history and the animals involved in many of these relationships cannot be easily categorized as strictly wild or strictly domestic. Nor is it credible that the range of species involved in these relationships took the same trajectory along the path that led to ‘complete’ domestication. The fluid nature of these relationships has led some authors to question whether the term ‘domestication’ (at least as it is traditionally perceived) can even be applied to some species. The unique biology and behaviour of, for example, pigs present special challenges to the study of their domestication that have caused some to question whether the threshold we term ‘domestication’ is really relevant to them (Jarman, 1976; Zvelebil, 1995).

Thus, any paradigm that relies on a strict wild/domestic dichotomy prevents a deeper appreciation of those animals whose lives are spent somewhere in between. More importantly, because this sort of dichotomous perspective rules out long-term evolutionary change as an explanation of the process of domestication, it therefore (however unintentionally) both obscures the existence of transitional forms and prevents any real understanding of the domestication process itself (Dobney and Larson, 2006).

It is perhaps best to model the process of domestication as a continuous trajectory divided into a series of diffuse stages during which differing selection pressures vary in intensity, subsequently allowing separate selective pressures to begin asserting themselves. Understanding
other naturally occurring and mutually beneficial/altering relationships that exist between other organisms (symbiotic, mutualistic, commensal, or even parasitic) is perhaps more useful in understanding the initial stage of this process, one probably not readily visible in the archaeological record and in which humans are likely to have played little or no direct role. Zeder (2006) has taken this idea a step further and envisioned the process as an evolving mutualism between populations of people and populations of plants and animals. At its most extreme, this idea suggests that humans, plants and animals have all evolved with a biological and/or cultural dependence on one another to the point where each is necessary for the other’s survival. Direct analogies can also be found in the plant world, as has been demonstrated recently by Fuller et al. (2010).

This dependency began with an initial phase of domestication, during which the relationship between humans and, for example, wild boar began to grow in intensity, and is likely to have predisposed them to eventual full domestication. As tamer wild boar ventured closer to human settlements (probably attracted by new scavenging opportunities created by human settlements in the form of waste and refuse), physiological, phenotypic and even genetic changes perhaps would already have begun. However, it was not until humans purposefully enhanced the selection pressure upon that behaviour (by captivity and control over reproduction, in turn leading to the enhancement and exaggeration of behavioural and phenotypic changes) that true pigs would become permanent members of human settlements and more readily recognizable (to our modern eyes) as domestic pigs. These pigs would have been significantly different from their wild progenitors at all levels of biological organization, from their genomes to their bones. Until recently, accessing the genetic evidence of the archaeologically preserved bones and teeth of long-dead organisms was not possible, and certainly not as well established as the zooarchaeological methods used to interpret the bones themselves.

Because domestication (like speciation) is a continuum, deriving a specific definition that differentiates wild and domestic animals is necessarily futile, but that fact makes it more difficult to recognize domestic animals in the zooarchaeological record. Animals at various points along the cline from wild to domestic may change their biology, behaviour and attitude towards humans in many different ways and to varying degrees, and it therefore follows that changes along the continuum cannot be analysed in an unequivocal manner. Instead, a more reasoned approach becomes necessary, one that employs a diversity of approaches and contextual arguments. In addition to the complexity of the domestication process per se, factors that provide the context for domestication (e.g. related climatic, environmental, geographical, chronological and cultural variables) must also be considered (Albarella et al., 2006a).

Any real understanding of domestication must initially be predicated on an appreciation of domestication as a process. Once it is understood first that wild and domestic are the terminal ends of a complex continuum, and secondly that plants and animals have played just as active a role as people in the process, then we can begin to answer the fundamental questions regarding when and how our modern world was populated by larger populations of domestic animals than people. Lastly, it is also worth recognizing the value of an approach that combines methods from a wide variety of academic disciplines, including archaeology and genetics. By adding multiple contexts, the understanding of domestication in general, and of pig domestication specifically, can only benefit.

Pig Domestication Across the Old World

A historical perspective

The distribution of pigs and their close relatives across Europe, North Africa and Asia (Groves, 1981; Oliver et al., 1993) means that, unlike other domestic animals whose wild ancestors are more regionally restricted (e.g. sheep and goats), a simple diagnosis of pig domestication based upon geography is impossible. The very earliest excavated and studied assemblages of animal bones from various European sites such as Danish shell middens, Swiss lake dwellings, Italian terramare settlements and others
included teeth and bones of pigs, and by the end of the 19th century it was recognized that at least two main forms, interpreted as wild and domestic pigs, were represented by the remains (Albarella et al., 2006a).

In his early attempts to distinguish between the two, Winge (1990) believed that prehistoric European domestic pigs (which, following earlier authors, he termed S. s. domesticus) were descendants of the wild boar of Europe, northern Asia and North Africa (S. s. ferus), while he thought modern domestic pigs in South-east Asia were derived from local wild boar that he suspected might be a different species (Sus vitatus). Before and immediately following the Second World War, work in the Near East began revealing very early evidence for agriculture stretching back to the beginning of the Holocene (~10,000 years ago). Along with sheep, goats and cattle, pigs appeared to be an important early domesticate, present on some of these early farming (Neolithic) sites (Flannery, 1983). As a result of these new data, the dominant view of the mid-20th century was that pigs (along with sheep, goats, cattle and domesticated cereals such as barley and wheat) were domesticated in the Near East and brought to Europe by immigrant farmers (Childe, 1958). A few, like Sauer (1952), preferred a South-east Asian origin for domestic pigs, but so little archaeological work had been done in this region that there was not enough evidence to support such a claim.

Although the majority of researchers believed in a limited geographical origin for most animal domestication – in the case of pigs, just the Near and Far East, later spreading west and east with Neolithic farmers – this hypothesis was challenged, and even Winge’s metrical separation of wild boar and domestic pigs was questioned (Higgs and Jarman, 1969). The picture was then further complicated by a case made for ‘intermediate’ or ‘semi-domestic’ pigs under extensive human control, suggesting that closer relationships between geographically widely distributed wild and the semi-domestic pigs could have occurred anywhere within the larger distribution, not just in the previously recognized domestication centres (e.g. Jarman, 1976; Zvelebil, 1995).

Within the last two decades, two competing models for the history of pig domestication have emerged:

1. Pig domestication took place in the early Holocene, during the early Neolithic (see Table 2.1) in a few geographically isolated locations (specifically the Near East and probably China) and the descendants of those early domesticated individuals were later dispersed with the spread of early farmers and stockherders (e.g. Rowley-Connwy, 2003).

2. Pig domestication took place within several additional (and independent) geographical regions outside the Near East and China and included Neolithic Europe (e.g. Zvelebil, 1995; Albarella et al., 2006a) and even the Jomon period in Japan (Nishimoto, 2003).

With the advent and application of new biomolecular and morphological techniques to both modern pig material and historical and archaeological remains, our understanding of both the geographical and temporal context of pig domestication and subsequent trajectories of livestock dispersal can now be refined, and in some cases radically altered. The following is a summary of the genetic and archaeological evidence for pig domestication as it currently stands. Figure 2.1 depicts nine possible pig domestication centres, which are discussed below.

A general perspective

Recent genetic research of modern extant Sus species has provided evidence for the second of the two hypotheses listed above. Mitochondrial sequences extracted from hair and blood samples of recent European and Asian wild boar and various breeds of domestic pig revealed a distinctive Asian clade and two European clades (Giuffra et al., 2000; Kijas and Andersson, 2001). The Asian clade consisted of Japanese wild boar, Chinese Meishan domestic pigs and some European domestic pigs. Of the two European clades, the first consisted of the majority of European and all Israeli wild boar, as well as most European domestic pigs (including a sample from the Cook Islands in the Pacific), and the second was made up of Italian wild boar. A molecular clock approach in these studies and a subsequent study (Fang and Andersson, 2006) demonstrated that the two clades separated long before the advent of domestication, and
Table 2.1. A list of archaeological terms used in the text, their definitions and the general time frames that are associated with each.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holocene</td>
<td>A geological epoch that began approximately 12,000 BP and continues today. A period intimately linked with the origins of agriculture and the rise of modern human culture</td>
</tr>
<tr>
<td>Mesolithic (or Epipalaeolithic)</td>
<td>A period beginning around 11,660 BP in the Near East and ending with the introduction of farming, the date of which varied in each geographical region</td>
</tr>
<tr>
<td>Neolithic</td>
<td>Followed the Mesolithic and is not a specific chronological period, but a suite of behavioural and cultural characteristics specifically linked with the origins and spread of farming</td>
</tr>
<tr>
<td>Pre-pottery Neolithic A and B (PPNA and PPNB)</td>
<td>The early Neolithic of the Near East, characterized by the lack of pottery (dating from 9600 to c.8000 BP)</td>
</tr>
<tr>
<td>Pottery Neolithic</td>
<td>The period of the Neolithic of the Near East, characterized by the first use of pottery (beginning around 8000 BP)</td>
</tr>
<tr>
<td>Linear Bandkeramic (LBK)</td>
<td>A major cultural entity of the European Neolithic, dating from ~7400 to 6900 BP. The geographical focus of the LBK was the middle Danube, the upper and middle Elbe and the upper and middle Rhine – considered the earliest Neolithic culture in Europe</td>
</tr>
<tr>
<td>Jomon</td>
<td>The longest and earliest recorded human cultural complex in Japan, lasting from ~12,000 to 6000 BP and characterized by semi-sedentary hunter-gatherer communities</td>
</tr>
<tr>
<td>Chalcolithic</td>
<td>Also known as the Copper Age or Eneolithic – this is a transitional period between the Neolithic and Bronze Age, linked with the beginnings of metallurgy. In the Middle East it begins around 6500 BP and lasts for approximately 1000 years</td>
</tr>
</tbody>
</table>

thus the genetics confirmed that at least two populations of wild boar (in western and eastern Eurasia) were domesticated independently, and that female Asian domestic pigs were later hybridized with European domestics (Giuffra et al., 2000).

A subsequent genetic study (Larson et al., 2005) of several hundred Sus mitochondrial DNA (mtDNA) sequences sampled across Europe and Asia revealed a surprisingly strong correlation between each sample’s geographical origin and the position of that sample on the resulting phylogenetic tree. This clear phylogeographical pattern (unique for any domestic animal) revealed a clear East–West cline, indicating an evolutionary origin for the genus Sus in Island South-east Asia (ISEA) and a subsequent westward dispersal of wild boar, culminating in the western corners of the European continent. This natural migration established a pattern of wild boar genetic variation across the Old World that today remains largely intact, and which proved significant when attempting to explore where (and in how many places) pig domestication occurred. The idea was that, if domestic signatures clustered with wild haplogroups were found in a specific region, then wild boar from that region must have either been domesticated, or at least contributed maternal DNA to domestic stocks potentially domesticated elsewhere. The results clearly demonstrated that not only did domestic pigs from Europe and Asia cluster with European and Chinese wild boar as expected, but also that some domestic pigs clustered with wild boar indigenous to Italy, India and peninsular South-east Asia. A separate population of feral pigs on New Guinea, Halmahera (Moluccan Islands) and Hawaii possessed yet another unique signature.

Though the phylogeographical structure evident in the phylogenetic tree was supported by the vast majority of samples, some of the pigs possessed haplotypes that did not match those of the region from which they were sampled. This was not entirely surprising given both the fact that pigs are behaviourally plastic and capable of migrating long distances on their own and the fact that humans have also
Fig. 2.1. A map depicting nine possible pig domestication centres. (1) Eastern Anatolia. Archaeological evidence from several sites in this region dated to 10,000 BP demonstrates a clear intensification of the relationship between humans and pigs leading to complete domestication. The first domestic pigs in Europe were from stocks originally domesticated in this region. (2) Modern China. Though the number of places where pigs underwent a similar transition to that in the Near East is disputed, pig domestication certainly began independently along the Yellow River in about 10,000–9000 BP. (3) Europe. Virtually all modern European pigs are derived from European wild boar, and, although this process appears to have begun at about 7000 BP, it is unknown whether it was independent or was kick-started by the introduction of Near Eastern domestic pigs into Europe. (4) South Asia. Though the archaeological evidence for domestication is weak, several modern domestic pigs from India and Bhutan share identical mitochondrial haplotypes with a native South Asian wild boar, suggesting either an independent domestication, or a significant mitochondrial introgression of Indian wild boar into introduced domestic stocks. (5) Peninsular South-east Asia. The South Asian circumstance of limited archaeological evidence but the presence of native wild boar DNA in domestic stocks is repeated here by the recent finding of a separate clade of pigs that possess local wild boar mitochondria. (6) Southern Yunnan Province, Northern Vietnam, Northern Laos. A divergent wild boar population in this region possesses a unique haplotype only found in the ‘Pacific Clade’ pigs spread across Island South-east Asia and the Pacific by early farmers. Curiously, no modern domestic pigs from this region possess that signature, most likely as a result of a more recent replacement by domestic pigs indigenous to China. (7) Italy. The Italian peninsula plays host to a genetically differentiated indigenous wild boar. Mitochondrial sequences identical to those of these wild boar have also been found in domestic and feral pigs in Sardinia and Corsica, suggesting introgression (intentional or otherwise) between imported domestic pigs and the Italian wild boar. (8) Domestic pigs originally from the island of Lanyu off the coast of Taiwan possess highly divergent mitochondrial and nuclear signatures, though thus far no wild boar from the island or any surrounding islands with matching signatures have been found. This suggests that these pigs were domesticated from a now extinct wild boar that was indigenous to the small islands in the East China Sea. (9) Sulawesi. The Sulawesi warty pig (Sus celebensis) is found only on the island of Sulawesi and a few outlying islands, where it was probably transported by humans. Some have suggested that these animals hybridized with incoming Pacific Clade pigs, and it remains possible, though there is no archaeological evidence, so far, that the warty pig was independently domesticated.
transported domestic pigs and occasionally even wild boar (Vigne et al., 2009). However, this observation meant that those pigs whose haplotypes did not match their region could be used as a proxy for human movement, and, in this context, two results from the tree were interesting (Larson et al., 2005).

First, as mentioned above, the distinctive pigs found on New Guinea, Halmahera and Hawaii, where no indigenous populations of wild boar were ever present before being transported by people, were distinct from the standard East Asian pig signatures found across China. This strongly suggested that these pigs were derived from a separate population of wild boar than those that had given rise to the ubiquitous East Asian breeds, and that they had then been taken to the islands as part of a dispersal by pig-herding farmers. Secondly, none of the domestic pigs in Europe possessed mtDNA signatures that matched those of wild boar from the Near East. The archaeological evidence for pig domestication in the Near East is well established (see below) though, if the first of the two hypotheses above were entirely true, then all European domestic pigs should possess Near Eastern signatures. The absence of any Near Eastern wild boar haplotypes in modern domesticated porkers from Europe suggested that, although pigs were definitely domesticated in the Near East, they must also have been domesticated in Europe, however independently, thus supporting the second of the two hypotheses (Larson et al., 2005).

As interesting as these conclusions were, because they were based entirely on samples that were no more than 200 years old, they meant that no empirical time frame for the domestication and migration of pigs could be ascertained. More simply, these data described where but not when domestication took place. Additional archaeological evidence (including the use of ancient DNA) was necessary to add a temporal dimension to these questions in the Near East, Europe, China, ISEA and Oceania.

**The Near East**

The zooarchaeological record of the Near East all but unequivocally demonstrates that the earliest pig domestication took place here. This evidence derives essentially from demographic changes that can be observed in Sus remains excavated from several key archaeological sites, principally in Eastern Anatolia (Turkey): Çayönü Tepesi, Hallan Çemi Tepesi, Hayaz Tepe, Tell Hallula and Görcütepe.

A unique stratigraphic sequence at the site of Çayönü Tepesi (which covers the entire Pre-pottery and Pottery Neolithic periods) reveals that pigs were killed at progressively younger ages through the entire sequence, which spans 10,200–7500 BP uncalibrated (Hongo and Meadow, 1998; Ervynck et al., 2002). In addition, visible changes to the skeleton and teeth (e.g. snout shortening and changes in conformation) and evidence for an increase in physiological stress (revealed by hypoplastic defects in the dental enamel) (Ervynck et al., 2002; Dobney et al., 2004) reveal the actual process of a closer human–animal relationship. The process appears to have been quite slow, occurring gradually over ~2000 years and apparently complete by the start of the Pottery Neolithic (~8000–7500 BP uncalibrated). Perhaps what is most interesting about the evidence from Çayönü is the fact that these changes appear not to have all been coeval; for example, there appears to have been a gradual shortening of the molar teeth through time, in contrast to a sudden shift to narrower molar teeth later in the sequence during the Pottery Neolithic period.

These intimately linked zooarchaeological data have been used to infer a more complex physiological and behavioural response in pigs, perhaps not originally driven by direct human intervention. In fact, Ervynck et al. (2002) suggested that possible early morphological and biometrical changes to the skulls of pigs at Çayönü were possibly related to changes in the rooting behaviour of wild individuals, perhaps attracted to human settlements by new food sources such as crops or human refuse. A long process of shifting pig–human relationships ensued, ultimately leading to direct control and full domestication of some wild boar by the Pottery Neolithic period (Albarella et al., 2006a).

Though excavations are still being carried out, Rosenberg et al. (1995, 1998) have investigated a huge collection of Sus remains recovered from the early Neolithic Turkish site of Hallan Çemi Tepesi (like Çayönü, situated in the Eastern Taurus mountains), and suggested
that a shift from a completely wild behaviour within a wild boar population to a way of living closer to humans may have taken place even earlier than at Çayönü. In a review of *Sus* data from sites more recent than Çayönü or contemporary with its later Pre-pottery phases, Peters et al. (1999) observed a decrease in the length of the third molar between Pre-pottery Neolithic B (PPNB) material from Çayönü and Late PPNB specimens from Gürcütepe (see Table 2.1). The authors claim this to be ‘unequivocal morphometrical evidence for the occurrence of domestic pigs’ in the Late PPNB (Albarella et al., 2006a). Peters et al. (1999) also highlight data from other Late PPNB sites, such as Hayaz Tepe and Tell Hallula, to substantiate their claim for the appearance of domestic pigs in that chronological period.

A genetics perspective on this region is difficult for two primary reasons, the first of which is that the dominant Muslim culture of the region has meant that very few, if any, modern pigs are kept, eaten or available to sample. Secondly, even if a modern population of Near Eastern domestic pigs existed, regions of the pig genome that correlate with the morphological changes described above have not yet been identified. The genetic window that has been opened has thus far focused primarily on the mitochondrial genome, though this is changing as high-throughput sequencing becomes a reality. The neutrally evolving mitochondrion is ideal for understanding population-level differences as it reveals a level of resolution invisible at the phenotypic level of an individual pig, and, though the insights into where and how many times pigs were domesticated that this genome has provided have been invaluable, understanding how the process resulted in domestic pigs will come once genes involved in the process have been identified and sequenced (see below).

Europe

The establishment of economic and cultural elements associated with Neolithic farming, including domestic pigs, took place across Europe between the 9th and 6th millenniums BP. How this process happened has been the focus of a great deal of research, the results of which have tended to focus on three primary hypotheses. According to these hypotheses, the European Neolithic may have resulted from either the direct migration of immigrant farmers from the Near East (known as demic diffusion; Childe, 1957; Pinhasi et al., 2005; Sampietro et al., 2007), the transmission of the idea of domestication and agriculture through established trade and exchange networks (cultural diffusion), or the independent development of agriculture (including the domestication of pigs and cattle) by indigenous European Mesolithic cultures (Renfrew, 1972; Clark and Price, 1981). It is worth pointing out that these explanations are not mutually exclusive (Zvelebil, 2000), as it may have been possible for both Near Eastern populations and ideas to have entered Europe, and then be taken up and expanded upon by local populations, though the degree to which either Near Eastern or ancient European populations have contributed to the make-up of modern European human genomes also remains contentious (Haak et al., 2005; Belle et al., 2006).

As mentioned earlier, Larson et al. (2005) not only revealed a complete absence of Near Eastern wild boar mtDNA signatures in modern domestic pigs sampled from Europe, but also showed that European domestic pig mtDNA haplotypes were identical to those sampled from modern/recent European wild boar populations. European wild boar populations must have begun contributing DNA to domestic pigs in Europe (even if those pigs were brought in from the Near East), though when this began remained an open question. In addition, the absence of Near Eastern signatures in modern domestic pigs strongly suggested that Near Eastern genetic lineages were either never introduced by farmers into Europe or, if they were, those mtDNA lineages have been all but completely replaced by those from pigs derived from European wild boar.

Larson et al. (2007b) attempted to answer these questions by sequencing mtDNA sequences extracted from the archaeological remains (bone and teeth) of *S. scrofa* from Europe and the Near East. The genetic sequences obtained in this study from ancient (and additional modern) wild boar were placed into a temporal framework that allowed for the
timing of movements of different mtDNA lineages of pigs to be determined. The relative geographical position and haplogroup affinity of 20 pre-Neolithic (13,000–7500 BP) wild boar samples from Europe were identified. The observation that all of these samples possessed European haplotypes (Larson et al., 2007b) supported previous observations of a phylogeographical boundary between Near Eastern and European wild boar haplotypes (Larson et al., 2005), and suggested that this biogeographical boundary was in place during the early Holocene. Additional evidence came from four wild individuals dated from Neolithic and post-Neolithic (Chalcolithic) sites in Romania close to the modern phylogeographical boundary that also possessed European haplotypes. Although the exact location of this boundary is difficult to establish, the appearance of a Near Eastern pig haplotype in two supposed ‘wild’ S. scrofa specimens from Mesolithic and Neolithic sites in the Crimea indicated that the distribution of Near Eastern haplotypes probably extended to the north shore of the Black Sea.

Unique Italian genetic sequences, originally identified by Giuffra et al. (2000), Kijas and Andersson (2001) and Lattuada et al. (2009), were found both in samples of Mesolithic Italian wild boar and in Mesolithic Croatian and Medieval Sardinian samples, thus extending the historical geographical range of this haplogroup and suggesting that Italian wild boar had been transported to Sardinia, though it was impossible to say whether the original pigs were wild or domestic (Larson et al., 2007a).

Later (Neolithic) S. scrofa mtDNA sequences from Europe clearly demonstrate that pigs with Near Eastern haplotypes did cross the phylogeographical boundary, the clear result of a human-mediated dispersal. Eleven pig specimens (identified by standard zooarchaeological criteria as domestic), from four Neolithic Romanian sites dating to 7500 BP, possessed a Near Eastern haplotype identical to that found in a single recent modern boar from Turkey and two from Iran. Five so-called wild specimens (based on their large size) from the same site possessed European haplotypes, suggesting that, though local farmers were herding pigs whose ancestors were in the Near East, they were hunting and eating local wild boar. The same Near Eastern haplotype was also identified in four specimens from the 8th millennium BP Linear Bandkeramik (LBK) site of Eilsleben in northern Germany, and in two samples from the mid Neolithic (very early 6th millennium BP Chasséen culture) site of Bercy in the Paris Basin. At Bercy, mtDNA sequences of European origin were also extracted from archaeological specimens identified by zooarchaeological criteria as domestic swine, making it the earliest site where both Near Eastern and European domestic swine have been identified together.

The results described above clearly demonstrate that Near Eastern-derived domestic pigs were physically moved westward with early farmers into Central and Western Europe during the Neolithic. However, after the mid-6th millennium BP, Near Eastern Sus lineages vanished, perhaps in as little as 500 years. Thus, by at least the 6th millennium BP, European wild boar had already been domesticated and the domestic form spread throughout Europe, replacing the original introduced pigs of Near Eastern origin. Whether this replacement was the result of continual hybridization by the only locally available source of wild boar, thus minimizing and then eliminating the genomic input of Near Eastern wild boar, or a more active selection against those pigs who displayed Near Eastern affinities is not yet known, though again, with the application of increasingly robust sequencing powers, this question will hopefully be answered in the next decade.

The zooarchaeological record provides additional important evidence for this complex process. Although pig remains are comparatively scarce relative to those of other domestic animals from Neolithic Romanian sites (Bilişescu et al., 2005) and from most LBK sites in central and western Europe (Tresset and Vigne, 2004), they increase in abundance (Tresset, 2003) and size (Tresset and Vigne, 2007) starting in the first part of the 7th millennium BP, well after the initial introduction of Near Eastern domestic pigs. There is also a clear increase in the frequency of physiological stress (as evinced by incremental defects in the dental enamel, known as enamel hypoplasia) between S. scrofa remains of Mesolithic and Neolithic date (Dobney et al., 2004).
Interestingly, the replacement of Near Eastern domestic pigs by European populations was not restricted to Europe. A series of archaeological mtDNA sequences extracted from the 7th to the 4th millennia bp through to the early Iron Age Sus remains from Armenian archaeological sites revealed a clear transition through time from Near Eastern to European haplotypes. This turnover, analogous to that seen in European sites, suggested that, by the 9th century bp, European domestic pigs had spread eastward to Armenia replacing the indigenous domestic pigs of Near Eastern ancestry (Larson et al., 2007a). This transition probably reflects the major reorganization that occurred in the various Neo-Hittite polities during the 3rd millennium bp, involving the large-scale movement of peoples and the expansion of trade and exchange networks during the later Iron Age (Smith, 2003).

The presence of the unique Italian haplotypes (mentioned earlier) in modern ‘wild’ pigs of Sardinia, which, like Corsica, was an island devoid of pigs before the arrival of Neolithic settlers (Vigne, 1992), supports the possibility that indigenous Italian wild boar may also have been separately domesticated (Albarella et al., 2006b). The Italian haplotype was also identified in numerous ancient samples, including Mesolithic wild boar from Pupicina cave in Croatia, early and mid Neolithic pigs from Grotta della Madonna cave in south-western Italy, a 4th millennium bp (middle Bronze Age) Sardinian site, and numerous medieval wild boar from Tuscany and Rome. These samples indicate that not only were indigenous Italian wild boar distributed beyond their current restricted region of Maremma in north-west Italy, but the presence of the Italian haplotype in Bronze Age central Sardinia also suggests either an independent domestication of native Italian wild boar or the incorporation of female Italian wild boar into domestic stocks that were subsequently imported to Sardinia by at least the end of the 4th millennium bp.

The zooarchaeological evidence supports these conclusions. In a summary of prehistoric pig exploitation in the Italian peninsula and Sicily, Albarella et al. (2006b) found a broadly consistent diachronic pattern of change in Sus body shapes across Italy. During the Mesolithic, Italian wild boar were relatively small (though their bones were quite large in relation to the teeth). This pattern persisted in the Neolithic, although shifts in pig exploitation at some sites were interpreted as the beginning of local domestication. A clear morphological separation between wild and domestic pigs is not evident until the Bronze Age, when domestic pigs decreased in size and wild boar appear to have increased (Albarella et al., 2006b).

The genetic and zooarchaeological records therefore reveal a complex temporal and geographical pattern of changes in Holocene Europe. These data provide unique insights and answers to specific questions about the nature of the European Neolithic revolution. First, the presence of Near Eastern haplotypes in Neolithic contexts in Romania, Germany, France and Croatia demonstrates that Near Eastern pigs were introduced by early farmers into Europe. Secondly, given the time frame of the initial introduction of Near Eastern domestics and the first appearance of domestic pigs derived from European wild boar, it appears that European pig domestication may not have been a truly independent event but rather a direct consequence of the initial introduction of Near Eastern domestic pigs (and other animals) into Europe by early farmers. If true, the process of pig domestication in Europe (and possibly the degree of intentionality among early farmers) appears fundamentally different from that in the Near East. Regardless of the specific cause or progression of European pig domestication, what is clear is that, once European wild boar were domesticated, they rapidly became the predominant lineage within European, and later South-west Asian, domestic swine.

China

China has been accepted as a primary and independent centre of early pig domestication, and recent studies in East Asia have highlighted the antiquity of plant and animal domestication and the generally greater amount of genetic variation found in East Asian wild boar and pigs relative to their European counterparts (Fang et al., 2005, 2009; Megens et al., 2008). Early agricultural activities practised by seasonally mobile cultivators were well established along the Yellow River and Inner...
Mongolia by ~8000 BP (Barton et al., 2009; Fuller et al., 2009; Liu et al., 2009), and plant cultivation may have even begun 2000 years earlier on the hilly flanks of the Yellow River Valley (Lu et al., 2009). In southern China, it was sedentary hunter-gatherers (Fuller and Qin, 2009) who first began cultivating rice along the Yangtze river at about 9000–8000 BP, a process that culminated in the dependence upon domesticated rice agriculture by ~6000 BP (Fuller et al., 2009). Although dogs may have been the earliest domesticated animals in these regions, archaeological evidence suggests that domestic pigs were soon prevalent in both northern and southern China by at least 8000 BP (Yuan and Flad, 2002; Flad et al., 2007). In both regions, however, pigs make up only a small percentage of the earliest mammal bone assemblages, which are primarily dominated by remains of hunted deer (Flad et al., 2007; Jing et al., 2008).

Long-standing claims for very early (~12,500 years BP) domestic pigs at, for example, the site of Zengpiyan near Guilin, have more recently been refuted by reanalyses of the zooarchaeological record by a new generation of Chinese zooarchaeologists (e.g. Yuan and Flad, 2002). Recently, systematic and traditional analyses of previously excavated Chinese zooarchaeological assemblages and subsequent data syntheses – along with the discovery of new sites – have led to the proposal of a new model for pig domestication in China: one that involves multiple and independent centres focused upon the Yellow River and on the Lower Yangtze river around 8000 years ago (Jiang, 2004). The sheer geographical scale of China, which has revealed a substantial diversity of Neolithic cultural entities scattered within a mosaic environment where at least two subspecies of wild boar exist (Groves, 1981, 2007; Groves and Grubb, 1993), makes such a model highly plausible. However, understanding the natural and cultural dynamics behind the early pig domestication process in China requires further integrated studies in which novel morphometric and genetic approaches have proved invaluable.

As mentioned previously, a very strong phylogeographical structure among wild boar was revealed using a phylogenetic tree, with a clear cline from East to West across the Old World (Larson et al., 2005). However, while the western Eurasian and ISEA branches of that tree consisted of well-supported clades, the portion of the tree that comprised samples from the East Asian mainland was defined by a large polytomy of clades and individual branches. As a result (and because of the paucity of Chinese samples in the study) it was impossible to say much more beyond the fact that pig domestication had taken place, at some point in the past, somewhere in East Asia. A series of larger genetic studies was therefore undertaken in an effort to refine the geographical and temporal context of East Asian pig domestication by adding more modern and ancient East Asian wild boar and domestic pigs sequences (Larson et al., 2007b, 2010). In addition to the genetic research, several complementary morphological proof-of-concept studies – involving tooth shape analyses of both recent and archaeological Sus specimens – were first piloted in an attempt to explore the potential complementarity of such an approach (Larson et al., 2007b; Cucchi et al., 2009, 2011). Though a true understanding of pig domestication is only possible using a combination of both approaches, the results of genetic and morphological approaches will be explored separately below.

A recent study (Larson et al., 2010) combined over 1000 modern wild and domestic pig samples with the results from an ancient DNA analysis that generated sequences from nearly 20 specimens spanning the last 9000 years. The data revealed 34 haplotypes found only in wild specimens, 99 haplotypes found only in domestic specimens, with 20 haplotypes shared between 86 wild and 581 domestic pigs. In addition, 25 separate clades that consisted of at least two haplotypes were present on the phylogenetic tree, although most of the haplotypes were arranged as a large polytomy referred to as the general cluster.

Contrasting the phylogenetic position of these S. scrofa sequences with their geographical provenance revealed several interesting patterns. The domestic specimens found within the general cluster were present throughout East Asia, and many of them represent the most common Asian pig haplotypes found today in globally distributed modern breeds (Larson et al., 2007b). In contrast, the general cluster haplotypes found
only in wild boar, and those that were shared between wild boar and domestic pigs, were more restricted geographically; they were found mostly in central China, or in the neighbouring Yunnan Province and Bhutan.

The geographical patterning of the specimens found in the 14 clades outside the general cluster made up only of wild boar and in the three clades made up of wild and domestic pigs showed a different pattern. Of the wild clades, four were restricted to islands including Japan, Okinawa, Taiwan and the southern Chinese island of Hainan. Three clades were found only in South Korea. One clade was restricted to North-east Asia (not including South Korea), one to Central China (though two samples are also found in northern Vietnam), and five clades were found only in samples from Indochina.

Of three clades that possessed haplotypes found in both wild and domestic samples, the first included both native wild boar and domestic pigs from India and Nepal (Larson et al., 2005; Tanaka et al., 2008). The second clade, referred to as the Pacific Clade by Larson et al. (2005), included six samples found in northern Vietnam, Yunnan Province and Laos. Despite additional sampling of domestic pigs from Central China and Indochina, the only domestic or feral pigs that belonged to this clade were found in Oceania. The third clade, referred to as MTSEA, given their mountainous and South-east Asian distribution (Tanaka et al., 2008), consisted of both wild and domestic samples found almost exclusively in Indochina, although one domestic sample from Bhutan and one sample described as wild from Taiwan were also found within this clade.

Network analysis revealed a separation of the clades that consisted solely of wild specimens from those of domestic or mixed origin, and the addition of 18 archaeological specimens to the network did not alter this pattern. Also, though three of the nine ancient haplotypes were novel, the six remaining haplotypes were identical to the most common haplotypes found within all modern East Asian pigs. In fact, the ancient samples derived from archaeological sites spanning 5000 years possess five of the seven most common haplotypes shared by modern wild and domestic East Asian pigs.

These genetic results (Larson et al., 2007b, 2010) clearly demonstrate that the most common modern domestic haplotypes found within central China are also the most common Asian haplotypes found across East Asia, in Australian feral pigs, and in modern European and American breeds, the latter as a consequence of the 18th-century drive to improve European breeds by hybridizing them with imported Asian pigs (Jones, 1998; Giuffra et al., 2000; Yang et al., 2009). The lack of fine geographical resolution of the mtDNA data across central China (probably resulting not just from a lack of genetic differentiation between wild boar populations in this area, but also from a history of human-mediated movement that has blurred whatever signal may have originally existed) precludes any conclusions regarding multiple centres of domestication here, although the identification and typing of distinguishing markers in the nuclear genome in ancient samples may generate enough resolution to address this question.

The position of ancient DNA sequences among the most common haplotypes on the network clearly supports recent zooarchaeological conclusions that purport modern Chinese pigs to be the direct descendants of the original populations of domestic pigs sampled along the Yellow River (Jing et al., 2008). Though this evidence does not rule out separate pig domestications from geographically and genetically differentiated wild boar in other parts of Central China, it does demonstrate a long-term genetic continuity between early and modern domestic pigs. In addition, neither modern nor ancient pig samples in this study share any genetic affinity with modern wild boar from Central China. The combination of a shared geographical distribution but maintenance of a strict genetic differentiation suggests not only that domestic haplotypes have not leaked into wild populations, but also that neither ancient nor modern herders of domesticated pigs made a lasting effort to incorporate the females of other separate wild populations into their domesticated stocks.

The evidence from the countries surrounding China, however, seems thus far to suggest that, regardless of how many times wild boar were domesticated in what is now China, there were no analogous processes of domestication
on the Korean peninsula, in Japan, on the Ryukyu archipelago or on the island of Taiwan. Although each of these regions harbours genetically differentiated populations of wild boar that probably migrated to these places during periods of lowered sea level, the domestic pigs that currently occupy these places derive from the Chinese domestication process described above (Larson et al., 2010).

There is, however, one notable exception to this pattern. Microsatellite and mtDNA from pigs originally from the tiny island of Lanyu, off the south-east coast of Taiwan, have revealed that this population is significantly different from all other wild and domestic pigs, though they are more similar to east Asian populations than to those from ISEA (Wu et al., 2007; Chang et al., 2009). More intriguingly, the island of Lanyu does not play host to a population of equally unique wild boar, suggesting either that the wild ancestor of these pigs was endemic to Lanyu and has since been exterminated, or that Lanyu domestic pigs were derived from an as yet undiscovered population of wild boar. Either way, this population represents the sole indigenous domestic variety found among islands and land masses east of China, and is deserving of a great deal more study.

Although ancient and modern DNA samples are proving to be one of the most useful tools for investigating domestication and human-mediated dispersal of domesticated animals, the mtDNA and nuclear markers used in palaeogenetics are often selectively neutral and, therefore, incapable of detecting responses related to rapid phenotypic changes in the skeleton associated with the domestication process (Bradley, 2006). In addition, the preservation of DNA is also a crucial limitation in palaeogenetics. The Near Eastern cradle of domestication, for example, is characterized by high and fluctuating temperatures that have reduced the availability of ancient sequences (Bollongino and Vigne, 2008). Molar shape analysis is a complementary tool that can be used to assess S. scrofa variation using geometric morphometrics to track phenotypic responses associated with both the process of domestication and its geographical origins.

A recent study (Cucchi et al., 2011) investigated Sus remains from three early Neolithic sites in mainland Northern, Central and Southern China (Xinglongwa in Hebei Province, Jiahu in Henan Province and Zengpiyan in Guangxi Province), two of which (Xinglongwa and Zengpiyan) have been previously used to support claims for early pig domestication (Nelson, 1998; Yuan and Flad, 2002). Pig remains from the late Neolithic site of Xishuipo (Henan Province) were also included in the analysis as comparative data as their domestic status is unequivocal. The evidence for pig domestication at Xinglongwa (in the far north-east of China) is based on an association between two complete S. scrofa skeletons and a wealthy human burial dated from the latest phase of occupation (8000–7400 BP) (Jing, 2006), and large collections of pig skulls found in subterranean buildings, all of which possess unambiguous, circular holes in the frontal bone associated with poleaxing.

As mentioned above, excavations in Southern China during the 1970s at the limestone cave of Zengpiyan in Guangxi Province resulted in claims for the earliest evidence of pig domestication in China (Chang, 1986). The site has recently been re-excavated (Institute of Archaeology, 2003) and a reinterpretation of the sequence has resulted in finalizing the dates of occupation as between 12,000 and 7000 BP. Reinterpretation of the vertebrate assemblage from Zengpiyan has revealed a subsistence economy most likely focused exclusively upon hunting and gathering throughout the occupation sequence, although a change in ceramic and lithic technologies observed in the latest phase indicates a potential shift towards cultivation (Institute of Archaeology, 2003). The Sus remains, however, are from small individuals, and thus their wild, domestic or semi-domestic status remains uncertain.

Two additional sites in East Central China (Jiahu and Xishuipo, Henan Province) are from a key region of the central Yellow River valley, both of which have produced important zooarchaeological collections. Jiahu has been conventionally radiocarbon dated to 9000–7700 cal (calibrated years) BP (Zhang et al., 1999), confirmed by thermoluminescence (TL) dating (Yang et al., 2005). It is perhaps one of the most significant Early Neolithic sites in China, and is of particular importance in our understanding of the development of Neolithic
civilization in the Yellow River basin (Zhang and Wang, 1998; Zhang et al., 2004), with evidence of early writing (Li et al., 2003) and early development of music (Zhang et al., 1999), and where evidence of the earliest rice domestication in Northern China remains hotly debated (Fuller et al., 2007). The late Neolithic site of Xishuipo is located on the Northern shore of the Yellow River and is dated to 5000–3000 BP (Kesner, 1991).

Shape analyses of more than 100 Sus lower second molars from both modern comparative and archaeological specimens were undertaken and, although the data set was relatively small, important conclusions could still be drawn. First, the data demonstrated that molar tooth shape can be used as an effective phenotypic marker, and can provide highly significant discrimination between wild and domestic forms (Cucchi et al., 2011). Secondly, the analysis revealed that the Zengpiyan pigs displayed a typically wild second molar morphotype, contradicting the earlier claims of the presence of pig domestication at Zengpiyan (Nelson, 1998) and supporting more recent claims that the pigs were in fact wild boar (Yuan and Flad, 2002). The Zengpiyan wild morphotype was not identical in shape or in size to that represented by the modern wild boar comparative data used in this study, however, suggesting that different wild Sus morphotypes existed, and may still exist, across China, a result supported by the genetic studies discussed above. Indeed the, North-eastern Chinese wild boar comparative specimens used in this study originated from the current geographical range of S. s. ussuricus, recognized as the largest subspecies of S. scrofa, whereas Zengpiyan is situated in (wetter and warmer) South China within the current range of the southern subspecies, S. s. moupinensis (Groves, 2007). This southern subspecies is today described as ‘fairly small but with a broad and high crowned skull’ (Groves, 2007), a description that fits with the dental traits identified by Cucchi et al. (2011) at Zengpiyan.

Morphometric signatures from the Sus remains from Xinglongwa in North-east China also revealed the presence of a clearly wild morphotype that showed strong phenetic relationships with the modern S. s. ussuricus specimens included in the study (Cucchi et al., 2011). These data suggested that the Xinglongwa pigs derive from locally hunted wild boar and that the wild boar phenotype in Northern China has been conserved without major morphological change for at least the last 8000 years. This, and other evidence at Xinglongwa, supports the idea of an early symbolic (as well as economic) value of wild boar, and the practices associated with this more intimate relationship perhaps led ultimately to an independent domestication of pigs in North-east China.

The story from the early Neolithic site of Jiahu appears to be different. Molar shape analyses by Cucchi et al. (2011) demonstrated that Jiahu pigs and domestic pigs from the later site of Xishuipo have clear domestic second molar morphotypes that, though similar, are not identical. Both data sets display a common secondary divergence trajectory indicating that both populations have undergone the same directional phenotypic evolution. On the basis of these data, it has been argued that as early as 10,600 BP, the pigs at Jiahu were clearly in the process of undergoing morphological change triggered by domestication (Cucchi et al., 2011).

The evidence from the recent genetic and morphological analyses confirms the suspected independent domestication of wild boar in China, though so far neither of the types of analysis has been able to definitively state how many different times in how many different places the process of ever-tightening dependence of wild boar and people took place. As more samples are analysed, and as the resolution generated by each of these techniques becomes ever greater, it seems likely that the ongoing uncertainty will be significantly reduced, perhaps to the point where definitive answers are uncovered. Evidence for additional centres of East Asian pig domestication outside China is emerging as well, and is discussed in the next section.

Peninsular South-east Asia, Island South-east Asia and Oceania

This region is one of the most important and interesting for studying the range of different relationships between pigs and humans, not
only because the Sus genus probably originated in ISEA, but also because the genetic diversity here is greater than anywhere else (Groves, 1981, 1983). Despite the near ubiquitous presence of Sus remains from Holocene archaeological sites in mainland East Asia, ISEA and Oceania, however, comparatively little is known about the temporal context of pig domestication, the species involved, or their involvement in the different human diasporas of the region (Hardjasasmita, 1987; Lucchini et al., 2005).

Pig domestication in mainland Asia is relatively simple given that only one species (S. scrofa) could have been involved. In ISEA, however, a variety of Sus species (S. scrofa, S. verrucosus, S. barbatus, S. celebensis, S. cebifrons and S. philippensis – see Chapter 1) are endemic, all of which were hunted by humans in the past and presumably had the potential to be domesticated. The taxonomy of these island suids remains contentious, as does their species identification and wild, feral or domestic status within the fossil record (Cucchi et al., 2009). To make matters worse, interspecies hybridization between introduced S. scrofa and other indigenous species has been claimed (Blouch and Groves, 1990). Groves (1981) has also stated on the basis of morphometrics that the feral and domestic pigs of New Guinea were hybrids of S. scrofa and S. celebensis.

The previously discussed genetic study (Larson et al., 2005) of modern wild and domestic Sus supported Groves’ (1981, 1983) conclusions stating that the genus Sus originated in ISEA and also revealed evidence for additional pig domestinations in South-east Asia. First, one haplotype indigenous to wild boar from peninsular South-east Asia was also typed in Australian feral pigs from northern Queensland. Australia was devoid of any Sus species until pigs were introduced by post-European contact settlers. The current feral population is therefore made up of escaped domestic pigs that are mostly European, though the presence of the South-east Asian haplotype complicates the picture by suggesting either that indigenous South-east Asian pigs were independently domesticated and transported, or were possibly moved into Australia as wild animals. Neither scenario is supported by any historical evidence, and the absence of the South-east Asian haplotype in between Australia and Thailand only further confounds the issue. Secondly, the data revealed another phylogeographical anomaly: the presence of a significant number of Sus samples on islands east of the major biogeographical boundary known as Wallace’s Line that possessed mtDNA signatures consistent with an East Asian mainland origin. These haplotypes, collectively referred to as the Pacific Clade (Larson et al., 2005), were present in every pig sampled from New Guinea, the eastern Indonesian island of Halmahera and in several specimens in remote Oceania as far east as Hawaii. Because these haplotypes clustered with others indigenous to East Asia, it followed that a population of wild boar, separate from those found in China, had been domesticated and moved through ISEA and into the Pacific islands (possibly associated with the Austronesian expansion), though, because no mainland samples possessed these haplotypes, determining precisely where this domestication had taken place was not possible.

Current archaeological evidence suggests that, once established, intensive, sedentary agriculture (including rice, millet and pigs) expanded across northern Asia in about 6500–5000 BP and across South-east Asia between 5000 and 4000 BP during the Late Neolithic (Chang, 1986; Liu, 2004; Bellwood, 2006). Though the archaeological evidence for the domestication of pigs in central China is relatively well established (see above), there is currently no indication of domestic pigs in peninsular South-east Asia until the end of the 5th millennium BP, when they are said to appear alongside the first evidence of sedentary agriculture (Higham, 1975, 2002b). Domesticated pigs and rice are also claimed to be present in South-east Asia (Thailand) no earlier than ~4000 BP (Higham, 1975, 2002b).

Despite the lack of archaeological corroboration, the significant genetic variation present in wild and domestic pigs in Indochina has previously been used to propose an independent centre of pig domestication somewhere in peninsular South-east Asia (Jing et al., 2008). More recent genetic evidence (Larson et al., 2007b) clearly demonstrated that some breeds of domestic pigs share haplotypes with two
clades of differentiated wild boar that are currently found only in Indochina, thus suggesting the domestication of wild boar lineages in this region, though the details and temporal context of this process remain unknown. What is clear, however, is that, once domesticated, Pacific Clade pigs were dispersed by people into ISEA, Wallacea and the Pacific.

The sole sequences that rooted the pathway of Pacific Clade haplotypes in mainland Asia were reported from two pigs from northern Vietnam (Hongo et al., 2002). Four more Pacific Clade sequences have been identified in modern wild boar from Laos and Yunnan Province, China (Larson et al., 2010). These samples expand the geographical range of wild boar possessing Pacific Clade signatures and add support to the hypothesis that the Pacific Clade is indigenous to peninsular South-east Asia. This evidence also supports the postulated Neolithic expansions of Austro-Asiatic language speakers along the major South-east Asian rivers from Yunnan (Higham, 2002a). The fact that no modern domestic pigs possessing Pacific Clade haplotypes have yet been found in mainland Asia is most likely a consequence of a replacement of native pigs by pigs introduced from China, a situation analogous to the previously discussed replacement of Near Eastern domestic pigs by European domesticates during later prehistory (Larson et al., 2007a).

There have been several demographic expansions of agricultural populations into peninsular South-east Asia that could have resulted in the importation of domestic pigs from (for example) central China. These include Austronesian speakers through ISEA and parts of the mainland coastal regions (Pawley, 2003), post-Neolithic expansions of Sino-Tibetan speakers (Van Driem, 1998; Pawley, 2003) and Austro-Tai or Miao-Yao groups from Southern China (Blench, 2005). Importantly, the replacement did not extend beyond mainland South-east Asia, leaving intact the populations of introduced domestic Pacific Clade pigs in ISEA.

With the exception of the wild boar sequences mentioned above, all other Pacific Clade pigs appeared to be scattered throughout ISEA and the Pacific (Larson et al., 2010), though four Pacific Clade pigs were identified in Sumatra and Java (islands on which indigenous wild populations of Sus existed well before the Neolithic), while Pacific Clade pigs also made up 15 of 19 specimens from eight islands east of Wallace’s Line in the Lesser Sunda chain and the Moluccas, as well as 100% of New Guinean samples. This pattern indicates that, like Australian ‘wild pigs’, the New Guinea feral pig population is also descended from mainland Asian domestic pigs, though in this case the original population was originally introduced by farmers carrying Pacific Clade domestic pigs travelling east over Wallace’s Line.

Perhaps more important in ascertaining the migration route that pigs and humans took to reach New Guinea is the fact that no Pacific Clade pigs have yet been identified from Taiwan (despite numerous samples from both native wild and domestic modern pigs and an ancient domestic sample), from the 40 wild samples from the Philippines (identified as endemic S. philippensis), or from the 17 introduced domestic samples from two central Philippine islands (Panay and Cebu).

In order to provide at least some temporal context for pig dispersal in Oceania, ancient DNA was also successfully extracted from five archaeological pig specimens from purportedly pre-European contact sites in Polynesia (from Tubuai, Hanamiai in the Marquesas and the Tangatatau rock shelter in Mangaia), the Reef Islands and Mussau. Although somewhat limited in number and geographical and temporal coverage, the sequences of these specimens demonstrated they were all Pacific Clade pigs, thus linking them with the Polynesian dispersal and, by association, with that of the earlier Lapita cultural complex – traditionally associated with the first peopling of remote Oceania (Kirch, 2000).

Analyses of a further six ancient mtDNA sequences extracted from archaeological Sus specimens from Liang Bua Cave (Flores, Lesser Sundas) demonstrated two separate human-mediated dispersals of domesticated S. scrofa (Pacific Clade pigs) and one involving an endemic Wallacean Sus species (S. celebensis). As S. celebensis is now endemic only to Sulawesi (assuming it was not naturally distributed beyond Sulawesi in the early Holocene), its presence at Liang Bua on Flores as early as 7000 BP (Van den Bergh et al., 2009) suggests its early Holocene translocation by humans,
perhaps as a domestic animal as previously suggested by Groves (1983). Another possible explanation is that *S. celebensis* was transported by hunter-gatherers to seed the island with game, just as the hunter-gatherers of the Levant introduced wild boar to Cyprus several thousand years before (Vigne *et al.*, 2009). Human-mediated translocation of other wild species in the region is evidenced by the introduction of the cuscus 20,000 ago into the Bismack archipelago (White, 2004).

Independent verification of the distinctiveness of pigs with the Pacific mtDNA signature has been shown by morphometric analysis of the lower third molar (M3) from the recent New Guinea and Flores pigs and archaeological pigs from the Flores site of Liang Bua using the same specimens analysed by Larson *et al.* (2007a). Molar shape analysis of the archaeological pigs from Holocene deposits at Liang Bua confirms the mtDNA evidence for the presence of two *Sus* taxa at the site: an endemic Wallacean type and several sharing the unique tooth morphotype of recent New Guinea pigs (Groves, 1983, 1995). In the light of the chron stratigraphic position of these remains (Van den Bergh *et al.*, 2009), these data show that two separate human introductions of *Sus* to the island of Flores occurred during the Holocene – an early one involving an endemic *Sus* species (*S. celebensis*) and a later one involving the arrival of domesticated *S. scrofa* most likely brought by migrating early farmers.

Groves’ claims for an interspecific hybrid (*S. s. vittatus* × *S. celebensis*) as the basis to explain the anthropogenic status and morphological divergence of New Guinea pigs (Groves, 1983, 1995) is not supported by either the morphometric or genetic data, all of which confirm the maternal affinity of the New Guinea pigs as a form of *S. scrofa*. Because mtDNA only reveals the maternal heritage, it remains possible that male Sulawesi warty pigs (*S. celebensis*) were involved in the creation of Pacific Clade pigs found on New Guinea. From a morphological perspective, ongoing research demonstrates a congruent molar shape divergence in modern specimens from Flores, New Guinea and Near Oceania, from their mainland South-east Asian relatives, which suggests that their molar shape variability has retained a common population signature. Random drift and directional adaptation linked to insular syndrome do not seem to have blurred the signal of their population history. The New Guinea *S. scrofa* morphotype is more likely the signal of its phylogenetic identity within the mtDNA Pacific clade, reflecting its domestic heritage.

The recent combination of genetic and morphometric approaches has significantly advanced our understanding of pig domestication in both South-east Asia and ISEA. These studies have conclusively demonstrated that peninsular South-east Asia was another centre for pig domestication, and that pigs domesticated here were then transported along an eastward trajectory into near and then eventually remote Oceania. This dispersal is linked with the spread of early farmers during the Neolithic period, and endemic ISEA *Sus* taxa, although hunted, appear to have played little or no part in this process, although further research will address this issue (Larson *et al.*, 2007b; Cucchi *et al.*, 2009).

**India**

The evidence for multiple pig domestication centres, particularly for the Near East, Europe, China and South-east Asia, is reasonably robust. However, there are several other centres of the Old World where evidence for pig domestication and/or introduction of domesticated pigs remains unresolved and where recent genetic data provide new evidence.

Although *Sus* remains are a widespread component of zooarchaeological assemblages throughout India and Pakistan, they represent such a minor proportion of the domestic and wild fauna recovered (Chattopadyaya, 2002) that the possibility of an authochthonous domestication of *S. scrofa* in India has rarely been considered. Recent genetic data have, however, suggested the possibility of an independent domestication of Indian wild boar. Modern mtDNA data (Larson *et al.*, 2005, 2010; Tanaka *et al.*, 2008) have demonstrated that domestic pigs in India and Bhutan possess unique and highly differentiated haplotypes identical to those found in Indian wild boar. This suggests that Indian domestic pigs were probably not derived either from migrating East Asian or South-east Asian populations or
that so-called wild boar samples from Morocco and Algeria possessed European haplotypes, while samples from Egypt and Sudan possessed Near Eastern haplotypes. These data could either indicate natural distributions of wild boar, extending from the Near East into north-eastern Africa, and from south-western Europe into north-western Africa, or the presence of feral populations of domesticated Near Eastern wild boar introduced into North Africa.

Additional unpublished sequences extracted from modern domestic pigs in Egypt (derived from a Coptic Christian butcher and claimed to be from traditional Egyptian varieties) have all produced common European domestic haplotypes. These sequences provide further evidence for a wholesale replacement of Near Eastern pig lineages by European domestic pigs that probably took place during later prehistory (Larson et al., 2007a). The application of additional DNA and geometric morphometric approaches to these questions will help to resolve these outstanding issues.

**North Africa**

Although the majority of the African continent is inhabited by numerous suid species (see Chapter 1), *Sus scrofa* populations have only been reported in North Africa along the Mediterranean. In Egypt, wild pigs were last reported in the late 19th century, and were present during the Christian era, reportedly from the Nile Delta, the Gizeh marshes the Fayum and Magrah oases, the Wadi el Natrun and the Wadi el Gharand (Manilus and Gautier, 1999). Domestic pigs clearly made up a portion of the ancient Neolithic (predynastic) Egyptian diet, and the northern delta has been claimed to be a centre of intensive pig breeding (Hecker, 1982). Whether these pigs were the descend-ants of a local domestication of an indigenous population of *Sus scrofa* remains unanswered, though the limited fossil and archaeological record for wild boar in Egypt implies that they were absent during prehistory (Manilus and Gautier, 1999), suggesting that domestic pigs were introduced, probably from the Near East.

More recent, though limited, genetic data (Larson et al., 2005, 2007a) demonstrated...
phylogenetic context is far simpler, but the resulting lack of paternal input into its inheritance pattern means that all mtDNA studies are limited to understanding the maternal evolutionary history. This is often not a problem, but because domestication in general, and pig domestication specifically, often involves a significant component of inter-population and even inter-species hybridization, as is evident in the European pig domestication and Island South-east Asian narrative above, studies that exclusively focus on mtDNA are unable to even ask questions related to the degree, timing and ramifications of that hybridization.

Two related approaches are changing that. First, newly emerging sequencing technologies are making it ever easier and cheaper to generate large numbers of both genetic sequences and individual single nucleotide polymorphisms (SNPs), as discussed in other chapters of this book. Secondly, comparative and functional genomic approaches using these data mean that many of the genes that are responsible for the phenotypic, behavioural and physiological differences evident between wild boar and domestic pigs are beginning to be found.

Though a number of other chapters in this book discuss the details of those genes, one specific locus is worth mentioning here. Following on from several publications that described the genetic variability in the pig melanocortin receptor 1 (MC1R) and correlated individual causative SNPs with specific coat colours, Fang et al. (2009) went one step further by typing more than 80 individuals from numerous locations across both Europe and China and, in doing so, identified 16 haplotypes. The most remarkable pattern that emerged from these data was not the fact that Asian wild boar harboured seven haplotypes while all European wild boar shared a single haplotype; because wild boar dispersed out of ISEA and into East Asia before continuing west towards Europe, they had spent a considerably longer period of time in East Asia, during which they had accumulated more genetic variation (Larson et al., 2005; Megens et al., 2008). What was unexpected was that, of all the five European and three Asian haplotypes found in domestic pigs, every one contained at least one non-synonymous mutation that altered the MC1R protein and, thus, the coat colour of the pigs.

These results first demonstrated that there is a strong selective pressure against non-camouflage coat colours in nature, and that humans must have exerted a significant positive selection for variation in order to generate large enough populations of all-black domestic pigs in order that the initial mutated sequences could act as the template on which additional mutations could be added. The fact that black domestic pigs exist in both Europe and China but achieve this coat colour through different mutations to the MC1R locus strongly suggests that the penchant for novelty is a universal trait, the consequences of which are evident in all our domestic animals. This kind of study goes a long way towards answering some of the basic questions that surround domestication, including the debate as to whether the process acts on standing variation found in wild populations or assists the spread and fixation of mutations that occur after domestication has begun (Fang et al., 2009).

What remains unknown, however, is when the domestic haplotypes began to appear and how quickly they spread through pig populations and across Europe and Asia. By typing the causative mutations in ancient material, it may soon be possible to reconstruct not just the coat colour, but also many other characteristics, including developmental and behavioural traits of archaeological pig specimens, thus adding the missing temporal context. This type of approach has already been carried out in horses, the results of which demonstrated that coat colour variation increased substantially 5000 years ago, at the same time that horse domestication becomes evident in the archaeological record (Ludwig et al., 2009).

Despite the progress that has been made over the past decade, a significant number of fundamental questions remain unanswered. The production of ‘improved breeds’ in the 19th century through the intentional mating of Asian and European pigs may have been just the most recent instance of hybridization (intentional or otherwise) between different species and subspecies of S. scrofa. As genetic studies move beyond the mitochondrial genome, assessing how much of an effect hybridization has had on global pig populations
becomes feasible. By increasing the power and resolution of genetic data sets, we will begin to further unravel 10,000 years of selection and interbreeding, the results of which will have a large impact on what we know, and how we can tackle issues associated with commercial pig production, the conservation of wild boar populations and even the appearance and spread of pathogens that infect both pigs and people. Thus, the combination of the new generation of genetic and morphological techniques has the power to inform the past, present and future of our long-standing relationship with pigs.

References


Molecular Genetics of Coat Colour Variation

Leif Andersson and Graham Plastow

1Uppsala University and Swedish University of Agricultural Sciences, Sweden; 2University of Alberta, Canada

Introduction

Genetic studies of coat colour variation in pigs were initiated soon after Mendel’s laws of inheritance were rediscovered in around 1900. Spillman (1906) first described the dominant inheritance of white colour, and this was followed by a study of the inheritance of the white belt phenotype in the following year (Spillman, 1907). Sewall Wright had already published the first review on the inheritance of coat colour variation in pigs in 1918 (Wright, 1918). This was followed by a number of studies that established the basic principles of pig coat colour genetics. H.O. Hetzer made a major contribution to this field when he published a series of papers during the 1940s and 1950s analysing the segregation of coat colour in a number of crossbreeding experiments (Hetzer, 1945a,b,c; reviewed by Legault, 1998). The classical genetics of coat colour variation in pigs has been reviewed by several authors (Ollivier and Sellier, 1982; Legault, 1998).

Before the era of molecular genetics, it was proposed that a set of major coat colour loci designated A, B, C, D, E, P and S (Table 3.1) controlled coat colour variation in mammals. Castle and Little (1909) were the first to introduce the concept of a set of major genetic factors affecting coat colour, and this was further developed by Sewall Wright in a series of seminal papers (e.g. Wright 1917, 1918). Subsequent molecular studies have to a large extent confirmed previous assumptions of gene homologies between species based on strikingly similar effects on the coat colour phenotype (Table 3.1). However, further genetic studies have revealed that mutations in a fairly large number of genes may affect coat colour. In the mouse, for example, 159 genes affecting pigmentation have been described so far (Montoliu et al., 2009).

The first molecular study on pig coat colour genetics was published in 1992, when close linkage between the gene for Dominant white colour and the ALB and PDGFRA genes on pig chromosome 8 was reported (Johansson et al., 1992). Since then, more than 20 molecular studies on pig coat colour genetics have been published and three major coat colour...
loci in the pig have been studied at the molecular level: the Dominant white (I) locus, the Extension (E) locus and the Agouti (A) locus. These studies have shown that most of the coat colour diversity present among major pig breeds, at least those of European origin, is explained by genetic variation at the Dominant white and Extension loci. The major aim of this chapter is therefore to review these molecular studies and their implications for our understanding of coat colour variation in pigs.

**The Dominant White (I) Locus – Extensive Allelic Diversity**

Dominant white is one of the most widespread coat colour variants in European domestic pigs, and major breeds such as Yorkshire/Large White and Landrace are assumed to be fixed for the Dominant white (I) allele, although this is not always the case (see below). Some pig meat products are marketed including the skin, and in many markets there is a strong consumer preference for products with white skin. This, at least in part, relates to the perception that the dark hair roots that are present after the dehairing of coloured breeds indicate a hygiene (mould) problem. Therefore, the common practice in pig breeding is often to use a dam that is homozygous for the Dominant white allele to ensure that the production animals have white skin. In skin-on markets, such animals require less processing post slaughter, representing significant cost savings, especially from yield loss (Cannon et al., 1996).

Wentworth and Lush (1923) proposed that Dominant white colour is caused by a single dominant gene, and this was later confirmed by Hetzer in his analysis of crosses between coloured and white breeds (Hetzer, 1945a,b,c). Hetzer also proposed the allele designation I for inhibition of colour. The molecular characterization of the Dominant white locus was initiated by the assignment of this locus to the centromeric region of chromosome 8 based on segregation analysis of an intercross between the European wild boar and Large White pigs (Johansson et al., 1992). The segregation analysis of the white colour in this cross was in accordance with expectation, i.e. a 3:1 segregation between white and coloured progeny in the F2 generation, but with one exception (Plate 1A); one of the Large White

---

**Table 3.1.** The classical coat colour loci in mammals and their corresponding genes as identified in the mouse.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>Effect on pigmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Agouti</td>
<td>Asip</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Brown</td>
<td>Tyrp1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Albino</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Dilution</td>
<td>Myo5a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Extension</td>
<td>Mc1r</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Pink-eyed dilution</td>
<td>Oca2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>Spotting</td>
<td>Ednrb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*a In mouse nomenclature, only the first letter in a locus symbol is capitalized, whereas in humans and pigs locus symbols are written with all capitals.

*b Mutations in other genes may give very similar dilution of coat colour.

*c Mutations in other genes give similar white spotting phenotypes. In the pig, the KIT gene is the major locus for white colour and white spotting.
founder animals carried an alternative allele, which was named Patch (\(l^p\)), and \(F_1\) animals that were heterozygous (\(l^p/i\)) showed a white spotted phenotype in contrast to the pure white phenotype observed for the \(l/i\) heterozygotes. The segregation data in the \(F_2\) generation confirmed that the Patch phenotype was caused by an alternative allele at the Dominant white locus (Johansson et al., 1992).

The assignment of the Dominant white locus to the centromeric region of chromosome 8 immediately revealed the KIT gene as a strong positional candidate locus (Johansson et al., 1992). A few years earlier the murine homologue, Kit, had been identified as the gene underlying Dominant white spotting (Chabot et al., 1988; Geissler et al., 1988). Kit encodes a tyrosine kinase receptor that has a crucial role for the migration and survival of melanocyte precursor cells originating from the neural crest, as well as for haematopoiesis and germ cell development. The Kit ligand (KITL) has previously been denoted Mast cell growth factor (MGF), and mutations in the gene encoding KITL in mice also cause pleiotropic effects on pigmentation, haematopoiesis and fertility (Rothschild et al., 1992). A few years earlier the murine Kit gene has recently been characterized, but no significant role in determining coat colour variation has yet been revealed (Okumura et al., 2006; Hadjiconstantouras et al., 2008).

Molecular studies revealed that both the Dominant white (\(l\)) and the Patch (\(l^p\)) alleles are associated with a duplication of the KIT locus (Johansson Moller et al., 1996). A small deletion was detected in intron 18 and segregation analysis showed that both the deleted form and the normal form were transmitted together with the \(l\) and \(l^p\) alleles. In order to determine the size of the duplication, a BAC contig (a set of overlapping bacterial artificial chromosome clones) was established across the Dominant white/KIT locus, and this revealed a 450 kb duplication that encompasses all coding exons, about 150 kb of the upstream region and about 200 kb of the downstream region (Giuffra et al., 2002). The sequences in the vicinity of the duplication break-points indicated that the duplication had occurred as a result of a recombination event between LINE sequences flanking the KIT locus. Based on comparisons with the characterization of the functional consequences of similar rearrangements associated with Kit mutations in mice, it was proposed that the duplication acts as a regulatory mutation (Giuffra et al., 2002). The duplicated KIT copy may show dysregulated expression because it is physically separated from some of the regulatory sequences located at a far distance from the coding sequence.

The observation that the Dominant white and Patch alleles carried the same duplication implied that at least one other causative mutation must exist in order to explain the phenotypic difference associated with these alleles. An examination of KIT transcripts in white blood cells revealed the presence of two splice forms in pigs carrying the Dominant white allele, one normal and one variant form lacking exon 17 (Marklund et al., 1998). Sequencing of exon 17 and the flanking introns revealed a single base substitution in the first nucleotide of intron 17. It was concluded that this must be the mutation leading to the skipping of exon 17 because the splice junctions are essential for a normal splice pattern (Marklund et al., 1998). Further analysis revealed that the Dominant white allele carries one copy of the KIT gene with a normal intron 17 sequence, while the second KIT copy has a mutant form of intron 17 consistent with the transcript pattern; in contrast, in Patch both KIT copies have the normal form of intron 17. Exon 17 codes for a part of the tyrosine kinase domain that is essential for KIT signalling. The data therefore indicated that the mutant KIT sequence, the one with the splice mutation, encodes a receptor with normal ligand binding but with no or drastically reduced signalling capacity. Thus, the combined effect of the duplication (most likely a regulatory mutation, as discussed above) and the splice mutation explains the unique features of the Dominant white allele. It has a more drastic effect on pigmentation in the heterozygous condition than any mouse Kit mutation, but homozygotes are still fully viable in contrast to many mouse Kit alleles that are lethal or semi-lethal in the homozygous condition. The reason why the Dominant white allele is fully viable despite a major effect on pigmentation is that the presence of at least one functional KIT copy per chromosome provides sufficient KIT signalling to avoid the severe pleiotropic effects on haematopoiesis and fertility observed in
animals lacking KIT function completely. However, the Dominant white allele is associated with a mild reduction in the number of red and white blood cells (Marklund et al., 1998; Johansson et al., 2005). Mild negative effects on fertility have not been documented but it cannot be excluded that such effects exist.

The presence of two causative mutations at the KIT locus suggested an evolutionary scenario where the duplication occurred first and resulted in a partially white coat colour. The occurrence of the splice mutation then enhanced the white phenotype. This is the most plausible explanation because comparative mouse data show that loss-of-function mutations at the Kit locus give a mild coat colour phenotype in the heterozygous condition but they are lethal in the homozygous condition, whereas regulatory mutations can be both fully dominant and fully viable, even in homozygotes.

The relatively complex structure of the KIT locus, including the duplication and the splice mutation, and the interest in accurately scoring the Dominant white genotype to ensure homozygosity in animals used for breeding called for a very accurate genotyping method. Pyrosequencing was employed to establish such a method (Pielberg et al., 2002, 2003). An alternative method based on the oligonucleotide ligation assay was subsequently described (Seo et al., 2004). These studies revealed extensive haplotype diversity at the KIT locus among commercial pig populations assumed to be homozygous for the Dominant white allele. Haplotypes with one, two or three copies were found, and the splice mutation occurred in none, one or two of the copies (Table 3.2). Furthermore, an allele carrying a single copy but with the splice mutation was found, and denoted I', because it is expected to be lethal in the homozygous condition owing to a severe defect in KIT signalling (Pielberg et al., 2003). The extensive genetic heterogeneity at the KIT locus in pigs strongly suggested that it is generated by unequal crossing over during meioses, but the rate of unequal crossing over at this locus has not yet been established. This explains why many white lines still segregate for Dominant white despite breeders having attempted for more than 200 years to breed the white allele to homozygosity. The results also imply that white pigs that become feral will return to wild type at the KIT locus because single-copy alleles lacking the splice mutation will be generated by unequal crossing over and then favoured by natural selection as pigmentation provides camouflage and protection against the damaging effects of UV light.

The Belt allele was assigned to the KIT locus on the basis of classical segregation analysis using an intercross between Hampshire (belted) and Piétrain (non-belted) pigs (Giuffra et al., 1999). The entire coding sequence of KIT associated with Belt had been determined in a previous study and did not reveal any putative causative mutation (Marklund et al., 1998). These results strongly suggest that the Belt phenotype is caused by a regulatory mutation, and it is well established that some of the regulatory elements controlling KIT expression during development are located more than 100 kb away from the coding sequence (Berrozpe et al., 1999).

Table 3.2. Summary of the characteristic features of alleles at the Dominant white (I)/KIT locus in pigs.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Designation</th>
<th>No. of KIT copies</th>
<th>No. of normal copies</th>
<th>No. of copies with splice mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>i</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Belt</td>
<td>I&lt;sup&gt;lo&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Roana</td>
<td>I&lt;sup&gt;n&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patch</td>
<td>I&lt;sup&gt;e&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Dominant white</td>
<td>1'</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dominant white</td>
<td>1'</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Dominant white</td>
<td>L</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lethal</td>
<td>L</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*The existence of this allele needs to be confirmed once the causative mutation(s) for the Belt allele has been identified.*
A fourth KIT allele, in addition to wild-type, Patch and Dominant white, was detected in the intercross between European wild boar and Large White pigs (Pielberg et al., 2002). Like the Belt allele, this allele was associated with a single KIT copy and no splice mutation, but it showed a roan phenotype (white and coloured hairs intermingled) rather than a belted phenotype. This represents a tentative additional KIT allele denoted Roan (R<sup>on</sup>) but at present it cannot be formally excluded that Belt and Roan represent the same allele showing different phenotypic expressions on different genetic backgrounds. This can be resolved easily once the causative mutation for the Belt phenotype has been identified. The alleles described so far at the Dominant white locus are summarized in Table 3.2.

The Extension (E) Locus – Contrasting Selection Pressures in Wild and Domestic Pigs

Classical genetics established an allelic series comprising four alleles at the Extension locus in pigs (Legault, 1998). These are, with this order of dominance, Dominant black (E<sup>B</sup>), wild-type (E<sup>E</sup>), black spotting (E<sup>p</sup>) and recessive red (e). In the F<sub>2</sub> generation of a cross between European wild boars (E<sup>E</sup>/E<sup>E</sup>) and Large White pigs segregating both at the Dominant white and Extension loci, wild type colour (E<sup>E</sup>/E<sup>E</sup>) and black spotting (E<sup>p</sup>/E<sup>p</sup>) segregated in accordance with the expected 3:1 ratio among the progeny that were homozygous i/i at the Dominant white locus, and E was assigned to pig chromosome 6 by linkage analysis (Mariani et al., 1996). The MC1R gene encoding the melanocortin 1 receptor was identified as the obvious candidate gene because an earlier study had shown that the Extension locus corresponds to the Mc1r gene in mice (Robbins et al., 1993; Table 3.1). MC1R signalling determines the switch between production of black/brown eumelanin and the production of red/yellow pheomelanin. MC1R signalling is activated by the binding of the ligand melanocyte-stimulating hormone (MSH) to the receptor leading to production of black pigment, while the agouti signalling protein (ASIP) encoded by the Agouti locus inhibits MC1R signalling and thereby favours the production of red pigment. A typical wild-type coat colour in mammals, like the one in the wild boar, is composed of a mixture of eumelanin and pheomelanin, and individual hairs may be striped black and yellow owing to the temporal activation of ASIP production during the hair cycle. Mutations causing constitutive activation of MC1R signalling cause dominant black colour, whereas loss-of-function mutations are associated with recessive red colour.

In an initial analysis, the entire MC1R coding sequence (except the first 120 nucleotides) was determined in pigs representing different Extension alleles (Kijas et al., 1998). This analysis immediately confirmed that the Extension locus also harbours the MC1R gene in pigs because a number of missense mutations showing complete association to the different Extension alleles were revealed. Two different sequences were associated with Dominant black colour, and it was apparent that one, denoted E<sup>Ed1</sup>, had an Asian origin, while the other, denoted E<sup>Ed2</sup>, clearly had a European origin. The MC1R sequence associated with the latter allele was identical to the one found in the European wild boar except for a single non-conservative missense mutation, Asp124Asn (Table 3.3), which is almost certainly the causative mutation (Kijas et al., 1998). The Asian Dominant black allele was associated with two missense mutations, Val95Met and Leu102Pro (Table 3.3), and it is assumed that the latter is the prime causative mutation because the same missense mutation in the corresponding position causes dominant black colour in cattle (Klungland et al., 1995). The e allele causing recessive red colour, for instance in Duroc pigs, is also associated with two missense mutations, Ala164Val and Ala243Thr (Table 3.3), and it has not yet been established if one of these or the combined effect of the two is the cause of red colour.

A surprising finding in this initial study was that the MC1R sequence of the E<sup>p</sup> allele for black spotting was identical to the one associated with the Dominant black allele (E<sup>Ed2</sup>) of European origin (Kijas et al., 1998). However, a subsequent study in which the entire coding sequence of MC1R was
Table 3.3. Sequence alignment of pig MC1R/E alleles (— indicates identity with the master sequence).

<table>
<thead>
<tr>
<th>Codon</th>
<th>4</th>
<th>17</th>
<th>21</th>
<th>22</th>
<th>95</th>
<th>102</th>
<th>117</th>
<th>121</th>
<th>122</th>
<th>124</th>
<th>164</th>
<th>166</th>
<th>243</th>
<th>301</th>
</tr>
</thead>
<tbody>
<tr>
<td>0102/ Eα</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0103/ Eα</td>
<td>C</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0104/ Eα</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0105/ Eα</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0201/ Ed</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0202/ Ed</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0203/ Ed</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0301/ Ed</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0401/ e</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0501/ Eδ</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0502/ Eδ</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0503/ Eδ</td>
<td>A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

*aThe nomenclature for MC1R alleles is based on the system proposed by Fang *et al.* (2009).

bFS = frameshift, sequence out of frame after codon 22.
determined revealed that the $E^p$ allele also carries an insertion of two C nucleotides at codon 22, in addition to the missense mutation Asp124Asn (Kijas et al., 2001; Table 3.3). The two-base-pair insertion constitutes a frameshift mutation and is thus expected to be a complete loss of function that should cause recessive red colour according to the current knowledge of $MC1R$ function. In fact, some of the pigs that are homozygous for this allele show a uniform red colour, although the great majority shows a black spotting phenotype (Plate 1A); how can this be explained? The fact that the CC dinucleotide insertion occurs in a stretch of six C nucleotides, creating a short eight-base-pair mononucleotide repeat, implied that this mutation may be somatically unstable during melanocyte development. In order to test this hypothesis, reversed transcriptase (RT)-PCR analysis was performed from skin with black spots as well as skin with red spots; the results showed that somatic reversions had occurred in areas with black pigmentation so that in all cases the 8C had shrunk to 6C, which restores the reading frame, and the Dominant black mutation at residue 124 becomes reactivated!

The $E^p/E^p$ genotype has a very variable phenotypic expression and may be associated with uniform red colour, black spotting on a red background colour or black spotting on a white background. One of the most extreme forms is found in Berkshire pigs, which show a uniform black colour but with white feet, nose and tail. Sewall Wright (1918) had already proposed that the black colour of Berkshire pigs constitutes an extreme form of black spotting due to the action of modifying genes. The presence and extension of black spots in $E^p/E^p$ pigs can be affected both by factors/mutations that enhance the frequency of somatic mutations and by factors/mutations that stimulate melanocyte migration subsequent to the occurrence of a somatic mutation. A very intriguing observation is that the black spots are consistently larger on a white background than on a red background colour (Plate 1A). A possible explanation for this could be that the red areas contain melanocytes producing red pheomelanin, while the white areas lack melanocytes completely, which allows the revertant melanocytes to expand more and form larger black spots. However, no functional data supporting this suggestion are yet available and it is still not known whether the white areas lack melanocytes or contain melanocytes that are unable to produce pigment.

It can be concluded that pigs with white background colour and black spots have a genotype that does not support pigment production in the absence of $MC1R$ signalling. The genetic difference between pigs with black spotting on a red or white background colour has not yet been resolved. These two phenotypes segregated in an intercross between European wild boars and Large White domestic pigs (Plate 1A) and a careful examination of the segregation data showed that this difference is not controlled by the $MC1R$ locus as full siblings sharing the same $E^p$ alleles may differ in background colour (Mariani et al., 1996). The ratio of the two phenotypes did not fit any simple Mendelian inheritance ratio and no linkage was detected. This implies either that the phenotypes have a more complex genetic background or that one or both of the parental strains are not fixed at a putative locus underlying this trait.

A significant spin-off finding of the characterization of the $MC1R$ locus was that it clearly indicated a substantial genetic distance between Asian and European pigs because the Asian $E^{D1}$ allele differed from the $E^+$ allele and all other $E$ alleles of European origin by two synonymous substitutions, in addition to the two non-synonymous substitutions. The implication was that Asian domestic pigs originate from Asian wild boars whereas European domestic pigs originate from European wild boars. Furthermore, the presence of the Asian allele for dominant black colour in some European breeds most likely reflects the documented introgression of Asian pigs into European stocks primarily during the 18th and 19th centuries (Kijas et al., 1998). This was subsequently confirmed in follow-up studies including the analysis of mtDNA (mitochondrial) and nuclear genes, which indicated a considerable genetic distance between Asian and European wild boars (Giuffra et al., 2000; Kijas and Andersson, 2001). This finding has been further corroborated in many subsequent studies (see Chapter 2).
A comprehensive screening of genetic diversity at the MC1R locus was accomplished by sequencing the entire coding sequence from pigs representing 31 populations of European pigs, 15 breeds of Chinese pigs, and European and Asian wild boars (Fang et al., 2009). This resulted in several interesting findings. First, the $E^+$ wild-type allele was only present in one of the European breeds, the Mangalitsa. Interestingly, this is one of the very few breeds in which the piglets are striped, which is a characteristic feature of wild boar piglets. This implies that MC1R is a domestication locus in pigs as almost all breeds are fixed for a mutant form of MC1R. Furthermore, it indicates that an MC1R wild-type sequence is required to produce the striping pattern in piglets. The authors have also seen dorsoventral striping in other crosses, suggesting that there are further alleles segregating in some breeds.

Secondly, a total of seven nucleotide substitutions was found among European and Asian wild boars, and all were synonymous, i.e. they did not change the amino acid sequence (Plate 1B). In sharp contrast, nine out of ten mutations only found among domestic pigs affected the coding sequence, and some alleles differed by up to three non-synonymous substitutions from their closest wild-type allele (Plate 1B). The ratio of the relative frequency of non-synonymous ($d_N$) and synonymous ($d_S$) substitutions was as high as 23.5 among domestic pigs, whereas the $d_N/d_S$ ratio was 0 among the wild boars. This provided evidence for purifying selection in the wild, most likely to maintain camouflage colour, but strong directional selection against camouflage at the farm. Possible explanations for this selection in domestic pigs are that: (i) a non-camouflaged coat colour may have facilitated early animal husbandry; (ii) coat colour variants have been used as markers to easily distinguish domesticated forms from wild forms; or (iii) simply because the phenotypic diversity created by coat colour variants has been appealing to humans as we can appreciate fashion.

This study of MC1R diversity among pig populations has important implications. It demonstrates how strong the selection pressure can be in domestic animals and that the history of domestication is sufficiently long for some alleles to represent the accumulation of several consecutive mutations with phenotypic effects. The $E^+$ allele is an example of an allele carrying two causative mutations, the missense mutation (Asp124Asn) and the two-base-pair insertion at codon 22. Domestication is very recent from an evolutionary perspective – within the last 10,000 years – which means that there has not been sufficient time to accumulate many synonymous substitutions. This is well illustrated by the fact that not a single synonymous substitution in the MC1R gene has yet been verified among European wild boars and European domestic pigs. This finding is consistent with other data indicating that the European wild boar went through a population bottle-neck before domestication (Fang and Andersson, 2006; Fang et al., 2009).

This MC1R study provided conclusive evidence that intentional human selection has acted directly to change the coat colour. The very strong selection signature refutes the possibility that the coat colour has changed simply as a result of relaxed purifying selection subsequent to domestication. Furthermore, MC1R has no known function outside the pigment cell, making it exceedingly unlikely that the coat colour has been altered as a by-product of selection for another trait, for instance behaviour, as suggested in some other studies (see Fang et al., 2009).

The Agouti (A) Locus – Still Poorly Characterized at the Molecular Level

Agouti is one of the classical coat colour loci in mammals and it determines, together with the Extension locus, the relative proportion of red/yellow pheomelanin and brown/black eumelanin in pigmentation (Table 3.1). The correct gene designation is now ASIP, as this gene encodes agouti signalling protein (ASIP), which is an MC1R signalling protein that interferes with the interaction between MSH and MC1R. Therefore, loss-of-function alleles at the ASIP locus cause recessive black colour, because MSH is free to bind MC1R throughout the hair cycle, whereas ASIP overexpression leads to dominant yellow colour, because ASIP
blocks MC1R. The ASIP locus in pigs has not been characterized to the same degree as the KIT and MC1R loci, but Drögemüller et al. (2006) have reported that the black-and-tan phenotype segregating in a Mangalitsa × Piétrain intercross is determined by an allele at the ASIP locus, and proposed that the allele is designated \( a^t \). No sequence difference in the coding region was detected between the \( a^t \) and the \( A^t \) wild-type alleles, but five ASIP transcripts showed differential expression between genotypes at this locus in dorsoventral skin, strongly implicating that \( a^t \) is caused by a regulatory mutation. It is well established that the expression of ASIP has a complex regulation and is affected by multiple regulatory elements spread over a fairly large genomic region (Vrielings et al., 1994).

**Other Loci**

As well as the three loci described in detail in the present review, Legault (1998) reviews data suggesting the existence of genetic variation at five additional loci that may affect coat colour variation in pigs: the Brown (B) locus, the Albino (C) locus, the Dilution (D) locus, the Red-eye (R) locus and the White head (He) locus. The three former loci are assumed to correspond to the classical B, C and D loci in other mammals (Table 3.1). In fact, a recent study demonstrated that brown colour in Chinese Tibetan, Kele and Dahe pigs is caused by a six-base-pair deletion in TYRP1, which encodes tyrosinase-related protein 1 (Lusheng Huang, personal communication). No conclusive evidence for genetic variation at the Albino (C) locus has yet been reported, but Searle (1968) proposed that dirty white colour in Mangalitsa pigs maybe due to an allele at this locus. Similarly, no conclusive evidence exists for genetic variation at the D locus in pigs (Legault, 1998), which corresponds to the MYO5A gene, which encodes for Myosin type Va (Table 3.1). Roberts and Krider (1949) identified an autosomal recessive allele (\( r \)) causing red eye pigmentation and dilution of coat colour in Hampshire pigs. Fernández et al. (2003, 2005) investigated genetic variation in the OCA2 gene, which encodes for oculocutaneous albinism II and corresponds to the classical pink-eyed dilution (\( P \)) locus (Table 3.1); they found that this locus had an additive polygenic effect on colour intensity in red Iberian pigs. Finally, Legault (1998) reviewed the evidence for a White head (\( He \)) allele showing dominant inheritance. An obvious candidate gene for this phenotype is the KIT locus, and it cannot be excluded that the phenotype reflects the expression of one of the many \( KIT \) alleles (see Table 3.2) on a certain genetic background.

In addition, linkage studies such as that of Hirooka et al. (2002) point to a number of ‘modifying loci’ that affect coat colour. However, these loci need to be confirmed by more detailed molecular studies. For example, a candidate for one of these modifying loci is the KIT ligand (see above in relation to Dominant white/KIT), although subsequent analysis of this locus failed to find any association with variation in coat colour (Hadjiconstantouras et al., 2008). This need is also well illustrated by the fact that the Belt phenotype, for instance, that is exhibited by Hampshire pigs was assumed to represent an allele at a separate locus until it was shown that it is controlled by an allele at the Dominant white locus (Giuffra et al., 1999). Similarly, Bushnell (1943) reported loose genetic linkage between the Belt and Extension loci, but this can now be refuted because subsequent molecular studies have shown that the Dominant white (Belt) and Extension loci are located on different chromosomes – chromosomes 8 and 6, respectively. Thus, access to molecular data provides an excellent opportunity to test hypotheses of inheritance patterns of coat colour variation.

**Coat Colour Variants as Breed-specific Genetic Markers**

The identification of the two major loci for coat colour, Dominant white and Extension, and the characterization of variation in these genes across breeds has provided a useful tool for the verification of breeds and the products derived from these breeds (e.g. Alderson and Plastow, 2004; also reviewed more generally by Dalvit et al.,...
Molecular Genetics of Coat Colour Variation 47

2007). This utility is of course based on the selection of physical uniformity of type (i.e. different breed-specific morphological characters such as ear type, head shape and coat colour) as part of setting specific breed standards during the development of a breed (Alderson and Plastow, 2004). Dalvit et al. (2007) indicated that the use of colour loci is ‘deterministic’ and therefore more straightforward than the use of ‘probabilistic approaches’. This has proven to be very effective in certain situations where the question of origin can be framed to match the information available for a breed, the likely source of non-conforming product, and the probability of finding rare but uncharacterized breeds that negate the conclusion. The need for such a tool increases as products become specified to specific regions or types of production that include a genetic element. Jambon Iberico in Spain and Kuro Buta in Japan are examples of pork products that specify breed type, and Dalvit and colleagues (2007) state that ‘breed traceability is a means to defend and valorize particular food products’. A number of different studies have investigated this application and the ‘keys’ derived from them continue to be applied in several different situations. For example, Carrion et al. (2003) discuss examples for wild boar and Iberian pigs as well as Kuro Buta (Berkshire) in Japan, while Alderson and Plastow (2004) give examples for Tamworth as well as Berkshire. Other examples include replacing test mating to remove ‘red alleles’ introduced into the Hampshire breed in the USA and the authentication of Iberian pork products using MC1R and other markers such as a synonymous substitution at the OCA2 locus (Fernández et al., 2004).

Some Remaining Questions to Resolve

The coat colour variation present among European pig breeds is primarily controlled by alleles at the Dominant white/KIT and Extension/MC1R loci. The causative mutations for the Belt and Roan alleles at the Dominant white/KIT locus and for the a’ allele at the ASIP locus remain to be identified. It is expected that this can be accomplished in the not too distant future because next-generation sequencing now makes it possible to very efficiently and cheaply scan the genome for sequence variants associated with phenotypic traits (Rubin et al., 2010). Another fascinating question to resolve is the genetic basis for the difference in background colour (white or red) in black-spotted pigs, as discussed above. It is also possible that additional coat colour loci contribute to coat colour variation within or between European pig breeds, as suggested by classical segregation analysis (Legault, 1998), for example pigs derived from a cross between Berkshire and Hampshire breeds show remarkable colour patterns that would not be expected from simple models, but these need to be verified by further molecular work.

It is also interesting that there has been independent selection for black colour and for different types of white-spotting phenotypes in European and Asian pigs. It is clear that most Asian – or at least Chinese – pig breeds are fixed for the Asian-specific Dominant black allele at the Extension/MC1R locus (Fang et al., 2009), but the genetic basis for white spotting and white colour in Asian breeds remains to be resolved. Evidence exists for a non-dominant white coat colour in Asian breeds, and the authors have seen evidence for similar loci in European pigs. It has been reported that neither a belt-like phenotype nor the white colour in Rongchang pigs is associated with the KIT duplication or the KIT splice mutation causing white colour in European pigs (Xu et al., 2006; Lai et al., 2007). It is possible that other Asian-specific KIT mutations underlie white colour, but other major candidate genes for these phenotypes are the microphthalmia-associated transcription factor (MITF) and endothelin receptor beta (EDNRB) genes. Interestingly, a recent screen for genetic polymorphisms in 57 porcine homologues of genes associated with coat colour variation in mice revealed that Chinese Jinhua pigs with a characteristic extensive white belt (two-end black) are homozygous for an EDNRB missense mutation that is rare or absent in other breeds (Okumura et al., 2006, 2010). It is anticipated that the genetic basis for white colour and white-spotting phenotypes present among Asian pig breeds will also be resolved in the not too distant future.
References


Hetzer, H.O. (1945c) Inheritance of coat color in swine. IV. Analysis of hybrids of Landrace and Large Black. 
*Journal of Heredity* 36, 309–312.

Genome scan reveals new coat color loci in exotic pig cross. *Journal of Heredity* 93, 1–8.


Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* 7, 822–830.


Okumura, N., Hayashi, T., Uenishi, H., Fukudome, N., Komatsuda, A., Suzuki, A., Shibata, M., Nii, M., 


Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., 
Sharpe, T., Ka, S., Hällböök, F., Besnier, F., Carlborg, Ö., Bed’hom, B., Tixier-Boichard, M., Jensen, P., 


Wright, S. (1917) Color inheritance in mammals: results of experimental breeding can be linked up with chemical researches on pigments – coat colors of all mammals classified as due to variations in action of two enzymes. \textit{Journal of Heredity} 8, 224–235.


Introduction

Morphological traits have not been of much importance in pigs. In contrast, inherited disorders have been of considerable importance, both from an agricultural point of view (e.g. malignant hyperthermia) and in terms of animal models for inherited human diseases. This chapter, therefore, concentrates on inherited disorders. However, it does also include mention of morphological traits that have been documented.

With the molecular revolution now bearing fruit, and, in particular, with the development of DNA markers covering all regions of all pig chromosomes (see Chapters 5, 7 and 8), knowledge of the genetic basis of morphological traits and inherited disorders in pigs will increase rapidly in the decades ahead. As described below, regularly updated information is available on the Internet. By this means, it will be possible for readers throughout the world to obtain the latest information on any trait or disorder mentioned in this review.

The Range of Possibilities

The spectrum of morphological traits and inherited disorders ranges from those that are definitely due to the action of just one gene to those that are due to the combined action of many genes and many non-genetic (environmental) factors. In between these two extremes are many traits and disorders which appear to run in families, but for which there is insufficient information to enable a conclusion to be drawn about whether one or more genes are involved. Unfortunately, the literature abounds with examples of traits and disorders that have been claimed to be due to just one gene, despite the data being so sparse that such a claim cannot be justified. Similar problems exist with claims of inheritance being recessive or dominant; in most cases, there is insufficient information to justify the claims that have been made. In the fullness of time, of course, additional data might support the initial claims. But we must be careful not to jump the gun.

During the 1980s and 1990s, there were considerable advances made in identifying the
molecular basis of inherited disorders in domestic animal species, and the pig was at the forefront of such studies. Indeed, the discovery of the molecular basis of malignant hyperthermia by Fujii et al. (1991) provides a marvellous, pioneering example of the power of molecular genetics to solve important, practical and economically important problems (see Chapter 15).

**Previous Reviews**

Several comprehensive reviews of inherited traits and disorders in pigs have been published over the years. The first major summary was by Smith et al. (1936). Since then, there have been comprehensive surveys of inherited disorders by Johansson (1964), Ollivier and Sellier (1982), Hanset (1991) and Robinson (1991), together with more wide-ranging reviews by Koch et al. (1957), Wiesner and Willer (1974), Huston et al. (1978) (congenital disorders) and Hamori (1983). More recent reviews with a focus on specific types of traits include those of Kocwin-Podsiadla (1998), Fukawa and Kusuhara (2001) and Misdorp (2003). It should be noted that some of these reviews are concerned with congenital traits and disorders, i.e. traits and disorders that are present at birth. Not all such traits and disorders are inherited.

In compiling the list of single-locus traits and disorders presented at the end of this chapter (Table 4.1), the information provided by Ollivier and Sellier (1982) and Robinson (1991) has been used as a convenient starting point. It is interesting to compare the conclusions drawn in these two reviews. Ollivier and Sellier discussed 76 traits and disorders, and concluded that 29 are determined by a single locus. Nearly a decade later, Robinson listed 64 traits and disorders, of which 33 were regarded as being due to a single locus. Of the 64 in Robinson’s list, only five were first reported after the publication of Ollivier and Sellier’s review. Thus, despite being more stringent, Robinson actually declared one more pre-1982 traits/disorders to be due to a single locus than did Ollivier and Sellier. To what extent do the two single-locus lists coincide? Somewhat less than one might have thought. In fact, only 20 traits/disorders are common to the two lists: nine of Ollivier and Sellier’s single-locus traits/disorders were excluded from Robinson’s list; and ten of Robinson’s single-locus traits/disorders were excluded by Ollivier and Sellier.

Of course, these statistics are simply a reflection of the difficulties involved in evaluating inadequate evidence concerning inheritance, as described above. Both lists were compiled with great care by very competent geneticists. The fact that their conclusions differ is a cautionary tale for anyone charged with the task of determining whether a particular trait/disorder is due to a single gene.

**Current Sources of Information**

While a list of reviews is useful, it is even more useful to have a single catalogue of morphological traits and inherited disorders that is regularly updated, and which is made available freely on the Internet. Human geneticists have long had access to such a resource, McKusick’s *Online Mendelian Inheritance in Man* (OMIM), available at http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim. This catalogue contains a wealth of information on thousands of morphological traits and inherited disorders in humans. It also contains a surprising quantity of information on pigs, because McKusick has always been interested in potential animal models of human disorders.

In 1978, the present author commenced compiling a catalogue of inherited traits and disorders in a wide range of animal species. Being modelled on, and complementary to, McKusick’s catalogue, this catalogue is called *Online Mendelian Inheritance in Animals* (OMIA) and is available at http://omia.angis.
OMIA includes entries for all inherited disorders in pigs, together with other traits in pigs for which single-locus inheritance has been claimed, however dubiously. Each entry consists of a list of references arranged chronologically, so as to present a convenient history of knowledge about each disorder or trait. For some entries, there is additional information on inheritance or molecular genetics. If the disorder or trait has a human homologue, the relevant OMIM numbers are included, providing a direct hyperlink to the relevant entry in McKusick’s online catalogue OMIM.

**An Overview of Single-locus Traits and Disorders**

At the time of writing, OMIA contains entries for 222 pig traits or disorders, collectively called phenes. Several of them concern coat colour, which is discussed in Chapter 3 of this book, and these are, therefore, not covered in the present chapter. Of the 222 phenes, 72 are potential animal models of human disorders (see Chapter 17 for detailed discussion of some of these phenes). For conciseness, the only phenes listed in the present chapter (Table 4.1) are the 55 non-coat-colour phenes for which there is strong evidence of single-locus inheritance. For each phene listed, the table includes the earliest report plus a recent reference, extracted from OMIA. If the relevant gene has been identified, it is listed, together with its genome location, where known. More comprehensive information, including a complete set of up-to-date references for each entry, can be obtained by accessing OMIA at http://omia.angis.org.au/.

**Conclusions**

The list of inherited morphological traits and disorders presented in this chapter provides an indication of the range of such traits and disorders that have been observed and studied in pigs. The molecular and gene-mapping revolutions now under way will lead to an explosion of knowledge in this area in the years ahead. To exploit fully the genetic variation that does occur, breeders and researchers need to be continually on the lookout for unusual animals, saving them where possible. If DNA can be sampled from several generations of a family in which a particular morphological trait or disorder occurs, and if careful records on the occurrence of the trait or disorder in that family have been kept, it will be an increasingly straightforward matter to identify the gene responsible.
Table 4.1. List of single-locus traits (other than coat-colour) and disorders documented in pigs. Further information on each entry can be obtained from OMIA (Online Mendelian Inheritance in Animals; http://omia.angis.org.au/).

<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aplasia of tongue</td>
<td>000055</td>
<td></td>
<td></td>
<td>Congenital absence of the median portion of the apex of the tongue. Nes (1958) presented data from several matings supporting single-locus autosomal recessive inheritance</td>
<td>Nes (1958)</td>
</tr>
<tr>
<td>Arthrogryposis multiplex congenita (AMC)</td>
<td>000069</td>
<td>SSC5</td>
<td></td>
<td>Persistent flexion of a joint. Also known as bent-stiff-legged, bentleg and congenital articular rigidity (CAR). Can be caused by various non-genetic factors such as ingestion of Jimsonweed or tobacco by pregnant sows. However, there is convincing evidence of a single-locus autosomal recessive form of this disorder (Lomo, 1985), and the disorder has now been mapped to chromosome SSC5 (Genini et al., 2006)</td>
<td>Hallqvist (1933); Genini et al. (2006)</td>
</tr>
<tr>
<td>Ataxia, progressive</td>
<td>001091</td>
<td></td>
<td></td>
<td>A progressive failure of muscle coordination, resulting in perverse movements. Also known as congenital motor defect or congenital ataxia. The central nervous system appears normal at birth, but older pigs show dysplasia of the cerebellar cortex. Rimaila-Parnanen (1982) provided convincing evidence of autosomal recessive inheritance</td>
<td>Rimaila-Parnanen (1982); Genini et al. (2007)</td>
</tr>
<tr>
<td>Blood group system A</td>
<td>001089</td>
<td>ABO</td>
<td>SSC6</td>
<td>This system has at least two alleles. Also known as the A–O blood group system. Mariani et al. (1996) mapped the locus to SSC6</td>
<td>Sprague (1958); Mariani et al. (1996)</td>
</tr>
<tr>
<td>Blood group system B</td>
<td>000120</td>
<td></td>
<td></td>
<td>This system has at least two alleles</td>
<td>Baker and Andresen (1964); Erhard et al. (1988)</td>
</tr>
<tr>
<td>Blood group system C</td>
<td>000121</td>
<td></td>
<td></td>
<td>This system has at least two alleles</td>
<td>Andresen and Baker (1964); Rasmussen (1982)</td>
</tr>
<tr>
<td>Blood group system D</td>
<td>000122</td>
<td></td>
<td></td>
<td>This system has at least two alleles</td>
<td>Andresen (1962); Erhard et al. (1988)</td>
</tr>
<tr>
<td>Blood group system</td>
<td>Reference(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Andresen and Irwin (1959a); Hojny and Nielsen (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Andresen (1957); Erhard et al. (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Andresen (1957); Erhard et al. (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Andresen (1957); Zeveren et al. (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>Andresen (1957); Andresen (1966)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>Andresen (1957); Hradecky et al. (1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Andresen and Irwin (1959b); Erhard et al. (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Andresen (1962); Marklund et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Nielsen (1961); Nielsen (1991)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Hala and Hojny (1964)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Hojny and Hala (1965); Mariani et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetans</td>
<td>Hjarre (1953); Evensen (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dermatosis

A well-characterized syndrome comprising skin lesions on the body, swollen feet (club foot), and multinucleate giant cells (MGC) in the lungs associated with fatal pneumonia. Occurs only in the Landrace breed. Breeding data support single-locus autosomal recessive inheritance (Flatla et al., 1961). Evensen (1993) showed that the giant cells in the lungs of affected pigs are derived from mesenchymal cells, with a monocyte/macrophage origin.
<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwarfism</td>
<td>000299</td>
<td>COL10A1</td>
<td>SSC1: 85743574–85750572</td>
<td>There have been several isolated reports of dwarfism in pigs, with the most convincing single-locus evidence being provided by Jensen et al. (1984). The two reports with the greatest clinical detail are those by Kaman et al. (1991) and Shirota et al. (1995). Nielsen et al. (2000) identified a missense mutation in the COL10A1 gene in a form of dwarfism in pigs that is analogous to Schmid metaphyseal chondrodysplasia (SMCD) in humans.</td>
<td>Petrov (1974); Nielsen et al. (2000)</td>
</tr>
<tr>
<td>Epitheliogenesis imperfecta</td>
<td>000348</td>
<td></td>
<td></td>
<td>Congenital absence of areas of skin. Also known as aplasia cutis. Affected animals usually die within 3 days, but some survive to adulthood. The chance of survival is indirectly proportional to the area of skin that is absent. A single-locus autosomal recessive disorder.</td>
<td>Nordby (1929); Benoit-Biancamano et al. (2006)</td>
</tr>
<tr>
<td>Gangliosidosis, GM2</td>
<td>000403</td>
<td></td>
<td></td>
<td>A lysosomal storage disease in which there is a build-up (storage) of GM2 gangliosides (a type of glycolipid) in various tissues, due to the lack of the enzyme hexosaminidase, whose task is to break down the GM2 ganglioside into its constituents. Characterized by progressive neuromuscular dysfunction and impaired growth from an early age.</td>
<td>Read and Bridges (1968); Kosanke et al. (1978)</td>
</tr>
<tr>
<td>Heterochromia iridis</td>
<td>000468</td>
<td></td>
<td></td>
<td>Difference in colour of the iris in the two eyes, or in different areas of one iris. In pigs, part (partial heterochromia) or all (complete heterochromia) of the iris lacks pigment. The latter form is called glass-eye. Reconciling data from Durr (1937) and Gelati et al. (1973), Ollivier and Sellier (1982) concluded that bilateral complete heterochromia iridis is autosomal recessive, and that heterozygotes show unilateral and partial heterochromia.</td>
<td>Durr (1937); Gelati et al. (1973)</td>
</tr>
<tr>
<td>Trait</td>
<td>Gene Symbol</td>
<td>Chr</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hind limb paralysis</td>
<td></td>
<td></td>
<td>Evidence for autosomal recessive inheritance of hind limb paralysis was presented by Berge (1941). Ollivier and Sellier (1982) cited a study by Ludvigsen et al. (1963) as providing evidence for autosomal recessive inheritance of paralysis of the hind limbs, associated with abnormal lumbar vertebrae. Because the Berge disorder did not involve vertebral abnormalities, it is possible that more than one locus can give rise to hind limb paralysis. Berge (1941); Ludvigsen et al. (1963)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypercholesterolaemia, spontaneous</td>
<td>LDLR</td>
<td>SSC2: 52433429–52437209</td>
<td>An excess of cholesterol in the blood. Also called cholesterolaemia. A strain of pigs that shows spontaneous development of hypercholesterolaemia has a mutation in the gene for apolipoprotein B ($\text{apoB}$), designated $\text{Lpb5}$. This mutant $\text{apoB}$ allele is associated with low-density lipoprotein (LDL) particles that are deficient in their ability to bind to the LDL receptor (Purtell et al., 1993), and that are therefore cleared from the circulation more slowly. The resultant hypercholesterolaemia, however, is not a single-locus trait. Indeed, there appears to be at least one other gene of large effect involved in this disorder (Aiello et al., 1994). A genome scan conducted by Hasler-Rapacz et al. (1998) showed that the gene for this disorder in pigs maps near to the centromere of chromosome 2, which is homologous to the region of human chromosome 19 containing the gene for the low-density lipoprotein receptor ($\text{LDLR}$), a strong candidate for involvement in this disorder. Sequence analysis of the $\text{LDLR}$ gene from homozygous normal and affected pigs showed that the disorder is due to a single missense mutation (resulting in the amino-acid substitution Arg84Cys). Lee et al. (1990); Pena et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypotrichosis, dominant</td>
<td></td>
<td></td>
<td>An autosomal dominant form of hairlessness (or hypotrichosis). The listing here of two separate loci for hypotrichosis is based on precedent (Ollivier and Sellier, 1982; Robinson, 1991) and on supposition of the present author. Breeding trials to test non-allelism have not been reported. Meyer and Drommer (1968)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name of phene</td>
<td>OMIA number</td>
<td>Gene (if known)</td>
<td>Location of gene</td>
<td>Description</td>
<td>Earliest and most recent reference</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Hypotrichosis, juvenile with age-dependent emphysema</td>
<td>001458</td>
<td>SSC15</td>
<td>SSC15</td>
<td>This disorder is characterized by hairlessness until puberty, and localized emphysema in the lungs following puberty (Bruun et al., 2008). Similarity of the phenotype with an integrin beta 6 knockout mouse disorder led Bruun et al. (2008) to show that this disorder in pigs maps to the location of this gene (ITGB6) in pigs, i.e. SSC15. However, no lesion in the porcine version of this gene or in the gene for the other subunit of integrin beta 6 (ITGAV) has yet been found</td>
<td>Bruun et al. (2008)</td>
</tr>
<tr>
<td>Hypotrichosis, recessive</td>
<td>001279</td>
<td></td>
<td></td>
<td>An autosomal recessive form of hairlessness (or hypotrichosis)</td>
<td>Roberts and Carroll (1931)</td>
</tr>
<tr>
<td>Legless</td>
<td>00587</td>
<td></td>
<td></td>
<td>The absence of all four legs. Also called streamlined. Johnson and Lush (1939) and Johnson (1940) provided convincing evidence of autosomal recessive inheritance of this lethal disorder</td>
<td>Johnson and Lush (1939); Johnson (1940)</td>
</tr>
<tr>
<td>Lymphosarcoma</td>
<td>000615</td>
<td></td>
<td></td>
<td>A malignant neoplastic disorder of lymphoid tissue. Also called lymphoma. Sometimes also called leukaemia, although, strictly speaking, this is a different disorder (namely, a malignant disorder of the blood-forming tissue). Clinical signs of lymphosarcoma in pigs include stunted growth, pot belly, enlargement of superficial lymph nodes, an increase in circulating lymphocytes, and death before 15 months of age. Post-mortem reveals sarcoma involving all lymph nodes, but primarily those draining the gut and lung (Head et al., 1974). McTaggart et al. (1979) provided convincing evidence that this disorder is autosomal recessive in pigs</td>
<td>McTaggart et al. (1971); Hejazi and Danyluk (2005)</td>
</tr>
</tbody>
</table>
A progressive increase in body temperature, muscle rigidity and metabolic acidosis, leading to rapid death. In pigs, malignant hyperthermia (MH) leads to rapid post-mortem changes in muscle, resulting in pale soft exudative (PSE) meat. MH can be triggered by a minor stress, such as loading, transport, sexual intercourse, high ambient temperature, or exposure to the anaesthetic halothane. Susceptibility to halothane-induced MH is an autosomal recessive trait in pigs. Together, sudden death and PSE constitute porcine stress syndrome (PSS), which became a major economic problem in many countries in the 1970s. In part, this was due to strong selection for increased leanness, which is associated with susceptibility to PSS. In 1991, a major breakthrough occurred when a Canadian research team led by David MacLennan (Fuji et al., 1991) showed that MH is due to a base substitution (C→T) in the 1843rd nucleotide of the gene for the calcium release channel in the sarcoplasmic reticulum of skeletal muscle. The base substitution causes an amino-acid substitution (arginine→cysteine) in the 615th position of the calcium-release channel, resulting in altered calcium flow. A PCR genotyping test based on the causative mutation has been used extensively in many countries, leading to the elimination of the mutation from many herds. See Chapter 16.

The RN (Rendement Napole) mutation was first documented by Leroy et al. (1990). In longissimus dorsi muscles, carriers of this dominant gene show lower pH, higher surface and internal reflectance values, lower protein extractability, lower water-holding capacity, lower Napole yield (yield after curing and cooking) and greater cooking loss. On the positive side, carriers have a lower shear force value, a stronger taste and smell and greater acidity. The primary cause of these differences is that the mutant allele results in higher stored glycogen content in muscle. The gene is located on the long arm (q21–22) of chromosome 15.
Table 4.1. Continued.

<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membranoproliferative glomerulonephritis type II</td>
<td>000636</td>
<td>CFH</td>
<td>SSC (Sus scrofa) chromosome number: position in Sscrofa9 (pig genome) assembly (from <a href="http://www.ensembl.org">http://www.ensembl.org</a>)</td>
<td>A progressive inflammation of the capillary loops in the glomeruli of the kidney in which the glomeruli become enlarged as a result of the proliferation of mesangial cells and irregular thickening of the capillary walls, due to massive glomerular deposits of complement component C3 and the terminal C5b-9 complement complex. This disorder is due to a deficiency of complement factor H, whose task is to restrict the formation of C3 convertase. The hypermetabolism of the excess C3 results in the deposits of complement in the kidney, and in hypocomplementaemia. The disorder is autosomal recessive in pigs. Affected piglets die at around 5 weeks</td>
<td>Jansen (1993); Hegasy et al. (2002)</td>
</tr>
<tr>
<td>Motor neuron disease, lower</td>
<td>000662</td>
<td></td>
<td></td>
<td>A distinctive locomotor disorder of weaners, characterized by progressive ataxia and paresis of variable severity. Histological examination reveals significant degenerate changes in lower motor neurons, lower (motor) spinal nerve roots, and myelinated axons of peripheral nerves and of ventral and lateral spinal columns (O’Toole et al., 1994b). Lipid-like inclusions and mitochondrial swelling suggest an underlying defect in lipid metabolism and/or mitochondrial metabolism (O’Toole et al., 1994b). Limited breeding data are strongly suggestive of autosomal dominant inheritance, and a breeding colony has been established at the Wyoming State Veterinary Laboratory (O’Toole et al., 1994a)</td>
<td>Wells et al. (1987); O’Toole et al. (1994b)</td>
</tr>
</tbody>
</table>
Neonatal diarrhoea, 001088  
F4 (previously K88)

Neonatal diarrhoea in piglets is often caused by strains of Escherichia coli bacteria that have a cell-surface antigen called K88 (renamed F4), which combines with a glycoprotein receptor on the wall of a piglet's intestine, enabling the bacteria to attach themselves to the intestine. The receptor is a type of transferrin. Once attached, the bacteria proliferate, releasing enterotoxins and thus producing diarrhoea, which can lead to high mortality. Certain piglets lack the intestinal receptor to F4, and are therefore resistant to F4 bacteria and hence to diarrhoea caused by F4 strains. There are several different antigenic variants of F4 (ab, ac, ad), and it seems that there is a separate receptor for each. The presence or absence of at least two of these receptors (for F4ab and F4ac) is determined for each receptor by one of two closely linked genes on chromosome 13. The immunological and population-genetics implications of segregation at these loci are very interesting; the result is selection against heterozygotes, as explained by Nicholas (2010, pp. 132–133). From a three-generation linkage analysis, Python et al. (2002) concluded that the receptor for F4ac is the same as the receptor for F4ab, and that this single receptor is encoded by a gene they have called F4bcR, which they have more finely mapped to SSC13.

Gibbons et al. (1977);  
Joller et al. (2009)

Nucleoside transport defect  
001236

Defect in the transport of nucleosides (purine or pyrimidine base attached to a ribose or deoxyribose sugar) across erythrocyte membranes. This disorder has not been recorded in pigs in vivo. However, in a kidney cell line, Aran and Plagemann (1992) created several different mutants that resulted in failure to transport thymidine and uridine. The faulty gene has not yet been identified.

Aran and Plagemann (1992)

Oedema  
000493

Abnormal accumulation of fluid in tissues and/or body cavities. Also called myxoedema, dropsy or hydrops. In reviewing the evidence summarized by Koch et al. (1957), Olivier and Sellier (1982) concluded that oedema is an autosomal recessive disorder, possibly to do with a thyroid defect.

Young (1952);  
Neeteson (1964)
<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene SSC (Sus scrofa) chromosome number: position in Sscrofa9 (pig genome) assembly (from <a href="http://www.ensembl.org">http://www.ensembl.org</a>)</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polydactyly with otocephalic monster</td>
<td>000811</td>
<td>None</td>
<td>In village pigs of Papua New Guinea, Malynicz (1982) reported an autosomal dominant disorder in which heterozygotes are polydactylous, and homozygotes are 'monsters', having club foot (dactylomegaly, or abnormally large digits) and otocephaly (absence of the lower jaw, with the ears united below the face)</td>
<td>Malynicz (1982)</td>
<td></td>
</tr>
<tr>
<td>Porphyria</td>
<td>000815</td>
<td>Porphyria is a general term for disorders resulting from a deficiency of any one of the six enzymes involved in the biosynthesis of protoporphyrin from aminolaevulinic acid. Some of the intermediates (loosely called porphyrins) are extremely photoreactive. Because a deficiency of any one of these enzymes results in a build-up of intermediates, photosensitivity is a common clinical sign in most species, but not in pigs. In this species, clinical diagnosis is based on discoloration (pink to brown) of the teeth. On autopsy, bones, kidneys and lymph nodes also show a brown discoloration. Another major clinical sign is haemolytic anaemia, due to a deficiency of haemoglobin, of which protoporphyrin is a vital component. There have been only isolated reports in pigs, with only one study describing the likely enzyme deficiency: Roels et al. (1995) attributed their cases of porphyria to a combined deficiency of the third (uroporphyrinogen-III-cosynthetase) and fourth (uroporphyrinogen decarboxylase) of the six enzymes involved in protoporphyrin biosynthesis. Their cases would therefore be a 'mixed' porphyria, comprising aspects of congenital erythropoietic porphyria and porphyria cutanea tarda</td>
<td>Clare and Stephens (1944); Roels et al. (1995)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progressive myopathy</td>
<td>000829</td>
<td>None</td>
<td>A progressive degeneration of the hind legs, resulting in collapse. Also known as creeper syndrome. Wells et al. (1980) provided evidence for autosomal recessive inheritance</td>
<td>Wells and Bradley (1978); Wells et al. (1980)</td>
<td></td>
</tr>
</tbody>
</table>
Protamine-2 deficiency 000834

In most mammals, protamine-2 constitutes the major component of basic protein in sperm nuclei. It binds to DNA during the elongation of spermatids. Pigs and cattle have the protamine-2 gene, but produce very little protamine-2. In the case of pigs, this is due to a deletion of 27 bases in the middle of the gene. It appears that all pigs are homozygous for this mutation. Presumably, the mutation occurred before the splitting of the porcine evolutionary lineage. Maier et al. (1990)

Renal cysts 001257

The occurrence in the kidney of closed epithelium-lined sacs containing a liquid or semi-solid substance. Wijeratne and Wells (1980) reported breeding data that provided convincing evidence of autosomal dominant inheritance. Wijeratne and Wells (1980)

Resistance to oedema disease 000862 FUT1 SSC6: 37631354–37632621

Not to be confused with oedema (see separate entry), oedema disease is a fatal infectious disease of weaners and growers, characterized by lack of coordination, a hoarseness of voice, weakness, flaccid paralysis and blindness. Although oedema of the eyelids, face and ears is diagnostic, this trait is seldom visible during clinical examination. Also called gut oedema or bowel oedema. Caused by opportunistic colonization by certain strains of E. coli. Bertschinger et al. (1993) provided substantial evidence from breeding experiments for single-locus autosomal recessive inheritance of resistance to experimental challenge with a causative strain of E. coli. The gene is called ECF18R (for E. coli receptor F18), with alleles B for susceptibility and b for resistance. Vogeli et al. (1996) mapped ECF18R to the middle of the Hal linkage group on chromosome 6. Noting that one of the genes in this linkage group is the S gene (for suppression of blood group system A), and that the likely human homologue of S is blood group system I-I (whose locus encodes the enzyme alpha-fucosyltransferase, FUT), Meijerink et al. (1997) isolated porcine homologues of FUT, and showed that they mapped to the same location as the S gene, i.e. the recombination fraction between FUT and S is zero. By sequencing one of the FUT genes (FUT1) in resistant and susceptible pigs, these same workers then identified two base substitutions that are in strong linkage disequilibrium with resistance/susceptibility. In the short term, these polymorphisms provide valuable markers for resistance. In the not-too-distant future, we will know whether FUT is actually the ECF18R gene. Bertschinger et al. (1993); Coddens et al. (2008)
Table 4.1. Continued.

<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory distress syndrome</td>
<td>000101</td>
<td></td>
<td>SSC (Sus scrofa) chromosome number: position in Sscrofa9 (pig genome) assembly (from <a href="http://www.ensembl.org">http://www.ensembl.org</a>)</td>
<td>Also known as Barker syndrome, because of the resemblance between affected piglets and Barker foals (Gibson et al., 1976). Characterized by very small thyroid glands, hairlessness, retarded ossification, delayed haemopoiesis and death soon after birth due to acute respiratory distress (Wrathall, 1976). Breeding data presented by Wrathall (1976) provide convincing evidence of autosomal recessive inheritance</td>
<td>Gibson et al. (1976); Wrathall (1976); Wrathall et al. (1977)</td>
</tr>
<tr>
<td>Sex reversal: XX male</td>
<td>000901</td>
<td></td>
<td></td>
<td>Switorski et al. (2002) provided independent confirmation of the autosomal recessive inheritance of SRY-negative karyotypically XX pigs with varying degrees of phenotypic maleness</td>
<td>Pailhoux et al. (1994); Switorski et al. (2002)</td>
</tr>
<tr>
<td>Sperm, short tail</td>
<td>001334</td>
<td>KPL2</td>
<td>SSC16: 18508929–18759937</td>
<td>Known as the immotile short-tail sperm (ISTS) defect, this disorder is due to an insertion of a retrotransposon in the KPL2 gene (Sironen et al., 2006)</td>
<td>Andersson et al. (2000); Kopp et al. (2008)</td>
</tr>
<tr>
<td>Syndactyly</td>
<td>000963</td>
<td></td>
<td></td>
<td>Fusion of the digits. Convincing evidence for autosomal dominant inheritance was provided by Simpson and Simpson (1908), soon after the rediscovery of Mendelism. Breeding data reported by Deltsfsen and Carmichael (1921) confirmed this conclusion. There have been no recent reports of breeding data for this disorder</td>
<td>Auld (1889); Leipold and Dennis (1972)</td>
</tr>
<tr>
<td>Thrombopathia</td>
<td>001003</td>
<td></td>
<td></td>
<td>A blood coagulation disorder due to failure of ADP release from platelets following stimulation by aggregation factors such as thromboplastin. Also called storage pool deficiency. Characterized by mild-to-moderate bleeding. Shown to be autosomal recessive by Thiele et al. (1986)</td>
<td>Daniels et al. (1986); Thiele et al. (1986)</td>
</tr>
<tr>
<td>Tremor type A III, congenital (X-linked)</td>
<td>000770</td>
<td>There are several different forms of congenital tremor syndrome, also known as dancing pig disease, trembles or myoclonia congenita. The common clinical feature comprises rhythmic tremors of the head and limbs, being most severe when standing and being absent when sleeping. Except when provoked by external stimulus, e.g. handling or sudden noise, tremors are rarely seen after the first few months of life and usually disappear by the second or third week. Affected pigs have decreased life expectancy due to decreased ability to nurse and to avoid being crushed by the sow. Done and Harding (1967) distinguished between those forms that show lesions in the central nervous system (type A) and those that do not (type B). Within type A, there are two forms caused by infectious agents (I and II) and two inherited forms (III and IV) (Done, 1968). Congenital tremor type III A, also known as cerebrospinal hypomyelinogenesis, is an X-linked recessive disorder.</td>
<td>Harding et al. (1973); Baumgartner and Brenig (1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tremor type A IV, congenital</td>
<td>001020</td>
<td>Background details are given in the previous entry. Congenital tremor type A IV, also known as cerebrospinal dysmyelinogenesis, is an autosomal recessive disorder.</td>
<td>Patterson et al. (1973); Blakemore and Harding (1974)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tremor, high-frequency</td>
<td>001200</td>
<td>SSC7q This disorder was described by Richter et al. (1995), from whom the following description is taken. The disorder is characterized by muscular weakness and a very intense tremor of the legs when standing and walking but not when at rest in a lying position. The intensity of tremor and muscular weakness progressively increases with age, resulting in pronounced postural instability. Also known as the Campus syndrome, after the boar whose offspring first showed the disorder. Breeding data show autosomal dominant inheritance. The disorder maps to SSC7q (Tammen et al., 1999).</td>
<td>Richter et al. (1995); Tammen et al. (1999)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Name of phene</th>
<th>OMIA number</th>
<th>Gene (if known)</th>
<th>Location of gene SSC (Sus scrofa) chromosome number: position in Sscrofa9 (pig genome) assembly (from <a href="http://www.ensembl.org">http://www.ensembl.org</a>)</th>
<th>Description</th>
<th>Earliest and most recent reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin D-deficiency rickets, type I</td>
<td>000837</td>
<td>CYP27B1</td>
<td></td>
<td>Vitamin D-deficiency rickets, type I. Vitamin D (cholecalciferol) is synthesized in the skin from 7-dehydrocholesterol by the action of UV radiation from sunlight. Cholecalciferol, however, has very little biological activity: it requires two hydroxylations in order to become (biologically) active. The first hydroxylation, catalysed by cholecalciferol 25-hydroxylase, occurs in the liver. The second of these hydroxylations occurs in the kidney under the action of the enzyme 25-alpha-hydroxycholecalciferol 1-hydroxylase. The resultant active form of vitamin D (called 1,25-dihydroxycholecalciferol or 1,25(OH)sub2D) is a steroid hormone that plays a vital role in whole-body calcium homeostasis. Vitamin D-deficiency rickets, type I (previously known as pseudo-vitamin D deficiency rickets) is an inherited deficiency of the 1-hydroxylase enzyme. As expected, this deficiency results in clinical signs indistinguishable from those seen in individuals suffering from non-genetic lack of vitamin D, most commonly resulting from a dietary deficiency of calcium or insufficient exposure to sunlight. The clinical signs of rickets (inherited and non-genetic) arise from defects in calcium homeostasis. The most noticeable effects include a failure of calcification of bones (leading to bowing of limbs) and delayed dentition. The molecular basis of this disorder in pigs was determined by Chavez et al. (2003), who showed that in Hannover pigs the clinical signs result from either of two deletions in the gene for cytochrome P450C1 (otherwise known as CYP27B1)</td>
<td>Meyer and Plonait (1968); Chavez et al. (2003)</td>
</tr>
<tr>
<td>Trait</td>
<td>Code</td>
<td>Description</td>
<td>References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>von Willebrand disease</td>
<td>001056</td>
<td>Von Willebrand factor (vWF) is a multimeric form of a plasma protein encoded by an autosomal gene (not yet mapped in pigs). vWF plays a vital role in platelet adhesion and clot formation. It also combines with factor VIIIIC (the product of the X-linked haemophilia A locus), forming factor VIII. vWF accounts for 99% of the mass of factor VIII; its role is to protect factor VIIIIC from degradation. von Willebrand disease (also called pseudo-haemophilia or vascular haemophilia) is an autosomal bleeding disorder resulting from deficient or defective vWF. The molecular basis of this disorder in pigs has not yet been determined.</td>
<td>Hogan et al. (1941); Sauger et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wattles</td>
<td>001061</td>
<td>Appendages suspended from the head. Also known as tassels or bells, they are fleshy masses of cartilaginous material covered with normal skin and suspended from the mandibular area. Roberts and Morrill (1944) provided good evidence for single-locus autosomal recessive inheritance.</td>
<td>Kronacher (1924); Roberts and Morrill (1944)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woolly hair</td>
<td>001256</td>
<td>Curly coat, as seen in the Canastrao breed. An autosomal dominant trait, which may be allelic with recessive hypotrichosis.</td>
<td>Rhoad (1934)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Zeitschrift für Zuchtungsbiologie B 37, 129–158.


Hala, K. and Hojny, J. (1964) Blood groups of the N system in pigs. Folia Biologica 10, 239–244.


Meijerink, E., Fries, R., Vogeli, P., Masabanda, J., Wigger, G., Stricker, C., Neuenschwander, S., Bertschinger, H.U. and Stranzinger, G. (1997) Two alpha(1,2) fucosyltransferase genes on porcine chromosome 6O11 are closely linked to the blood group inhibitor(s) and *Escherichia coli* F18 receptor (ECF18R) loci. *Mammalian Genome* 8, 736–741.


Introduction

There has been enormous progress in the molecular characterization of the genome of the pig, and indeed many other mammalian species, in the 12 years since the publication of the first edition of *The Genetics of the Pig*. A partial genome sequence has been available for a number of years (Wernersson *et al.*, 2005); a complete genome sequence is imminent (see Chapter 8); and technical developments in sequencing methodology, so-called NextGen sequencing, are set to further revolutionize discovery and characterization of genetic variation in the domestic pig and its close relatives.

However, there have also been enormous conceptual developments in our understanding of mammalian genomes in that same time period that reinforce the need for continued extrapolation or inference of properties of the pig genome from other mammals, such as humans or the mouse. While it is reasonable to assume that the pig has a typical mammalian genomic structure and molecular biology, given the universality of some features across all mammals studied, and also given the concordance with other mammals in those features specifically examined in the pig, this review will cite studies specific to the pig wherever possible and will extrapolate only where necessary.
Perhaps the most profound conceptual leap arises from the realization that there is at least an order of magnitude more transcripts than the approximately 20,000 protein coding genes in the genome of mammals, and that the transcription of non-protein- (or non-) coding RNAs (ncRNAs) is pervasive. Indeed, a substantial proportion of recognized transcripts are for ncRNA, many of which are initiated from the 3′ UTRs (three prime untranslated mRNA regions) of protein-coding loci (The FANTOM Consortium et al., 2005). The recognition that the majority of the mammalian genome is transcribed, much of it from both DNA strands (The FANTOM Consortium et al., 2005), profoundly affects our understanding and interpretation of genome sequences of even the best characterized mammals, let alone the less intensively studied pig. It is now very clear that information in the genome does not lie only in short collinear, mainly protein-encoding exonic sequences comprising 2% of the genome and separated by 98% of long, informationally devoid deserts. Instead, it appears that information is encoded in a multifunctional, overlapping, even bidirectional fashion, dispersed across the genome and described by Kapranov et al. (2007) as an interleaved model.

One important component of the ncRNA spectrum of the pig genome not even mentioned in the first edition of this chapter, namely microRNA (miRNA), is now partially characterized in pigs and can be compared with similar sequences in other mammals. Because miRNAs have important regulatory and developmental functions and are amenable to manipulation to control gene expression and even to attack pathogens as short interfering (or short inhibitory) RNA (siRNA), they will receive commensurate weight in this review.

Since 1998, porcine endogenous retroviruses (PERVs), another component of the pig genome, have received a level of investigation and original paper publication entirely disproportionate to their representation in the genome because of concerns that they might represent a zoonotic hazard to xenotransplantation of porcine tissues into humans. These also will be reviewed in detail.

**Nuclear Genome Size, ‘Gene Number’ and Developmental Complexity**

For many years, the C-value paradox (Hinegardner, 1968; Britten and Davidson, 1971; Thomas, 1971) described the lack of correlation of genome sizes in eukaryotic organisms with their developmental complexity. In vertebrates, at one extreme, the puffer fish, Fugu rubripes, and related tetraodontoid fish have particularly small haploid genomes of only $4 \times 10^8$ bp, the smallest recorded for any vertebrate (Brenner et al., 1993). The Fugu genome, which is 7.5 times smaller than a mammalian genome, contains less than 10% repetitive DNA. At the other extreme, the genome of the lungfish, Protopterus aethiopicus, is 350 times larger than that of the puffer fish at $1.4 \times 10^{11}$ bp. Although many of the more extreme genome sizes, such as that of the lungfish, are now thought to be artefacts of polyploidy, the C-value paradox focused attention on the fact that much of the genome of complex organisms, especially the non-coding and often repetitive fraction that comprises 98% or more in many organisms, had no known function, and indeed was often dismissed as junk DNA owing to the fact that it could vary so substantially in amount between species. If it was doing anything, it was not obvious. Although the genome sizes of mammals, by comparison, are relatively uniform, they still contain a very large proportion of non-coding and repetitive sequences.

If total genomic DNA content does not explain differences in developmental complexity, perhaps the answer might lie in the number of protein-coding sequences. Before the completion of the human genome sequence in 2002, there was much speculation about the number of protein-encoding genes in the genome (e.g. Fields et al., 1994), with initial estimates from about 50,000 to 150,000. The now generally accepted and, to some biologists, disappointingly low figure of about 20,000 for humans and other mammals has created a new paradox, termed the G-value paradox by Hahn and Wray (2002). Not only do mammals have far fewer protein-coding genes than was originally thought, the notably simpler nematode worm, Caenorhabditis elegans, has almost as
many at 19,300 (Stein et al., 2003), and even protozoa such as *Tetrahymena thermophila* have many more protein-coding genes (about 27,000) than mammals (Eisen et al., 2006). Clearly, simple counts of protein-encoding genes do not scale with increasing developmental complexity. Although alternative splicing does provide an enormous increase in the number of protein-encoding transcripts and the complexity of their regulation, which partly relieves the G-value paradox, the discovery of the functional importance of many ncRNAs and pervasive, virtually genome-wide, transcription have led to changes in how we should even think about question of genome size and organismal complexity. Crude comparisons of genome size are not meaningful; nor are crude comparisons of counts of protein-coding sequences. However, a more inclusive definition of the word ‘gene’ to include the many newly discovered classes of non-coding but regulatory RNAs, as well as a recognition of the interleaved rather than collinear pattern of encoding information in DNA, will not only further contribute to relieving the C-value and G-value paradoxes, but also help to explain the 98% or more of the mammalian genome that has previously had no clearly assigned function.

So how do the size and other properties of the nuclear genome of the pig match up with the genomes of other mammals? The estimated genome size, based on flow karyotyping (Schmitz et al., 1992) is $2.72 \times 10^9$ bp for an X-bearing haploid nucleus and $2.62 \times 10^9$ bp for a Y-bearing haploid nucleus. John and Miklos (1988) have reported an approximate twofold range in DNA content in mammals from $2.1 \times 10^9$ bp in the muntjac deer, *Muntiacus muntjak*, to $5 \times 10^9$ bp in the aardvark, *Orycteropus afer*. This places the pig in the low-to-mid range of mammalian genome sizes, although differences in measurement methodology mean that some of these estimates must be interpreted cautiously. Wernersson et al. (2005) have made detailed comparisons of their partial porcine genome sequence with human and mouse sequences and have shown that for all classes of sequence comparison – exonic, 5' and 3' UTR, intronic, intergenic and miRNA – the pig sequence is more similar to the human sequence, despite a similar divergence time of artiodactyls, primates and rodents from a common ancestor.

### Repetitive DNA, GC Content and Isochores

Repetitive sequences in the genomes of many species were first recognized on the basis of reassociation kinetics over 40 years ago. When genomic DNA was made single stranded by heating in solution, it was recognized that some components of the genome were able to reassociate into double-stranded DNA very rapidly when the solution cooled, implying that they were present at very high copy numbers in the genome – sometimes millions of copies. On this basis, it was estimated that 40–60% of the genome of mammals comprised repeated sequences. Repetitive DNA can be broadly classified into two categories, namely tandemly repeated and dispersed repeated DNA. Tandemly repeated DNA ranges from highly repeated sequences, which often consist of small repeat units, to less highly repeated sequences, which may sometimes be functional, for example the ribosomal RNA genes. In the dispersed repeated category are transposable elements, some of which retain the capability of further movement, and pseudogenes.

Caesium chloride density-gradient ultracentrifugation was routinely used in the past for isolating DNA based on density. As the base composition of highly repeated DNA can differ substantially from the overall composition of the genome, many repeated fractions could be recognized and isolated on the basis of their different densities, occurring most obviously as distinct satellite bands in the CsCl gradients. More subtly, Bernardi et al. (1985) used CsCl gradient ultracentrifugation to partition the nuclear DNA of warm-blooded vertebrates into up to six density classes (L1 to L3, H1 to H3) determined by increasing GC content, which they termed isochores. Protein-coding sequences are found in the heaviest isochores, with about 50% of human protein-coding sequences found in H2 and H3, which comprise only 8% and 4%, respectively, of the genome. The clustering of coding genes into
GC rich isochores correlates with the clustering of protein-coding sequences into cytogenetically observable R bands, which are GC rich.

The spectrum of sequences contributing to the Wernersson et al. (2005) incomplete genome sequence for the pig comprised approximately 34% of repetitive sequences, of which 11.3% were SINE (short interspersed nuclear element) sequences, 16.14% LINEs (long interspersed nuclear elements), 2.8% LTR (long terminal repeat) elements (including endogenous retroviruses), 0.9% DNA transposons, 1.47% satellites, 0.62% simple repeats and 0.53% low complexity repeats.

The mean GC contents of the pig sequences (Wernersson et al., 2005) were found to be 40.7% for introns, 49.6% for coding, 41.8% for 3′ UTR, 59.2% for 5′ UTR and 39.6% for intergenic sequences. The GC compositions across these classes are very similar to those found in humans and mice, and are consistent with earlier predictions about the clustering of coding sequences in GC-rich domains in the genome. It is noteworthy that introns, 3′ UTRs and intergenic sequences all have low GC contents of around 40%, whereas coding sequences have approximately 50% GC. The GC content of 5′ UTRs is close to 60%. Reuter et al. (2008) discuss the reason why 5′ UTRs might evolve biased GC contents.

**Centromeres and centromeric DNA repeats**

Kinetochores are complex eukaryote structures that assemble on centromeric DNA and are essential for regular chromosome disjunction. Some eukaryotes, such as the yeast, *Saccharomyces cerevisiae*, have short centromeric sequences, which specifically recognize centromeric proteins, and form point centromeres. Other eukaryotes, including mammals, have regional centromeres based on very long (megabase) tracts of highly repeated, usually AT-rich DNA, consisting of 120–180 bp repeats (Meraldi et al., 2006; Przewloka and Glover, 2009). However, it is not the direct interaction of these repeat sequences within these tracts with the kinetochore proteins that recruits the kinetochore, but epigenetic modification of the histones associated with this DNA (Dalal, 2009). A particular histone-H3 variant, called CenH3 (also CenP-A in mammals), replaces H3 to identify centric specific nucleosomes. These specially marked nucleosomes then initiate the recruitment of over 40 proteins comprising the kinetochore. As kinetochores normally assemble in the same position, except in rare and exceptional cases of neo-centromere formation, there must be some signal for this epigenetic mark to be made, but this is not yet clearly recognized (Dalal, 2009). Pericentric transcription in the late G2 early M phase of mitosis is correlated with an increase in CenH3 levels. The requirement of the usually transcriptionally quiet heterochromatic environment for centromere formation is supported by several cases of neo-centromeres that have occurred in gene-poor regions with little transcriptional activity.

Two distinct centromeric repeat families have been characterized in the pig. Jantsch et al. (1990) have described the Mc1 family of repeats, restricted to the metacentric chromosomes in the pig karyotype, and the Ac2 family, whose distribution is restricted to the acrocentric chromosomes. The repeat unit of the heterogeneous Mc1 family is about 100 bp. Some family members occur in the pericentromeric region of all metacentric chromosomes, whereas others are found only on specific metacentric chromosomes. The Ac2 family is homogeneous and found in the subterminal pericentromeric regions of all acrocentric chromosomes. The Ac2 family is derived from a highly conserved 14 bp repeat unit. Jantsch et al. (1990) have speculated that the association of the centromeres of porcine acrocentric, but not metacentric, chromosomes observed during the pachytene of meiosis may allow interchromosomal exchanges that have maintained the homogeneity of the Ac2 family. Miller et al. (1993) have independently isolated what appears to be another member of the Mc family of repeats and showed that it hybridizes to the centromeres of all chromosomes except the Y under conditions of low stringency, but only to certain metacentric chromosomes under conditions of high stringency.

Apart from the centromeric repeats, McGraw et al. (1988) have isolated a 3.8kb repeat element with 80% sequence identity to an element independently isolated by Mileham et al. (1988). This element is male specific,
at least at the level of sensitivity of dot blot and Southern hybridization, with more than 200-fold greater copy number in males than females. It contains no internal repetitions and no open reading frames. Akamatsu et al. (1989) subsequently isolated and sequenced a 1.25 kb repetitive element visible in EcoR1 digests of both male and female porcine genomic DNA. This contains imperfect internal repetitions of about 44 bp and has been used as a gender-neutral probe for verifying the presence of target DNA in embryo sexing assays.

Dispersed repetitive elements

Transposable elements comprise an important group of dispersed repetitive elements, which have important biological features, including the capacity for insertional mutagenesis and at least a theoretical capacity for horizontal transmission. They can be classified into two categories, namely DNA-mediated elements (class 2 transposons or DNA transposons) and RNA-mediated elements (class 1 transposons, that is retrotransposons and retroposons). In mammals, RNA transposons comprise 90–95% of transposable elements, whereas in insects, plants and protozoa, DNA transposons predominate (Feschotte and Pritham, 2007). The glimpse of the porcine genome provided by the Wernersson et al. (2005) partial sequence indicates that 92.6% of porcine transposable elements are retrotransposons. If SINEs are included in the RNA-mediated class, this proportion increases to 95.2%.

DNA-mediated elements

DNA-mediated elements transpose via a DNA intermediate, generally using a cut-and-paste mechanism, although some use a rolling circle method of replication and others use methods not yet understood (Feschotte and Pritham, 2007). Twelve superfamilies of DNA transposons have been recognized (Feschotte and Pritham, 2007), although only five of these, comprising 120 families, have been found in humans. Comparisons across many well-characterized mammalian species, including humans, mice, rats and dogs, indicate that DNA transposons have been relatively inactive in mammalian genomes for the past 40–50 million years, although there are some rare exceptions, such as species of bats, which have a distinct spectrum of DNA transposons, some of which still appear to be active (Feschotte and Pritham, 2007).

DNA transposons were discovered independently in Drosophila and maize as a result of the mutational effects of insertion of the elements, but have only been found in mammals through sequence analysis. Morgan (1995) first discovered sequences in humans that are quite closely related to the mariner transposon of Drosophila, and are known also to be present in a wide range of arthropods and other invertebrates. Mariner-like elements (MLEs) are about 1250bp in length with inverted repeats of 20–40 bp. Auge-Gouillou et al. (1995) have identified MLEs in humans, mice, rats, Chinese hamsters, sheep and cattle. Sequence analysis of the human, bovine and ovine MLEs, which were the most closely related from this group, indicated that they were more similar to the MLEs found in Hyalophora cecropia, a lepidopteran, than to the mariner elements of Drosophila. Indeed, the mammalian and lepidopteran MLEs were more similar to each other than to MLEs from the two insect groups, although it should be pointed out that the mammalian elements were isolated with degenerate primers for Cecropia MLEs. It has been estimated (Auge-Gouillou et al., 1995) that there are more than 100 MLEs in the human, ovine and bovine genomes. Most mammalian MLEs are thought to be the decaying and non-functional remnants of ancient horizontal transmission, although Reiter et al. (1996) have found indirect evidence that MLE transposase could be expressed in humans.

In pigs, MER1 type and MER2 type elements comprise approximately 0.9% and 0.3% of the total sequence respectively (Wernersson et al., 2005); the MER (medium reiterated repeat) classes are arbitrary DNA transposon classes recognized by the repetitive DNA screening software RepeatMasker (Smit et al., 1996–2010).

RNA-mediated elements

RNA-mediated elements (retrotransposons, and also retroposons) comprise the other major category of transposable elements and
transpose via an RNA intermediate, usually with the original inserted copy remaining in position. They use reverse transcriptase, an enzyme that synthesizes DNA from an RNA template, and makes a DNA copy from the RNA intermediate before integration of the element into the genome. The two main categories of retrotransposons are the retroviruses, which have LTRs, and the LINEs, which do not. Retrotransposons encode their own reverse transcriptase, as well as integrase and other enzymes necessary for transposition. Retroposons do not encode reverse transcriptase and thus are non-autonomous, depending on reverse transcriptase activity provided by retroviruses or LINEs. SINEs have neither LTRs nor reverse transcriptase and generally have more or less degenerate poly(A) tails. Like pseudogenes, they must depend on reverse transcriptase and an integration mechanism parasitized from LINEs or retroviruses (Singer, 1982; Ohshima and Okada, 2005).

**RETROVIRUSES.** Endogenous retroviruses, mostly defective or incomplete, comprise a substantial proportion of the genomes of most mammals. Analysis of the initial human genome sequence assembly (IHGSC, 2001) found that there are about 450,000 retrovirus-like elements in the human genome comprising about 8% of the total DNA, although a large proportion of these are retroviral fragments, and very few complete retroviruses are present. The preliminary analysis of the incomplete genome sequence of the pig (Wernersson et al., 2005) has found that 2.8% of the genome sequences consist of retroviral-like elements. It is unclear whether this lower estimate in pigs reflects a biological reality or is an artefact of the incompleteness of the sequence or its assembly, or is due to ambiguity of the definition of what comprises a retroviral sequence. Respective comparison between pigs and humans for LINEs (16% versus 21%) and SINEs (11.3% versus 13%) but not for DNA transposons (1.5% versus 3%) suggests that perhaps pigs may actually have a lower occurrence of endogenous retroviral DNA in their genomes than humans.

As a class of retrotransposon, retroviruses encode their own reverse transcriptase, integrase and other enzymes necessary for transposition. Complete retroviruses also encode an env gene, which specifies the envelope proteins required for packaging infectious retroviral particles. Although these are generally capable of completing the infectious viral cycle and thus are capable of horizontal transmission, most complete endogenous viruses are stably inherited and are vertically transmitted only. Todaro et al. (1974) were the first to show that the PK-15 type C retroviruses secreted by pig cells in culture are present in multiple copies in the DNA of all pig tissues and cells. Thus, these retroviruses are endogenous and transmitted vertically. Benveniste and Todaro (1975) claimed that a related virus was found only in close wild relatives within the Suidae, such as the bush pig and the warthog, and was absent from the peccary and other artiodactyls such as cattle.

Interest in porcine retroviruses has been greatly intensified because they are considered to be a potential zoonotic hazard whose potential impact could be amplified if xenotransplantation of porcine tissue into humans were to become a clinical reality. A literature too large to review completely in this chapter has developed. Briefly, PERVs are members of the Gammaretroviridae. Members of two subfamilies, \( \gamma_1 \) and \( \gamma_2 \), have been extensively characterized. The \( \gamma_2 \) viruses have also been called PERV-E, owing to their similarity to the human endogenous retrovirus (HERV)-E. There are estimated to be up to 50 copies per host genome. These endogenous viruses have homogeneous envelope sequences; most are defective and they are believed to be of low infectious potential to humans (Klymiuk et al., 2006), so are now considered to be of little relevance to xenotransplantation.

\( \gamma_1 \) PERVs have generated an enormous level of interest ever since Patience et al. (1997) showed that at least some of them were capable of infecting human cells in vitro. \( \gamma_1 \) retroviruses are classified into PERV-A, B and C subfamilies on the basis of very distinct envelope gene sequences. There are up to about 50 copies of more or less intact endogenous retroviruses per host genome, but most of these are defective (for example Lee et al., 2002), with
A and B most common and C sometimes absent. Tropism, that is, the preference for replication in cells from some species or tissues rather than others, is determined by the envelope gene sequence. Novel viruses and novel envelope sequences can be generated by recombination, and indeed the PERV most capable of infecting human cells is a de novo A/C recombinant (Oldmixon et al., 2002). The possibility of recombination between human and porcine retroviruses has been argued as a remote but potentially catastrophic risk, but such recombinants have not been observed. Pseudotyping, the packaging of a pig retroviral genome in a human retroviral envelope or vice versa, a prerequisite for recombination, has not been observed. Indeed, in vivo infection of humans with porcine retroviruses has never been observed. In vivo infection of humans with porcine retroviruses has never been observed. The development of lines of PERV-C free pigs such as the Westran (Moran, 2008), incapable of producing A/C recombinants, together with the use of siRNA molecules targeting endogenous retroviruses in transgenically modified pigs (Dieckhoff et al., 2008; Ramsoondar et al., 2009), is contributing to increasing confidence in the safety of xenotransplantation.

**LINES.** LINES are a class of retrotransposon and are distinguishable from endogenous retroviruses mainly by their lack of env sequences and LTRs. Although LINEs encode reverse transcriptase, only a very small proportion of the elements in the genome contain a fully functional reverse transcriptase gene which has escaped mutational inactivation. Most LINE elements are truncated and stranded in their current genomic location, where they are doomed to gradually decay into random sequence due to unconstrained mutation. One LINE family, the mouse L1 family, is very old, and homologous elements have been found in plants and even in protists, leading to the suggestion that this is the ancestor even of retroviruses, which acquired their infectious capability secondarily. About 21% of the human genome consists of LINE sequences, with an estimated 850,000 copies (IHGSC, 2001). In the pig, Wernersson et al. (2005) reported that about 16% of their incomplete porcine genome sequence comprised LINES, of which LINE1 (14.5%) was predominant. In humans and mice, LINE elements predominate in the G+C-poor G-bands. Miller (1994) described an early use of a porcine LINE, which he called L1Ss, in PCR genotyping of pigs for the Hox 2.4 gene. Thomsen and Miller (1996) subsequently found that this porcine LINE is uniformly distributed throughout the euchromatic part of the porcine genome, with a slight bias towards G-bands.

**SINES.** The use of the term SINE will be restricted here to elements that transpose via an RNA intermediate, although it is occasionally used in the literature for any small interspersed element. Because of their small size, SINEs do not encode reverse transcriptase, although some are transcribed. SINEs were originally discovered in humans (Alu repeats) and mice (B1 and B2 repeats) owing to their very high copy number in the genome: the Alu repeat occurs about 500,000 times in the human genome. SINEs are generally less than 200 bp in size and are derived from small cellular RNA species, which have been reverse transcribed and then integrated into the genome. In effect, they are processed pseudogenes that have attained extremely high copy number. Structural RNAs, such as transfer RNAs (tRNAs), 5S ribosomal RNA (rRNA) and 7SL RNA (a component of signal recognition particle ribonucleoprotein), have independently become the target for reverse transcription and integration in humans and mice to form SINEs. It is believed that certain mutations within structural RNAs predispose them to become templates for reverse transcriptase. In particular, nucleotide changes in the 3′ region of an RNA that cause self-complementarity may lead to the formation of hairpin loops, and this intra-strand priming allows these RNAs to act as templates for synthesis of a DNA copy (cDNA – complementary DNA) by reverse transcriptase. In the presumably rare circumstances of the expression of reverse transcriptase in a cell, these mutated structural RNAs form effective templates for the synthesis of multiple cDNAs that are able to integrate into the genome.

In the pig, SINEs are ubiquitous. Three families of SINEs have been recognized, with the PRE-1 (porcine repetitive element-1) family
(Singer et al., 1987) being the most abundant of these. The consensus PRE-1 element is 233 bp in length, and Singer et al. (1987) have estimated that there are 50,000–100,000 copies interspersed throughout the porcine genome. Ellegren (1993a,b) has estimated from database searching that PRE-1 occurs about once every 12 kb in the porcine genome, implying that there are 250,000 copies if the element is randomly distributed. Other estimates (Frengen et al., 1991; Takahashi et al., 1992) are as high as 500,000. Yasue et al. (1991), using a Hinc fragment containing three PRE-1 elements, discovered evidence for an uneven distribution of PRE-1 along the chromosomes, surprisingly favouring the vicinity of the centromeres. They found no correlation with the distribution of Q-bands. In contrast, Sarmiento et al. (1993) have shown by in situ hybridization that a PRE-1 element, which they called C11, is distributed uniformly throughout the porcine genome, but spares the centromeres of acrocentric chromosomes. The element used by Sarmiento et al. (1993) was cloned from within the porcine MHC (major histocompatibility complex) and showed a pattern of apparent non-random association with other porcine repetitive elements within the MHC, but not elsewhere in the genome. It is difficult to reconcile these observations unless there are subfamilies of the PRE-1 element with different patterns of genomic distribution. Thomsen and Miller (1996) have also examined the distribution of PRE-1 and have found no reproducible clustering of label, but, in agreement with Sarmiento et al. (1993), signal was absent or reduced in the centromere regions, particularly of acrocentric chromosomes; however, they did find some evidence of heterogeneity of distribution of particular elements within the PRE-1 family, with one element apparently preferentially located in strong R-bands.

The 3′ terminus of a PRE-1 element consists of a poly(A) tail of variable length. Miller and Archibald (1993) also found that (CA), (GT), repeats occur at the 3′ termini of PRE-1 in 12% of cases, whereas Alexander et al. (1995) found 24% of their randomly isolated (CA),, (GT), microsatellite loci to be associated with PRE-1. Wilke et al. (1994) found about 36% of their (CA),, (GT), repeats from small insert clones to be associated with PRE-1, with 80% of the (CA),, (GT), repeats occurring at the 3′ end of the SINE. As described elsewhere, both the (A),, (T), and (CA),, (GT), repeats provide a valuable source of genetic markers. The consensus PRE-1 element contains internal imperfect direct repeats of about 38 nucleotides and an RNA polymerase III split promoter, which is found in all the structural RNAs that are transcribed by RNA polymerase III. Singer et al. (1987) found PRE-1 to be transcribed in liver and thymus, but did not test any other tissue. The PRE-1 consensus sequence is homologous to tRNA\textsuperscript{arg} genes (Takahashi et al., 1992), with a similar evolutionary origin to the mouse B2 repeat that is derived from tRNA\textsuperscript{ser} (Daniels and Deininger, 1985).

Alexander et al. (1995) reported two novel families, ARE-1P and ARE-2P (artiodactyl repetitive element-1 (or 2) porcine), during sequencing of clones containing AC, microsatellite sequences. Related AREs were subsequently found in cattle. Alexander et al. (1995) estimated the porcine haploid copy number of ARE-1P as about 3.6 × 10\textsuperscript{4} and about 2.4 × 10\textsuperscript{4} for ARE-2P, with 8 × 10\textsuperscript{3} copies apparently being dimers of ARE-1P and ARE-2P. These dimeric elements generally enclose a (CA),, (GT), microsatellite repeat. True stretches of polyadenylation were not observed in all ARE-1P and ARE-2P elements, although the high adenine content implies that a poly(A) tail may have been present but has degenerated. Direct repeats within the ARE SINEs were absent, in contrast to PRE-1. The evolutionary origin of the ARE SINEs is not immediately apparent, although the tRNAs and 7S1 RNA can be ruled out. The ARE SINE consensus sequences lack an RNA polymerase III split promoter, implying that they are transcribed by RNA polymerase II (Alexander et al., 1995), which transcribes protein-coding genes and the small nuclear ribonucleoprotein genes.

The PRE-1 family is absent from the Bovidae and more distantly related lineages, such as mice and humans (Takahashi et al., 1992). Yasue and Wada (1996) have shown that PRE-1 is present in warthogs and the collared peccary at about the same level as in domestic pigs. This specificity to the Suidae and Tayassuidae implies that it originated after their divergence from the ruminants. By examining the differences among 22 PRE-1 sequences,
Yasue and Wada (1996) estimated the time of divergence of these elements from an ancestral element as about 43.2 million years before present, setting an upper limit on the age of these three species. In contrast, the presence of ARE SINEs in pigs, cows and possibly sheep implies that these sequences were undergoing retroposition before the radiation of Suidae and Bovidae (Alexander et al., 1995), and thus are much older than PRE-1.

Wernersson et al. (2005) estimate that 11.3% of the porcine genome sequence comprises SINEs, with the ancient MIRs (mammalian-wide interspersed repeats) – which are tRNA-derived SINEs which predate the mammalian radiation – making up 1.75% of the genome. Several laboratories have developed assays for detection of porcine DNA (in the context of meat substitution or mixing) which are based on SINE elements, including Walker et al. (2003). Nikaido et al. (1999) have reported on the use of SINEs (and LINEs) in resolving the monophyletic phylogenetic relationship of pigs and peccaries to the exclusion of hippopotamus.

**Processed Pseudogenes.** Pseudogenes are genomic elements which are non-functional but have recognizable similarity to functional genes. Processed pseudogenes arise from reverse transcription of mRNA followed by integration of the cDNA copy. Processed pseudogenes are smaller than their parental genes because they lack introns, which are spliced out during transcription, but they do possess post-transcriptionally acquired poly(A) tails. They also lack the 5′ and 3′ regulatory sequences necessary for the controlled expression of the parental genomic copy of the gene, including promoter elements. Other than by coincidentally acquiring a new promoter, processed pseudogenes are not transcribed. Freed of adaptive constraints, they decay relatively rapidly owing to the accumulation of mutations. Goncalves et al. (2000) have estimated that there are up to 33,000 processed pseudogenes in the human genome. Remarkably, there are apparently only 51 in the chicken genome, which is attributed to the absence of LINE1 elements (Hillier et al., 2004), as the main chicken LINE, CR1, cannot reverse transcribe polyadenylated mRNAs.

Harbitz et al. (1993) described the first porcine example of a processed pseudogene, derived from the glucosephosphate isomerase (GPI) locus. This is obviously an ancient element, because it contains 181 transitions, 78 transversions, 11 deletions and one insertion, and has several stop codons relative to the functional cDNA, with which it retains only 83% nucleotide identity. Assuming a neutral mutation rate of 0.7–1.0% per nucleotide per million years, Harbitz et al. (1993) estimated that the pseudogene arose 20 million years ago. The pseudogene has been truncated at the 3′ end and lacks the poly(A) tail that typifies most processed pseudogenes. Only a single chromosomal site has been found for this GPI pseudogene.

A more recent example of processed pseudogenes in pigs is provided by ribosomal protein SA (RPSA), which is involved in the binding and internalization of prion proteins. Two processed pseudogenes and a single non-processed pseudogene have been identified and mapped in the pig (Knorr et al., 2007). The processed pseudogenes, RPSAPI and RPSAP3, have three and 13 indels, 36 and 57 substitutions and six and eight premature stop codons, respectively, with respect to the functional gene. The pseudogenes contain LINE and SINE-like repeats as well.

The Pseudofam database (Lam et al., 2009) has catalogued pseudogenes, including non-processed pseudogenes, in ten species with complete genome sequences, including the mouse, rat, human, chimpanzee and dog, but unfortunately not yet the pig. In humans, for example, there are 2.6 pseudogenes on average per parent gene, and about 7% of genes have pseudogenes. In general, housekeeping genes are more likely to have pseudogenes than other loci.

**Minisatellites.**

Minisatellites are tandemly repeated elements, with the tandem repeat groups dispersed throughout the genome. Together with microsatellites, they contradict the artificial dichotomy of repeats into tandem versus dispersed classes. Minisatellites have repeating units of 9–64 or
more nucleotides, and were first described in humans (Jeffreys et al., 1985) but have subsequently been discovered in numerous other species. The number of core repeat units varies enormously at particular sites, providing a potentially useful source of genetic markers. Because of this hypervariability in tandem copy number, minisatellites are also called VNTRs (variable number tandem repeats) (Nakamura et al., 1987).

Coppieters et al. (1990) discovered a porcine minisatellite, consisting of 15/18 bp repeat units, while attempting to clone the serum protein $PO2$ locus. The core sequence showed a high degree of similarity to the minisatellite core sequence found by Jeffreys et al. (1985) in humans. When used in low-stringency hybridization, the clone containing the core repeat units detected a highly variable fingerprint pattern in pigs, proving multiple locations of the element in the genome. It also detected a variable fingerprint pattern in horses and rabbits.

Davies et al. (1992) subsequently described a minisatellite located within the porcine glucosephosphate isomerase locus, with a repeating unit of 39 bp (AGGACCAGGTC ATGTACAGGTAGGCAGAGCTGGTCTG), flanked by divergent repeats of 53/54 bp. The core unit had at least 14 perfect tandem copies. The repeat unit has no homology with other tandem repeat sequences. The minisatellite apparently has but a single genomic location, because a fingerprint pattern was not obtained even under conditions of low hybridization stringency.

Signer et al. (1994) reported a porcine minisatellite isolated from a porcine genomic library using the human core sequence of Jeffreys et al. (1985) as probe. The porcine minisatellite hybridized to multiple genomic locations under conditions of low-to-intermediate stringency, but interpretable single-locus banding patterns could be obtained by high-stringency hybridization and washing. The element was found by linkage analysis and in situ hybridization to map to the pseudoautosomal region of the X and Y chromosomes. No estimation was made of the size of the core unit, variation in number of the repeat units or number of genomic sites of related elements, although allelic variants ranged from 3 to 18 kb.

Coppieters et al. (1994) isolated a porcine minisatellite using a CAC/CAT triplet probe. Several clones were independently isolated, which contained the same 30 bp repeat unit (consensus GATGAGGATGGGGGATTTGG AGATGGATGGA). The minisatellite maps to the telomeric region of chromosome 14q. It was estimated that there are more than 100 tandemly repeated units.

Despite their hyperpolymorphism, minisatellites have fallen into disfavour as genetic markers in recent years, mainly because they are not amenable to PCR, owing to the large repeat size and high repeat number. Their preferential location in telomeric regions of chromosomes (Royle et al., 1988) was another reason for abandoning them as general genetic markers.

**Microsatellites**

Like minisatellites, microsatellites are VNTRs. However, they have the advantage of small repeat units (1–6 nucleotides) and a reasonably small copy number, which makes them very suitable for PCR amplification. The extremely high number of locations, which appear to be randomly located throughout the genome, makes them ideal genetic markers, and the linkage maps of most mammalian species, including the pig, have initially been built largely of microsatellite markers.

Certain categories of microsatellite repeats predominate in mammalian genomes, with (AC)$_n$ repeats being the most widely exploited for mapping in mammals. Moran (1993) provided an early report on the incidence of all possible classes of microsatellite repeats in the pig by systematically searching for repeat clusters in excess of about 20 nucleotides for mononucleotide repeats through to tetranucleotides for all porcine sequences lodged in the GenBank database. The interpretation of the results is complicated by the fact that microsatellites were found in both cDNA and genomic DNA sequences of the approximately 181 nuclear genes represented at that time. Nevertheless, an approximate feel for the range and distribution of types of repeats could be gained. Eight (A)$_n$,(T)$_n$ mononucleotides and one (C)$_n$,(G)$_n$ repeat were found. For the dinucleotides, seven (CA)$_n$,(GT)$_n$ repeats and
three (GA)\textsubscript{n}·(CT)\textsubscript{n} were recognized. Five trinucleotides, representing only four of the 20 possible types, were found. Finally, nine tetranucleotides, representing only six of 60 possible categories, were found. The distribution of the elements within each category was 27.3% mononucleotides, 30.3% dinucleotides, 15.1% trinucleotides and 27.3% tetranucleotides. The (A)\textsubscript{n}·(T)\textsubscript{n} mononucleotides, at 24%, and the (CA)\textsubscript{n}·(GT)\textsubscript{n} dinucleotides, at 21%, were by far the most common categories of specific repeats found. Ellegren (1993b) has also searched the DNA sequence databases for mononucleotide and dinucleotide repeats only, recording the occurrence of runs of ten or more nucleotides. He also found (A)\textsubscript{n}·(T)\textsubscript{n} and (CA)\textsubscript{n}·(GT)\textsubscript{n} repeats to predominate, but found the (A)\textsubscript{n}·(T)\textsubscript{n} repeats to be about five times more common than (CA)\textsubscript{n}·(GT)\textsubscript{n} dinucleotides. Using data from genomic sequences only, he estimated that (A)\textsubscript{n}·(T)\textsubscript{n} mononucleotides occurred about once every 7 kb and (CA)\textsubscript{n}·(GT)\textsubscript{n} repeats occurred about every 30 kb, implying 400,000 copies of (A)\textsubscript{n}·(T)\textsubscript{n} repeats and 100,000 copies of (CA)\textsubscript{n}·(GT)\textsubscript{n} in the porcine genome. Ellegren (1993b) found that one-quarter of the (A)\textsubscript{n}·(T)\textsubscript{n} repeats were associated with PRE-1 SINEs. Neither Moran (1993) nor Ellegren (1993b) reported any (AT)\textsubscript{n}·(TA)\textsubscript{n} repeats, although Zhang et al. (1995) later identified and mapped an (AT)\textsubscript{n}·(TA)\textsubscript{n} repeat associated with a PRE-1 SINE in the 3' UTR of the soluble angiotensin-binding protein gene.

Winterton et al. (1992) have estimated that there are 65,000–100,000 (CA)\textsubscript{n}·(GT)\textsubscript{n} repeat loci in the porcine genome, from the proportion of positive clones in a genomic library. Using in situ hybridization, they showed the chromosomal distribution of these repeats to be more or less uniform across all chromosomes, except that they were under-represented in the centromeric regions, the q arm of the Y chromosome, the telomeres, 16q21, which is a C band positive interstitial site, and the NOR region on 10p. Thus (CA)\textsubscript{n}·(GT)\textsubscript{n} microsatellite repeats seem to be less common in regions containing other categories of repeated DNA.

The origin of (A)\textsubscript{n}·(T)\textsubscript{n} repeats in mammalian genomes can be easily explained in terms of reposition events, which incorporate poly(A) tails into the chromosomal DNA (Ellegren, 1993a), although it is far from certain that all (A)\textsubscript{n}·(T)\textsubscript{n} mononucleotides are generated in this way. The gradual decay of poly(A) tails can explain the origin of at least some of the less common related repeats found, and clearly is responsible for the (AT)\textsubscript{n}·(TA)\textsubscript{n} repeats mapped by Zhang et al. (1995). However, when it comes to the more common dinucleotide categories, such as (CA)\textsubscript{n}·(GT)\textsubscript{n} and (GA)\textsubscript{n}·(CT)\textsubscript{n}, it is unclear whether this generalization can be extended further. In support of the evolution of (CA)\textsubscript{n}·(GT)\textsubscript{n} repeats from degenerate poly(A) tails is the frequent observation by many groups of the contiguity in the porcine genome of the SINE element, PRE-1 and (CA)\textsubscript{n}·(GT)\textsubscript{n} repeats, as mentioned previously. Indeed, Alexander et al. (1995) discovered two new elements, ARE-1P and ARE-2P, in the process of examining the sequences flanking their other (CA)\textsubscript{n}·(GT)\textsubscript{n} sequences, finding them present in 6% of cases. Assuming that these are degenerate poly(A) tails, this means that either there is a very large number of heterogeneous uncharacterized SINE elements, for which no consensus sequence has been derived or, alternatively, that the other 70% of (CA)\textsubscript{n}·(GT)\textsubscript{n} repeats have evolved by some other mechanism. Because replication slippage is widely accepted as the mechanism for the generation of diversity of microsatellite length, it is conceivable that any non-coding sequence containing two or more dinucleotide repeats is capable in theory of growing via this mechanism to the size where it would be recognized as a microsatellite locus. However, the lesser propensity of small repeats to generate new alleles would argue against this hypothesis.

While microsatellites are still very valuable genetic markers for certain applications, such as analyses of population structure (Cowled et al., 2008), they have been usurped by single nucleotide polymorphisms (SNPs) in high-resolution genetic mapping and molecular breeding value estimation, where the availability of SNPs in huge numbers on easy-to-use and relatively cheap genotyping platforms makes them the marker of choice.

**Telomeric repeats**

From very early studies of chromosome behaviour, it was observed that the ends or telomeres of chromosomes had special properties, which
protected them from the degradation and fusion seen with recently broken chromosome ends. It was also realized that conventional DNA polymerases, which require priming with a short RNA fragment, would be incapable of synthesizing DNA right to the ends of discontinuously growing strands, implying that there would be a gradual shortening of the chromosomal ends during chromosome duplication. It is now recognized that mammalian chromosomes are capped by arrays of the hexamer (TTAGGG), (Blackburn, 1991), which can be repeated for up to 10–15 kb in humans and for 100–150 kb in rodents. This repeat is very rare elsewhere in the genome. Interstitial sites generally are points of evolutionary chromosomal rearrangement, such as the telomere to telomere fusion that gave rise to human chromosome 2 (Ijdo et al., 1991) and the interstitial site near the centromere of mouse chromosome 6 (Yen et al., 1995).

The 3′ end of the telomere in mammals is single stranded with an overhang of 50–100 nucleotides. This single-stranded overhang loops around to invade the double-stranded part of the telomere to form a structure called the t-loop (Palm and de Lange, 2008), which hides the end of the telomere and protects it from damage. However, the t-loop alone is insufficient protection, and a number of proteins also bind to the telomere to protect it from nucleases and to distinguish it from a broken chromosome end. Three of the proteins of the six-protein shelterin complex, Telomere Repeats binding factors 1 and 2 (TRF1 and 2) and Protection of Telomeres 1 (POT1), specifically recognize and bind to TTAGGG, with TRF 1 and 2 binding first and then recruiting the other three factors (Palm and de Lange, 2008).

In germline cells, the telomere repeats are synthesized by a distinct replicative mechanism from normal genomic DNA, using a ribonucleoprotein reverse transcriptase enzyme called telomerase (Morin, 1989). Telomerase contains RNA complementary to the telomeric repeats which it synthesizes, and thus is not dependent on the telomeric DNA template. Telomerase is not expressed in somatic cells in mammals. Consequently, the duplication of chromosome ends by conventional polymerases will lead to a reduction in the number of copies of the telomeric repeats, and it has been observed that the number of copies of telomeric repeats decreases as an animal ages. However, attempts to use telomere length to estimate the age of animals have generally been disappointing.

Meyne et al. (1989) demonstrated the conservation of the (TTAGGG), telomeric sequence among 91 species of vertebrates, including fish, amphibians, reptiles, birds and mammals, by in situ hybridization of the repeat to the telomeres of all chromosomes in all the species examined, although the pig was not among the 70 mammalian species tested. Subsequently, Gu et al. (1996) have demonstrated, using primed in situ synthesis (PRINS), that this sequence resides at the telomeres of all porcine chromosomes. They also found an interstitial site at chromosome 6 q21–q22, at a point where there is a disjunction in the syntenic relationship between pig 6 and human 1p and 19q chromosomes. This site of hybridization gradually disappeared with increasingly stringent conditions of hybridization of the primer, implying that the internal site is non-functional and has undergone gradual mutational decay. Gu et al. (1996) interpreted this interstitial position as evidence for a tandem fusion in the origin of porcine chromosome 6.

**Expressed repetitive sequences**

Mammals, like other organisms, contain some repeated sequences that are expressed and have important cellular functions. In some cases, the repetition of the gene is necessary to ensure adequate levels of expression of the gene product. In other cases, duplication of the gene locus has permitted evolutionary experimentation as the duplicate is able to diverge and evolve new functions. Infrastructure RNAs, such as rRNAs, are an important class of expressed repeats.

**Ribosomal RNA (rRNA)**

rRNA is a structural component of ribosomes and thus comprises a vital component of the cellular protein synthesis apparatus. In eukaryotes, the ribosome consists of a small subunit
Molecular Genetics

(40S) containing the 18S rRNA and a large subunit (60S) containing the 28S, 5.8S and 5S rRNA subunits. The 28S, 18S and 5.8S genes are clustered and expressed coordinately. Cells must express rRNAs at extremely high levels to meet the cellular demand for protein synthesis, and have a special enzyme, RNA polymerase I, dedicated specifically to this task. rRNA generally comprises about 80% of the RNA found in a cell. Multiple copies of the rRNA genes expedite transcription. In humans, there are 150–300 copies of the 18S/5.8S/28S rRNA cluster distributed between five autosomes, with the number of copies varying between individuals and chromosomal sites. The positions of the tandemly repetitive ribosomal gene clusters, which are separated by untranscribed spacers, are cytologically visible at features called secondary constrictions or nucleolus organizer regions (NORs). Their expression is controlled by a promoter that directs transcription of all members of the cluster as a 45S pre-rRNA transcript, which is processed into the 18S, 5.8S and 28S subunits.

In the pig, three sites of the rRNA 18S/5.8S/28S cluster, called RNR1, RNR2 and RNR3, are located on chromosome 8p1.2, 10p1.2–1.3 and 16q2.1 respectively (Czaker and Mayr, 1980; Miyake et al., 1988; Popescu et al., 1989; Bosma et al., 1991).

The 5S (about 120 bp) rRNA genes are also tandemly repeated and separated by untranscribed spacers. However, in humans, they are transcribed by RNA polymerase III and map to quite different locations from the 28S/18S/5.8S cluster, having a major site on human chromosome 1q42.11–42.13, with a minor site nearby at 1q31 (Lomholt et al., 1995a). Likewise the 5S rRNA genes in pigs map to 14q23 (Lomholt et al., 1995b), nowhere near the other rRNA genes.

Ling and Arnheim (1994) have characterized the promoter region of the 18S/5.8S/28S cluster in the pig and compared it with equivalent mammalian promoters. This revealed conservation of the regulatory regions found in other mammalian promoters. The only unusual feature was that there was a C nucleotide 16 bp upstream of the transcription start point in the pig, whereas G is conserved in all other eukaryotic promoters.

Transfer RNAs (tRNAs)

Like 5S rRNA, tRNAs are transcribed by RNA polymerase III. As for rRNAs, the very high cellular demand for tRNAs can be met only by transcription from multiple copies of the genes. Generally there are ten to several hundred genes for each tRNA. Interestingly, the repeat clusters contain genes for several different tRNAs. For example, in the rat, tRNA\textsuperscript{Asp}, tRNA\textsuperscript{Gly} and tRNA\textsuperscript{Glu} form a cluster, which is tandemly repeated about ten times. The members of the tandem repeats are not identical, and some may be pseudogenes. Very little has been reported about the tRNA genes of the pig, but it can be safely assumed that they follow a typical mammalian pattern of dispersed distribution and duplication which will be fully revealed by a complete genome sequence.

Other multigene families

In mammals, there are many protein-encoding gene families in which the members show developmental or tissue specificity of expression. For example, the globin gene family consists of members coordinately expressed at embryonic, fetal and adult stages, and includes pseudogenes. The \(\alpha\)-actin family consists of members whose expression is specific for skeletal muscle or for cardiac muscle. In some cases, the members of these gene families are located very closely in the genome, whereas in other more heterogeneous gene families, the members may be dispersed. The individual members of these gene families are generally structurally similar to the single-copy genes discussed later.

FUNCTIONAL GENES.

In the pig, numerous multigene families have been recognized. For example, the tumour necrosis factor alpha (TNFA) and tumour necrosis factor beta (TNFB) loci are located within the MHC and separated by a distance of only 2.5 kb (Chardon et al., 1991). ATP\textsuperscript{1A1} and ATP\textsuperscript{1B1}, the alpha 1 and beta 1 polypeptides of the Na\textsuperscript{+}, K\textsuperscript{+}-ATPase, also comprise dispersed members of a putative gene family on chromosome 4 (Lahbib-Mansais et al., 1993), which is conserved in humans and mice.
The casein family of milk proteins has been thoroughly characterized in the pig and retains a similar structure to that found in other mammals. Four members, namely α S1 (Alexander and Beattie, 1992a), α S2 (Alexander et al., 1992), β (Alexander and Beattie, 1992b) and κ (Levine et al., 1992), have been mapped to chromosome 8 and the first three at least have been shown to be very tightly linked (Archibald et al., 1994).

Two very well characterized porcine multigene families, namely the protease inhibitors (PI) and the immunoglobulin heavy chain (IGH) gene families, coincidentally map near each other on porcine chromosome 7. PI1, PO1A, PO1B, PI2, PI3 and PI4 form a tight linkage group adjacent to IGH1, IGH2, IGH3 and IGH4 (Gahne and Juneja, 1986; Juneja et al., 1986; Van de Weghe et al., 1987; Vogeli et al., 1987; Stratil et al., 1990; Cizova et al., 1993) on 7q24–26 (Gu et al., 1994; Musilova et al., 1995). PO1A and PO1B, originally called postalbumin 1A and 1B, respectively, appear to be cysteine protease inhibitors (Gahne and Juneja, 1986). PI1, PI2, PI3 and PI4 are serine protease inhibitors, with different loci and even alleles at particular loci being identified as trypsin or chymotrypsin inhibitors. Stratil et al. (1990) noted that there are many weakly staining alleles and null alleles at the PI3 locus, which is consistent with the evolutionary expectation within a multigene family containing redundant members.

With increased scrutiny of the porcine genome, more gene families are being discovered and soon all will be revealed by a complete genome sequence. An interesting example is provided by the kallikreins, which are serine proteases expressed in a diverse range of tissues, and are often involved in activating proteins from an inactive precursor. In humans, there are 15 members of this gene family mapping to HSA 19q13.3–q13.4. In the mouse there are even more, between 24 and 37, of which 14–26 are pseudogenes, all co-located on MMU7. Fernando et al. (2007) have reported 13 members of this family in the pig, mapping to SSC6q12-21. Orthology could be determined for all members of the family between human and pig, except for KLK2 and KLK3, which are absent from the pig and also from the mouse. The organization of the members of the gene family is similar across these three species, with all genes in all species oriented in the same transcriptional orientation from telomere to centromere.

miRNA families (miR) constitute a totally novel class of gene families. These families consist of members that differ by only one or two nucleotides, but given the small size of miRNAs, at about 22 nucleotides, there is limited scope for variation. Reddy et al. (2009) reported evidence in the pig for all three members of the miR-1 family, miR-1a, miR-1b and miR-1c, which are present in a diverse range of species. Evidence was found also for all ten members (a to j) of the let-7 family in the pig, although there is some nomenclatural confusion for miRNAs, as miR-98, reported by Reddy et al. (2009), differs by only one nucleotide from the let-7 family and should really be considered a member of it.

PSEUDOGENES. Pseudogenes can arise from duplication of a locus followed by an inactivating mutation, such as a frameshift, the incorporation of a stop codon or the loss or mutation of promoter elements. Like the processed pseudogenes already described, these duplicate elements are no longer subject to adaptive evolutionary constraint and diverge relatively rapidly from the parental sequences. Such divergent, non-functional loci are frequently recognized as ‘degenerate’ members of gene families.

Vage et al. (1994) have reported a pseudogene within the porcine MHC. The MHC class II DRB locus was found to have an expressed version, called DRB1, and a pseudolocus, called DRB2, present in about 60% of chromosomes. DRB2 has a single nucleotide deletion in codon 54, which would cause a frameshift and total loss of function. Ironically, the strong selection pressure for MHC diversity has meant that the functional DRB1 was found to be much more variable than the pseudogene. Brunsberg et al. (1996) have confirmed the findings of Vage et al. (1994) and have discovered an additional DRB pseudogene. Both were detected using primers that span intron–exon boundaries. Thus, the pseudogene is clearly a degenerate member of a gene family. Similarly, Mege et al. (1991) have described a pseudogene within the interferon family.
which they called psi IFN-alpha II-1, which has multiple frameshift mutations. More recently, Knorr et al. (2007) have described RPSAP2, a non-processed pseudogene of the ribosomal protein SA (RPSA), believed to have arisen by segmental duplication. This pseudogene, which has introns unlike the processed pseudogenes of RPSA described earlier, has seven indels, 25 substitutions and ten premature stop codons compared with the RPSA coding sequence. It also has seven SINE (tRNA-Glu) and three DNA transposons inserted into its UTRs and introns. Clearly the evolutionary experimentation following gene duplication can lead in some cases to functional novelty, but is probably more likely to lead to loss of function and decay of the gene.

**Copy number variants and segmental duplication**

Although duplications or deletions of relatively large segments of chromosomes have been known ever since the Bar eye phenotype in Drosophila was attributed to a cytologically visible duplication, genome-sequencing projects have revealed that copy number variants (CNVs), frequently attributable to segmental duplications, are common in the genomes of many organisms and may have important phenotypic consequences. Fadista et al. (2008) have addressed the occurrence of CNVs on parts of SSC4, 7, 14 and 17, using a preliminary assembly of parts of the porcine genome sequence, to gauge the frequency of CNVs greater than 1 kb in a sample of 12 boars. They identified 37 validated CNVs, with five overlapping segmental duplications and several others overlapping known genes, and putatively reflecting segmental duplications and several others overlapping known genes, and putatively reflecting gene family expansion or deletion. Use of a more complete and better assembled genomic sequence and sampling of more animals will obviously reveal many more of these features in the porcine genome. Fadista et al. (2008) estimated that about 0.18% of the porcine genome is tolerant of copy number variation. Du et al. (2009) have found evidence that segmental duplications in pigs have disproportionately affected genes involved in lipid metabolism, putatively related to the history of selection for fatness in pigs.

**Single-copy DNA**

Many single-copy protein-encoding genes have now been cloned and characterized in the pig, and it is well beyond the scope of this review to attempt to cover all of them. Some will be discussed in other chapters. Porcine genes appear in all respects quite typical for mammalian genes. The porcine calcium release channel (CRC) locus, which was the subject of an intensive search by several laboratories, has an extremely large message of over 15 kb (Fujii et al., 1991), but so also do the human and rabbit homologues (Zorzato et al., 1990). The size of the entire genomic copy of this locus, including introns, has not been determined, but is likely to be about 120 kb, given that 25,780 nucleotides of genomic DNA are required to specify only 3305 nucleotides of the porcine message from nucleotide 4624 to 7929 (Leeb et al., 1993). Protein-encoding sequences account for only 13% of the genomic DNA in this region, again emphasizing the large proportion of the genome that is non-coding and of unknown function, even in the vicinity of coding genes. Seemann et al. (2007) have reported the detection of 18,621 protein-encoding RNAs in the Sino-Danish PigEST database, which is likely to be only slightly short of the number of protein-encoding genes in this species.

**Gene Structure and Function**

**Protein-encoding genes**

The porcine growth hormone gene demonstrates the structural and functional features of a typical protein-encoding gene. The porcine gene retains extensive sequence similarity with the human, rat and bovine homologues, even in the non-coding promoter, intronic and untranslated regions (Vize and Wells, 1987). It is instructive to review the regulatory and structural features of this locus in the sequence of Vize and Wells (1987). Many regulatory features held in common with other mammalian genes can be seen. Starting at position −30 from the cap site is the sequence TATAAAA,
the Goldberg-Hogness box, required for correct positioning of RNA polymerase II and initiation of transcription. An adenine nucleotide at position +1, the cap site, marks the start of the messenger RNA, which is capped by the addition of 7-methylguanylate in an unusual 5′–5′ linkage by a capping enzyme. This capping protects the RNA from degradation. The initiating methionine codon, ATG, which marks the start of translation of all eukaryotic genes, commences at position +64. Thus the porcine growth hormone message has a 5′ UTR (untranslated region) of 63 nucleotides. Only ten coding nucleotides occur before the start of the first intron. It is virtually universal for the splice donor site, that is the first two nucleotides of an intron, to be GT, but here the 242 bp intron commences with GC, which is very unusual. However, intron 1 terminates with a typical splice acceptor dinucleotide, AG. Exon 2 is 162 nucleotides long. Intron 2, which is 210 nucleotides long, has typical splice donor and acceptor sites, namely GT and AG, respectively, as do all the other introns in this gene. Exon 3 is 117 nucleotides in length and is followed by intron 3 of 197 nucleotides. Exon 4 is 162 nucleotides in length, followed by intron 4 of 278 nucleotides. Exon 5 consists of 198 nucleotides that specify amino acids followed by the translation STOP codon, TAG. Following this coding sequence is a 3′ UTR of 102 nucleotides. The sequence AATAAA, which is a polyadenylation signal, is 80 nucleotides after the end of the stop codon. This directs an enzyme called poly(A) polymerase to add a poly(A) tail to the end of the message at a point 16 nucleotides downstream from the end of the signal. The addition of the poly(A) tail is not template determined and is essential for the stability of the mRNA. Of the 1745 nucleotides from the cap site to the start of the poly(A) tail, there are 1092 nucleotides of non-coding DNA in UTRs and introns, comprising 63% of the sequence. Although some of this non-coding DNA has recognized function, for example the polyadenylation signal, it is worth emphasizing again that, as for other loci in mammals, only a relatively small proportion specifies amino acids, and this proportion becomes even smaller if all sequence from the start of the 5′ to the end of the 3′ regulatory sequences are considered.

Of course, there are other 5′ and 3′ regulatory sequences that are essential for the tissue-specific and temporally regulated expression of any gene, such as growth hormone. In general, these include positive and negative proximal promoter elements, which are binding sites for transcription factors, which either assist or hinder directly or indirectly the binding of RNA polymerase II. Additionally, there are enhancer elements, which are large sequences, often quite distant, and either upstream or downstream from the gene, which increase the level of transcription, and silencer elements, which are short, often repeated, sequences that downregulate expression.

Somatically rearranged genes

Unlike conventional genes, the immunoglobulin and T cell receptor loci show extensive patterns of somatic rearrangement as a normal part of the diversity-generating mechanism, which enable the huge range of immune responses found in mammals. In essence, gene families of constituent parts at these loci are assembled by DNA breakage and joining events into more-or-less unique combinations within committed lineages of immune cells. The control of the transcription of these newly assembled functional units and the processing of transcripts occurs as for conventional genes.

The antibody molecules produced by B cells comprise light chains fitting into two major categories, \( \lambda \) and \( \kappa \), and heavy chains, consisting of five major types, \( \alpha \), \( \delta \), \( \varepsilon \), \( \gamma \) and \( \mu \), which correspond to the five classes of antibodies, IgA, IgD, IgE, IgG and IgM. The light-chain families map to two locations in human and mouse, whereas the five heavy-chain ‘loci’ map to a single location, which, as mentioned previously, has been identified in the pig. To make a \( \lambda \) or \( \kappa \) light chain, a B cell must undergo a process of combinatorial (V-J) joining in which a unique combination of four exons is assembled by chromosomal breakage and joining events. These exons are called L (leader) and V (variable), of which there are several hundred adjacent but different pairs, J (joining), of which there are four to five variants, and C (constant), of which there is normally one copy. Many hundreds of different
light chains can be specified in this way. Heavy-chain diversity is generated in a similar but slightly more complex way. As for light chains, there are about 250 different L and V exon pairs, followed by about 12 D (diversity) exons, followed by about four J exons, followed by multi-exonic constant regions corresponding to all of the major classes and subclasses of antibodies. The construction of a functional heavy-chain gene requires breakage and rejoining events to assemble a unique combination of L-V-D-J-C coding regions via two events, namely D-J joining followed by V-DJ joining. Again, many hundreds of different heavy chains can be produced in this way, which in combination with the many hundreds of different possible light chains yields great antibody diversity.

In T cells, which are responsible for cell-mediated immune responses, the antigen receptors are composed of α, β, γ and δ subunits, which are encoded at different chromosomal locations. The normal adult receptor is a δβ dimer in humans and mice, but γδ dimers are more prominent in pigs and are found on up to half the T cells. The functional copies of these genes are assembled by a joining mechanism similar to immunoglobulin chains, with α and γ chains having V and J segments, and β and δ having VDJ segments, in addition to the constant segments. Thus, both for antibodies and for T cell antigen receptors, somatic rearrangements are a normal part of the cycle of expression of these gene products, necessitated by the enormous diversity of products required and the impossibility of encoding all possible products in a conventional fashion.

Although most characterization of these immune gene systems has been performed on mice and humans, many elements of the process have been verified in pigs. Lammers et al. (1991) have sequenced most of the V and all of the J and C regions of the porcine κ and λ light chain genes and have found that the J region is more closely conserved across species than the constant regions. The porcine κ light chain is unusual in that the constant region terminates with the dipeptide Glu-Ala after the Cys residue that terminates all other mammalian light chains. Kacskovics et al. (1994) have identified five putative subclasses of porcine IgG loci (1, 2a, 2b, 3 and 4) on the basis of C-region sequences from a single animal. Brown and Butler (1994) sequenced the constant region of the porcine IgA heavy chain and have shown that unlike the human and the mouse, which have two or more loci, the pig has only a single C gene, with the sequence and structure conserved relative to other species particularly towards the carboxy terminus of the gene product. Bosch et al. (1992) have similarly characterized a porcine clone for the μ heavy chain. Sun et al. (1994) have cloned and sequenced 65 H chain V regions expressed with IgG, IgA or IgM C regions. They found that the porcine V regions belonged to a homogeneous family with greater than 80% nucleotide identity among members, and that there were probably fewer than 20 copies of the V regions in the pig. Surprisingly, the J region was also very homogeneous, suggesting that a single J region is preferentially used. In other mammals, there are three families of V genes, with mice having 100 to 1000 copies, humans 100 to 120 copies and rabbits 100 copies. The relatively depauperate V status of pigs raises questions about how they can generate sufficient antibody diversity.

The cDNA clones of the porcine T cell receptor α, β, γ and δ chains have been characterized (Thome et al., 1993; Grimm and Misfeldt, 1994). Thome et al. (1993) found evidence for at least three different constant segments for the porcine γ chain family. There is substantial conservation of constant segment coding sequence within the four subunits across pigs, cattle, mice, sheep and humans.

Non-coding RNAs (ncRNAs)

Following the discovery of very large numbers of ncRNAs in the mouse transcriptome (The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team et al., 2002), it is now recognized that most of the genome in mammals is transcribed in a controlled manner. Indeed, transcription from non-coding sequences is at least four times greater than from coding sequences (Mercer et al., 2009), although individual protein-coding transcripts are expressed at a much higher level than non-coding transcripts. The approximately 34,000 or so long non-coding
but messenger-like RNA molecules that have been discovered in the mouse will presumably have orthologues in other mammalian species, and perhaps should be thought of as a large new class of genes to which we need to ascribe function. Opinions vary as to the proportion that is functional. However, those that are expressed in a highly regulated way, with developmental and tissue specificity and even precise subcellular location (Mercer et al., 2009), clearly meet the criteria for functionality, despite the fact that there is generally poor sequence conservation of these elements between species. Pheasant and Mattick (2007) have argued that it is quite possible that the majority, and possibly all, of these sequences are functional. The database RNAdb (Pang et al., 2005) catalogues such ncRNAs from various species. It is believed that ncRNAs generally function as trans-acting regulatory factors.

Seemann et al. (2007) have searched for ncRNAs in the Sino-Danish PigEST resource, and identified over 1000 of them. Consistent with findings in other mammals, they found that ncRNA transcripts were much more abundant in brain tissue than in other less developmentally complex tissue.

**Micro RNAs (miRNAs)**

miRNAs are short regulatory RNA molecules that are about 22 nucleotides in length. They can regulate gene expression by interfering with translation or causing degradation of target mRNA molecules, which they do by base-pairing with the target molecule, most frequently in the 3′ UTR. miRNAs arise from longer transcripts and are trimmed to size. The precursor pri-miRNA is from several hundred bp to 1000 bp long and is capped, polyadenylated and spliced, and may actually incorporate a number of different miRNA molecules that are released following processing. The pri-miRNA contains a number of hairpin loops of about 70 nt in length, which are cut and then processed into functional miRNA. He et al. (2008) have identified miRNAs encoded by long ncRNAs, estimating that up to 800 are present in the mouse ncRNA population.

Most miRNAs are initially reported on the basis of bioinformatic scans, but given the small size of miRNAs these scans are susceptible to false discovery. Wernersson et al. (2005) identified 51 miRNA sequences in their partial porcine genome sequence by similarity to miRNAs in closely related species, but this necessarily excludes miRNAs that are specific to the pig. Kim et al. (2006) subsequently reported an additional nine miRNAs identified by homology search. Reddy et al. (2009) found 120 miRNAs using 454 sequencing. Comparisons with known miRNAs in miRBASE showed that 24 were previously recognized porcine miRNAs and 96 novel miRNAs. They confirmed expression of 22 conserved miRNAs and four pig-specific miRNAs in a range of pig tissues. Huang et al. (2008) identified 775 miRNA candidates in silico in the partial porcine genome sequence, of which 296 gave positive signals on a cross-species miRNA microarray using developing skeletal muscle, but only five of these were validated by real-time PCR; 255 of the 296 had not been previously reported. McDaneld et al. (2009) have also examined the profile of miRNA expression in developing porcine skeletal muscle, reporting 12 novel potential miRNAs. Sharbat-Tehrani et al. (2008) applied a novel concatameric cloning method to identify ten porcine miRNAs, of which nine were novel. Anselmo et al. (2009) identified large numbers of miRNAs in porcine cells infected with the Alphaherpesvirus responsible for Aujeszky’s disease, of which about 95% were of host origin. They were particularly interested in identifying miRNAs of viral origin that regulated host gene expression in the porcine cells, as Herpesviruses have been shown to produce such miRNAs. Tomas et al. (2009) identified 175 potential miRNAs of viral origin in porcine cells infected with Aujeszky disease virus, along with about 200 miRNAs of host origin. Feng et al. (2007) mapped six evolutionarily conserved porcine miRNA loci using a radiation hybrid panel, demonstrating that the porcine map locations were consistent with human positions for each miRNA, thus providing an alternative or at least complementary validation tool to expression analysis. miRBASE (Release 14) lists 721 confirmed miRNAs for humans and only 77 for the pig, implying perhaps a degree of scepticism about the published porcine miRNAs, but also demonstrating that there is ample scope for further discovery of porcine miRNAs.
Of course, the discovery of the miRNA is only the first step. The next important step is discovering the target for miRNA binding. If identifying the miRNAs themselves is challenging enough using bioinformatic approaches, because of the small size of the mature miRNA molecules, the challenge for finding the targets is much greater. Perfect complementarity is only required for the first six to eight nucleotides at the 5′ end of the miRNA and the 3′ end of the target and, in some cases, even this perfect match is not absolutely necessary. Liu et al. (2010) have pointed out that algorithmic methods for detecting miRNA targets have false discovery rates of up to 30%. Allelic variation in miRNA target sites has been associated with phenotypic effects in domestic animal species, although none have been verified in pigs. Clop et al. (2006) reported that a muscular hypertrophy phenotype in Texel sheep was the result of the creation of an illegitimate target for miR-1 and miR-206 in the 3′ UTR of the myostatin locus. Davis et al. (2005) have implicated maternal expression of five different miRNAs in the callipyge muscular hypertrophy phenotype in sheep. Cargill et al. (2008) have identified a mutation in the 3′ UTR of synaptogjanin that could disrupt let-7 and miR-98 binding sites, and is possibly responsible for the polled trait in cattle.

miRNAs have also generated enormous interest as it is possible to design inhibitory RNA molecules to mimic the inhibitory activity of the naturally occurring miRNAs, but with the possibility of targeting almost any molecule. As described previously, such constructs have been used to genetically modify pigs to suppress the expression of porcine endogenous retroviruses, but clearly they can be used in a much broader context.

**Base Composition and Methylation**

Protein-encoding genes in humans are found preferentially in GC-rich isochores. The GC content of human exons also tends to be about 10% higher than in introns. Wernersson et al. (2005) have found that the GC content and distribution in the pig are very similar to those in humans, as discussed above, with a 10% higher GC content in coding regions compared with introns. These findings have confirmed the results of earlier glimpses of porcine GC content. For example, the GC content of ten of the microsatellite-containing clones (excluding the repetitive microsatellite component) characterized by Robic et al. (1994), none of which is known to encode any function, was 43% (1607 from 3768 nucleotides), also consistent with the expectation that porcine non-coding regions would be GC poor. In contrast, the growth hormone locus has a genomic DNA GC content of 61% and a cDNA GC content of 59% GC. Again the expectation of high GC content in coding DNA is met. However, not all porcine coding genes have a high GC content. The genomic sequence of the relaxin gene, containing about 1 kb of 5′ flanking non-transcribed sequence, about 300 nucleotides of 3′ untranscribed sequence and a single, very large intron of 5495 nucleotides (Haley et al., 1987), is only 36% GC, and this figure rises to only 40% for the relaxin cDNA.

Variation in nucleotide composition can have more subtle manifestations than simple variation in nucleotide content. The cytosine of the dinucleotide CpG is normally methylated throughout vertebrate genomes as part of a transcription suppression mechanism, except in clusters of this dinucleotide forming so-called CpG islands, which somehow escape methylation and allow expression of the adjacent gene. The methylated CpG is particularly prone to mutation to TpG. In all vertebrates examined, the CpG dinucleotide is about five times less frequent than would be predicted on base composition, because of this mutational pressure.

**Number of Expressed Genes in the Pig**

When the first edition of this chapter was written, no estimate had been made of gene number in the pig, nor indeed had any accurate count of gene number been made for any species. The generally accepted figure for vertebrates at that time was about 50,000–100,000, regardless of variation in genome size. Antequera and Bird (1993), on the basis of there being an estimated 45,000 and
37,000 CpG islands in the human and mouse genomes, respectively, and that 56% of human genes and 47% of mouse genes are preceded by CpG islands, came up with an overall estimate of gene number in both species of about 80,000. Numerous other estimates of gene number were bandied about for humans in particular, and mammals in general, going as high as 150,000. These inflated numbers were brought down to earth with a thump with the completion of the first human genome sequences, which clearly showed that the number of protein-encoding genes, plus the already recognized infrastructural RNAs, such as rRNAs and tRNAs, was about 20,000. This number was subsequently verified for all other mammals for which complete sequences became available. The current best, but certainly downwardly biased, estimate of gene number in the pig is 18,621 (Seemann et al., 2007) and is based on the content of the PigEST resource database in 2007. It is to be expected that this will drift slightly upwards towards the generally accepted value for all mammals as the final elements of the porcine genome sequence come to light.

Ironically, the more recent recognition that almost the entire genome is transcribed, and that there are up to 34,000 non-coding but quite possibly functional RNAs in mammals, means that both the concept of gene number and the actual gene count require serious rethinking. Crudely, one might wish to simply define the ncRNAs as originating from genes and add their number to the count for protein-coding sequences, giving an overall figure of greater than 50,000. However, this is a gross oversimplification, ignoring the complexities of the interleaved model of genome structure. What do you count as a gene, for example, when a mRNA is transcribed from one strand of DNA and a miRNA is transcribed from the other and regulates expression of the protein encoded by that mRNA?

Mitochondrial Genome

Mitochondria are cellular organelles believed to be derived from ancient intracellular bacterial symbionts. In addition to their unique role in synthesis of ATP (oxidative phosphorylation), these organelles have a very peculiar genetic system. They have an autonomously replicating DNA genome, of about 16.5kb in mammals, but the mechanism of replication is quite unlike that for nuclear DNA. Replication commences with initiation of heavy (H)-strand synthesis at a specific origin and results in the formation of a displacement (D) loop with a newly synthesized H strand of about 680 bases, with very few H strands achieving full length. Initiation of light (L)-strand synthesis occurs from a specific origin but only after this region has been exposed by H-strand synthesis. Transcription is initiated from promoters situated within the D-loop region, with the primary transcripts processed to give 12S and 16S rRNAs, tRNAs and a number of mRNAs which are not capped but are polyadenylated. Virtually all of the DNA in the mitochondrial genome is functional, specifying rRNA, tRNA or mRNA, with very few, and in some cases no, non-coding nucleotides between adjacent genes. The D loop does not encode any known function.

Most proteins in the mitochondria are nuclear encoded but some are mitochondrialy encoded and translated by a separate system in which the rRNAs and tRNAs are also mitochondrialy encoded. In the mammalian mitochondria, proteins are synthesized using a unique mitochondrial code, which employs different initiation and stop codons from the universal code, and messages are decoded by a different mechanism from that used by nuclear genes. This decoding mechanism uses a minimal set of only 22 tRNAs with different wobble anticodons from those employed in nuclear protein synthesis. Formyl-methionine is used to initiate a peptide, as in prokaryotes. However, mammalian mitochondrial rRNA has only low homology with prokaryotic rRNA.

The complete mitochondrial genome sequence has been available for several mammalian species, including human (Anderson et al., 1981), cattle (Anderson et al., 1982), mouse (Bibb et al., 1981) and rat (Gadaleta et al., 1989), for many years. The first complete mitochondrial sequences for pigs became available in 1998 (Ursing and Arnason, 1998; Lin et al., 1999; Kijas and Andersson, 2001). Earlier mitochondrial sequence analyses for the pig were restricted to the cytochrome
b gene (Irwin et al., 1991; Honeycutt et al., 1995) and the cytochrome c oxidase II gene (Honeycutt et al., 1995), where the porcine sequences were generated as part of a large phylogenetic study. Additionally, the D loop region (Ghivizzani et al., 1993) of the pig has been extensively characterized. The porcine D loop is much larger than in the cow and, surprisingly, displays greater organizational similarity to the mouse and human D loop than to the bovine D loop. Within the porcine D loop, there is a ten-nucleotide repeat, with copy number ranging from 14 to 29, giving rise to the highest level of mitochondrial heteroplasmy documented in mammals, apparently as a result of replication slippage in the tandem repeat. Mitochondrial sequences have been used extensively for phylogenetic studies of Sus scrofa and its relatives (Giuffra et al., 2000; Kijas and Andersson, 2001; Okumura et al., 2001; Kim et al., 2002; Gongora et al., 2004; Larson et al., 2005; Lucchini et al., 2005), revealing among other things the separate domestication of pigs in Europe and Asia.

Nuclear mitochondrial sequences (numts) comprise an interesting class of pseudogene. In humans, there is an unusually large number, estimated at 350–612 copies comprising about 0.012% of the genome, which have been transferred into the nuclear genome (www.pseudogene.net). Domestic cats also have large numbers of numts, between 38 and 76. Undetected numts can compromise phylogenetic analyses of mitochondrial DNA (mtDNA) sequences as their evolutionary history will be quite distinct from the conserved functional mtDNA, although phylogenetic signals can be obtained from the presence or absence of numts in some lineages. Numts have not been reported in S. scrofa, although Funk et al. (2007) found evidence for a cytochrome B numt in the pygmy hog, Sus (Porcina) salvanius.

Conclusions

This review of the molecular genetics of the pig has aimed to cover studies specific to the pig, but to put this information in the broader context of the genomic structure and function common to all mammals. It is very clear that the pig, like all mammals, has only a small proportion of its genome of around 2% dedicated to encoding proteins. Formerly, much if not most of the remainder was often dismissed as junk DNA with no function. With the recognition of genome-wide transcription and large numbers of transcriptionally regulated and apparently functional non-protein-coding RNAs, a more sophisticated view of genome structure, gene number and functional genome content is beginning to develop, which increases the challenge of functionally cataloging the entire genome of the pig. We received small relief in this Herculean task with the recognition that the number of protein-coding genes is much smaller than originally expected in the pig and other mammals, but this was largely taken back with the recognition of the huge number of ncRNA genes, most of which remain functionally uncharacterized.

The protein-coding component of the genome of the pig is already almost completely characterized, and this task will soon be complete. The recognition of allelic variants (SNPs) is well under way. The molecular tools for understanding and exploiting genetic variation in economically important traits in pigs are largely in place, but we still have a long way to go in understanding the control of genetic variation in complex traits.

Acknowledgements

I wish to continue my thanks to Lee Alexander, Allan Lohe, David Grimm, Joan Lunney, Dinah Singer, Feng Gu, Leif Andersson, Esther Signer, Marilyn Mennotti-Raymond and Yizhou Chen, Frank Nicholas and Leif Andersson for helpful information and advice on the first edition of this chapter. I wish to thank past and present PhD students, Muladno, Weiyi Zhang, Jaime Gongora, Jun-Heon Lee, Seung Soo Lee, Denbigh Simond, Jaci Aldenhoven, Fabricia Nascimento and Karma Nidup, who have sustained my interest in pig genetic research by their contributions. I acknowledge helpful editorial contributions from Anatoly Ruvinsky and Max Rothschild.
References


Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. and Georges, M.


lation structuring in the rangelands of eastern Australia, applications for designing adaptive manage-

Czaker, R. and Mayr, B. (1980) Detection of nucleolus organizer regions (NOR) in the chromosomes of the


Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M. and Charlier,
C. (2005) RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Current Biology*
15, 743–749.

endogenous retrovirus (PERV) expression by PERV-specific shRNA in transgenic pigs. *Xenotransplantation*
15, 36–45.


Ellegren, H. (1993a) Variable SINE 3′ poly(A) sequences, an abundant class of genetic markers in the pig

Ellegren, H. (1993b) Abundant (A)n.(T)n mononucleotide repeats in the pig genome, linkage mapping of the


Porcine kallikrein gene family, genomic structure, mapping, and differential expression analysis.
*Genomics* 89, 429–438.

Review of Genetics* 41, 331–368.

Genetics* 7, 345–346.

Froegen, E., Thomsen, P., Kristensen, T., Kran, S., Miller, R. and Davies, W. (1991) Porcine SINEs, character-

Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V.K., Weiler, J.E., O’Brien, P.J. and MacLennan,

is a unique genus, 19th century taxonomists got it right first time round. *Molecular Phylogenetics and
Evolution* 45, 427–436.

nucleotide sequence of the *Rattus norvegicus* mitochondrial genome, cryptic signals revealed by

evidence for tight linkage between the structural loci of these inhibitors. *Animal Genetics* 17, 135–137.


Molecular Genetics 99


Introduction

The immune system protects swine against infection and coordinates immune responses to an unknown array of foreign antigens, vaccines and pathogens. Any meaningful attempt to correlate the classical genetics of disease resistance of swine with differences in immune system genetics must consider the genetic origin of innate and adaptive immune receptors (see reviews in Summerfield, 2009). The immune system adds an additional level of complexity to understanding genetics as a result of the differentiation of lymphocytes, i.e. the T cells and B cells, and their receptors. Unlike other somatic cells, lymphocytes utilize a combination of genetic mechanisms, including recombination-activating gene (RAG)
protein-mediated somatic recombination, gene segment rearrangement, junctional diversity and somatic mutation mechanisms to generate a vast repertoire of T cell receptor (TCR) and B cell receptor (BCR) phenotypes with which to sample the antigenic environment (Table 6.1) (Paul, 2008; Barreiro and Quintana-Murci, 2010). Immunoglobulin (Ig) serves as the surface antigen receptor for B cells, i.e. the BCR. Recognition of antigen is through the variable regions of both the heavy and light Ig polypeptide chains that comprise the BCR. Without the use of a combination of genetic recombination mechanisms, the TCR and BCR repertoire (>10^{10} for each) would require more DNA than is available in the entire typical mammalian genome (Venturi et al., 2008). Immune receptors like BCRs and TCRs provide an additional genetic challenge; they are the result of somatic hypermutation (SHM) that is regulated by genes thought to follow Mendelian genetics like most genes that control innate immune receptors.

An equally important genetic element is the major histocompatibility complex (MHC), which is also highly complex, not due to somatic recombination and mutation events, but rather resulting from the large number of allelic variants at many of the >300 loci in the complex (Horton et al., 2004; Kelley et al., 2005; Paul, 2008; Marsh et al., 2010). The human leucocyte antigen (HLA) is recognized as the most variable region in the human genome (Horton et al., 2008). The MHC antigens, in swine the 'swine leucocyte antigens' (SLAs), are vital to the overall immune response because of their role in binding and presentation of foreign antigens to the TCR. The allelic variants at these loci serve to increase the range of foreign antigens that can be recognized and presented to the TCR. The major class I MHC antigens are expressed on all cells, whereas the class II MHC antigens are primarily expressed on the surface of so-called professional antigen-presenting cells (APCs), such as B cells, macrophages and follicular dendritic cells (Schook and Lamont, 1996; Lunney et al., 2009). Other loci in this region serve to process and transport foreign antigens to the MHC molecule for effective presentation to the TCR. Additional 'non-classical' MHC molecules, such as HLA-G (human leucocyte antigen-G), present ligands to natural killer immunoglobulin receptors (KIRs), which themselves are very polymorphic receptors determining innate immune responses (Table 6.1) (Middleton and Gonzalez, 2010). The single pig KIR gene, KIR2DL1, maps to SSC6q where it is linked by LILRA, LILRB and LILRC, members of the leucocyte immunoglobulin-like receptor (LILR) family, innate immune receptors with immunomodulatory functions (Sambrook et al., 2006).

It is sometimes difficult for young animal scientists to understand differences between the genetics of adaptive immune receptors and those controlling simple Mendelian traits. The genetic complexity provided by loci encoding the MHC, TCR or BCR is the basis for the subtitle ‘the Science of Self–Non-self Discrimination’ adopted by Jan Klein in his classic immunology textbook (Klein, 1982). This chapter will focus on the genetics of the porcine MHC, TCR and BCR (Table 6.1), as well as the many genetic loci that regulate innate and adaptive immune responses.

Table 6.1. The genetic diversity of the antigen receptors of the adaptive immune system.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Origin of specificity</th>
<th>Origin of genetic diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Germ line</td>
<td>Somatic</td>
</tr>
<tr>
<td>B cell receptor (BCR)</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>T cell receptor (TCR)</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Major histocompatibility complex (MHC) I</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>MHC II</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Non-classical MHC</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Overview of the Major Antigen Receptors of the Immune System

Innate versus adaptive immunity, essential genetic components

To understand immunogenetics it is first necessary to understand the difference between the immediate, or innate, immune response and the delayed acquired, or adaptive, immune response (Paul, 2008; Barreiro and Quintana-Murci, 2010). Innate responses exhibit broad antigen specificity and have fixed receptors that are the first line of defence against bacterial and viral infections. These receptors recognize ligands, such as bacterial lipopolysaccharide (LPS), bacterial nucleic acids and bacterial flagellin DNA, that are collectively called pathogen-associated molecular patterns (PAMPs). Many have leucine-rich repeat (LRR) domains, which are indispensable for recognition of PAMPs. Most of these innate receptors are not lymphocyte restricted but are found on many cell types, and especially on APCs. The adaptive response is exquisitely antigen specific; the TCRs and Igs are encoded in the germ line; but antigen-dependent responses require gene rearrangements and result in the development of positive memory with enhanced recall responses (Lunney et al., 2004; Paul, 2008). Innate immunity is regulated by effector proteins such as antimicrobial peptides, acute-phase proteins, complement, cytokines and chemokines expressed by monocytes, macrophages, granulocytes, neutrophils and natural killer (NK) cells. Acquired immunity is regulated for each cell (T cell, B cell, APC, dendritic cell (DC)) by cytokines and chemokines, Igs, perforins and granzymes, and regulatory cell subset interactions (Ezquerra et al., 2009; Gerner et al., 2009; Summerfield and McCullough, 2009).

Genetic diversity of innate immune receptors

Based on comparative studies, several families of innate immune receptors were identified, e.g. Toll-like receptors (TLRs), first found as developmentally important for Drosophila (Lemaitre et al., 1996; Medzhitov and Janeway, 2000; Takeda et al., 2003). There are ten or more TLR molecules that recognize complex patterns of molecules derived mainly from microbes; genes corresponding to all ten TLRs have been cloned in pigs (Table 6.2) (Shinkai et al., 2006a; Zhu et al., 2008; Uenishi and Shinkai, 2009). Many TLRs (TLR1, TLR2, TLR4, TLR5, TLR6, TLR10) are localized on the cell surface and recognize compounds derived mainly from microbes: TLR2 senses Gram-positive bacteria through molecules such as peptidoglycans, lipoteichoic acid and zymosan from yeasts or fungi; TLR4 monitors Gram-negative bacteria by recognizing LPS; TLR5 recognizes microbial flagellin. Most TLR molecules form homodimers to transmit signals; however, TLR2 constructs a heterodimer with TLR6 or TLR1 to bind di- and tri-acylglycerides, respectively. The TLRs that are expressed on the membranes of intracellular organelles such as endosomes recognize nucleic acids or derivatives of nucleotides and are thought to be involved in antiviral defence. TLR3 recognizes double-stranded (ds) RNA; TLR7 and TLR8 recognize single-stranded RNA (ssRNA); TLR9 is a receptor for unmethylated CpG DNA, which is generally found in bacterial or viral genomes. No transcript corresponding to murine TLR11, TLR12 or TLR13 has yet been reported. A thorough review of porcine TLRs has recently been published by Uenishi and Shinkai (2009).

Shinkai et al. (2006a,b) reported single nucleotide polymorphisms (SNPs) in genes for TLR1, TLR2, TLR4, TLR5 and TLR6 by screening samples from 11 different pig meat breeds. They detected 136 SNPs in these five porcine TLR genes, which included 63 non-synonymous SNPs, i.e. SNPs causing amino acid substitutions. Bergman et al. (2010) reported that the numbers of SNPs detected in TLR1 and TLR2 were significantly lower in the wild boars than in the domestic pigs. The authors suggest that this may be because of a bottleneck before domestication in the European wild boar population, resulting in less diversity in European versus Asian pigs in both wild and domestic populations (Larson et al., 2005). For TLR1, TLR2 and TLR6 genes, SNPs were disproportionately located in the sequences encoding TLR ectodomains; non-synonymous
SNPs were more likely to be heterozygous than SNPs in other regions (Shinkai et al., 2006b). For TLR1, TLR4 and TLR5 in pigs, cattle and humans, SNPs affecting the coding amino acids are concentrated in the sequences encoding ectodomains, particularly the regions encoding LRRs (Uenishi and Shinkai, 2009). Polymorphisms in TLR3, TLR7 and TLR8 were observed particularly in their ectodomains (Morozumi and Uenishi, 2009). Comparative TLR studies have predicted disease effects (Jann et al., 2009). Numerous infections alter the expression of TLRs, e.g. porcine reproductive and respiratory syndrome virus (PRRSV) infection alters mRNA expression of TLR2, TLR3, TLR4, TLR7 and TLR8 in at least one lymphoid tissue or cells (Liu et al., 2009). Examples of TLR involvement in pig pathogen associated responses are given in Table 6.2.

The NLR (nucleotide-binding domain, LRR-containing) proteins are central regulators of immunity and inflammation; several NLRs activate the inflammasome complex and result in inflammatory cytokine expression, regulation of NFκB (nuclear factor kappaB) activation, etc. (Ting et al., 2010). Jozaki et al. (2009) investigated the functional variance of missense polymorphisms in ligand recognition by the porcine NLR family gene, NOD2. The T1949C polymorphism, located in the region encoding the hinge domain of the molecule, diminished the functional response of porcine NOD2 to muramyldipeptide (MDP), whereas the A2197C polymorphism, in the LRR region, significantly augmented the response of porcine NOD2 to the ligand (Table 6.2). Jozaki et al. (2009) noted that the 1949C allele was rare among pig breeds, suggesting that it is a disadvantage

Table 6.2. Pathogen responses associated with TLRs (Toll-like receptors) and NLRs (nucleotide-binding domain, LRR (leucine-rich repeat)-containing proteins) in pigs. Adapted from Uenishi and Shinkai (2009).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Map locationa (SSC: position in Mb)</th>
<th>Effect on immune response</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLR2</td>
<td>8: 30.3</td>
<td>Increased expression of TLR2 in the intestinal tissues of gnotobiotic pigs after treatment with <em>Escherichia coli</em> or <em>Lactobacillus fermentum</em></td>
<td>Willing and Van Kessel, 2007</td>
</tr>
<tr>
<td>TLR2, TLR6</td>
<td>8: 83.6</td>
<td>Recognition of <em>Mycoplasma hyopneumoniae</em> cell bodies in porcine alveolar macrophages</td>
<td>Muneta et al., 2003</td>
</tr>
<tr>
<td>TLR3</td>
<td>15: 58.9</td>
<td>Activation of TLR3 – increased expression of IFN (interferon) α, decreased PRRSV (porcine reproductive and respiratory syndrome virus) infection of macrophages</td>
<td>Miller et al., 2009</td>
</tr>
<tr>
<td>TLR5, TLR9</td>
<td>10: 14.9</td>
<td>Upregulation after feeding of <em>Salmonella enteritica enteritica</em> serovar Typhimurium or <em>S. e. e</em> serovar Choleraesuis</td>
<td>Burkey et al., 2007</td>
</tr>
<tr>
<td>TLR7</td>
<td>X: 9.3</td>
<td>Enhanced transcriptional activation of TLR-7-induced genes in swine macrophages</td>
<td>Fernandez-Sainz et al., 2010</td>
</tr>
<tr>
<td>TLR7, TLR8</td>
<td>X: 9.3</td>
<td>NK (natural killer) cells activated by TLR agonists cytotoxic against foot-and-mouth disease virus-infected cells</td>
<td>Toka et al., 2009</td>
</tr>
<tr>
<td>TLR9</td>
<td>13: 39.7</td>
<td>Porcine circovirus 2 infection, CpG-TLR9 signalling cytokine inducers and inhibitors</td>
<td>Hasslung et al., 2003; Vincent et al., 2007; Wikstrom et al., 2007</td>
</tr>
<tr>
<td>NOD1</td>
<td>18: 40.8</td>
<td>Immunobiotic lactic acid bacteria</td>
<td>Tohno et al., 2008a,b</td>
</tr>
<tr>
<td>NOD2</td>
<td>NRb</td>
<td>MDP (muramyldipeptide) response</td>
<td>Jozaki et al. 2009</td>
</tr>
</tbody>
</table>

a Jann et al., 2009.
b NR = not reported.
to pig responses to microbes; in contrast, the 2197C allele was widely distributed among Western breeds, thus suggesting a causal relationship between molecular function and polymorphisms in pattern-recognition receptors (PRRs). The authors concluded that the 2197C allele would confer an advantage in modern pig breeds and serve as a useful marker of improved disease resistance.

The structure of other innate receptors is diverse (Fritz et al., 2006; Pancer and Cooper 2006). Innate immune receptors are often ‘adjuvant receptors’ that control the activities of accessory cells of the immune system and thus allow adaptive immune receptors to function (McKee et al., 2010). Among the best known are products of mycobacteria used in Freund’s adjuvant and synthetic deoxyoligonucleotides (ODNs), e.g. CpG-ODN. For example, newborn piglets are unable to mount adaptive immune responses unless they first encounter PAMPs provided by colonizing bacteria or bacterial DNA because CpG-ODN is recognized by innate immune receptors such as TLR9 (Butler et al., 2002, 2005a). Uenishi and Shinkai (2009) predicted the use of TLR ligands as immune modulators, particularly for T and B cell responses. The high frequency of SNP variations among TLR alleles will affect ligand recognition and thus lead to differences in pathogen responses among pig populations.

Overall, the receptors for the innate immune system are largely encoded in the germ line; thus under classical Mendelian genetic inheritance. In any case, whether an immune receptor is innate or adaptive, its function depends on expression and expression depends on a complicated series of germ line encoded regulatory genes. So functional swine genetics will be reflected by a symphony of genetic influences, some somatic, some rigidly germ line, and all of which are under the control of regulatory germ line genes.

**Germ line versus somatic generation of B and T cell receptors (BCRs and TCRs)**

TCRs and BCRs are not encoded in the germ line but are the result of somatic gene rearrangement and, in the case of BCRs, also of somatic hypermutation (SHM), which is mediated by activation-induced cytidine deaminase (AID) (Venturi et al., 2008; Thomas et al., 2009). The somatic processes that control BCR and TCR generation are controlled by nuclear enzymes such as recombination-activating enzyme (coded by RAG), terminal deoxynucleotidyl transferase (Tdt), AID and various constitutive DNA repair enzymes. Except for certain constant region genes, receptor formation and the expression of BCRs and TCRs depend primarily on regulatory genes, not on germ line genes. Thus, one may predict that only the regulator genes controlling the somatic process ascribe to classical Mendelian genetics. However, such a simplistic demarcation is itself an oversimplification (Flajnik and Du Pasquier, 2004). For example, the BCRs of fetal B cells that lead to ‘natural antibodies’, and the invariant γδ TCRs that dominate fetal life might be considered innate receptors. It seems better that the distinction should be based on function, not structure.

In addition to being exclusively germ line encoded, TLRs and other families of innate immune receptors, as well as MHC antigens, are controlled by codominant genes, and both alleles can be expressed on any one cell. This sharply differs from TCRs and BCRs expressed on lymphocytes, for which gene expression is controlled by allelic exclusion, i.e. only the gene products of one allele are expressed on any one T cell or B cell; thus, lymphocytes are monospecific. This evolutionary development is easy to comprehend because a B or T cell with dual receptors might recognize both a dangerous pathogen and one’s own self-antigen.

**The Porcine MHC or Swine Leucocyte Antigen (SLA) Complex**

**Organization of the SLA complex genes**

The swine MHC, the SLA complex, is one of the most gene-dense regions in the swine genome. It consists of three major gene clusters, the SLA class I, class III and class II regions, that span ~1.1, 0.7 and 0.5 Mb, respectively, on chromosome 7 (SSC7p1.1–q1.1), making the swine MHC the smallest
among mammalian MHC so far examined, and the only one known to span the centromere (Fig. 6.1) (Chardon et al., 2000; Lunney et al., 2009). The entire SLA region of the very common Hp-1.1 (H01) haplotype has been mapped and sequenced (Velten et al., 1998; Chardon et al., 2001; Renard et al., 2001, 2003, 2006; Shigenari et al., 2004); a total of >230 loci have been identified, with at least 190 genes predicted to be functional (VEGA Pig, 2010).

The MHC genes are critical for generation of most adaptive immune responses. Initiation of adaptive immune responses to most protein antigens, the thymus-dependent (TD) antigens, requires dual recognition. The BCR recognizes

---

**Fig. 6.1.** Comparative genomic organization of the human and swine major histocompatibility complex (MHC) class I and class II regions. The human leucocyte antigen (HLA) class I and II map is adapted from Horton et al. (2004) and the swine leucocyte antigen (SLA) class I and II map is based only on one fully sequenced haplotype (Hp-1.1, H01) (Renard et al., 2006). Note: not all the genes are shown here and the scale is approximate. The number and location of expressed SLA class I or II genes may vary between haplotypes. Adapted from Lunney et al. (2009).
the intact foreign antigen, but the full-blown B cell response that results in secretion of antibodies requires a second signal delivered from a TCR-dependent T cell response. The TD antigen is processed into peptides that are presented to T cells by MHC class II antigens. The MHC class I antigens bind eight to ten amino acid peptides, generated by intracellular breakdown of self and foreign proteins, and stimulate adaptive cytotoxic (CD8+ T cell) immune responses when they present these peptides to the TCR (Yaneva et al., 2010). The heterodimeric proteins (SLA class I and β2-microglobulin) are constitutively expressed on the surface of virtually all nucleated cells and function mainly in presenting peptides to CD8+ cytotoxic T cells. They also interact with NK cells to prevent NK-mediated cytotoxicity (Kwiatkowski et al., 1999). Non-classical MHC I receptors are less or non-polymorphic receptors and recognize other chemistries, including glycolipids derived from bacteria. The SLA class II antigens are expressed mainly on APCs and function in presenting exogenous peptides to CD4+ helper T cells (Ezquerra et al., 2009).

**SLA class I and II genes**

The SLA map includes seven classical class I genes and three non-classical class I genes (Fig. 6.1a). The overall genomic organization of the SLA class I region is quite different from that of the HLA class I region, yet they share common orthologous segments with strong conservation in the order and organization of the genes. Most differences appear to be due to segments where HLA expansion has occurred (Renard et al., 2003; Kelley et al., 2005). The constitutively expressed classical SLA class I genes are SLA-1, SLA-2 and SLA-3, while the rest are pseudogenes. Increasing evidence also suggests that some SLA haplotypes have a duplicated SLA-1 locus, designated SLA-1′ (Smith et al., 2005a; Ho et al., 2009a). Because the SLA class I genes are encoded by three loci that means that potentially six different genes can be expressed in a heterozygous individual. Indeed, Tanaka-Matsuda et al. (2009) sequenced the genomic region carrying SLA classical class I genes of an SLA I heterozygous pig (Hp-28.0/Hp-62.0) and reported up to six classical class I genes in Hp-28.0, including an additional SLA-1′ (like) gene, newly designated SLA-12. Thus the SLA complex has copy number variance for the classical class I loci; this varies depending on SLA haplotype (Hp) (Tanaka-Matsuda et al., 2009). The non-classical class I genes, SLA-6, SLA-7 and SLA-8, are located at the centromeric end of the class I region. The SLA class I region also harbours the MHC class I chain-related genes (MIC); MIC2 is predicted to be functional while the MIC1 gene appears to be a pseudogene (Fig. 6.1a) (VEGFA Pig, 2010).

Phylogenetic analyses showed that the SLA class I genes displayed much more sequence homology to each other than to the HLA class I genes (Renard et al., 2003; Smith et al., 2005a; Ho et al., 2009a,b). SLA class I genes were named with numbers to avoid the implications that any of these loci were more homologous to the HLA-A, HLA-B or HLA-C genes of the HLA system. Sequence comparisons indicated that the SLA-1 and SLA-3 genes are more similar to each other, as is the SLA-1′, than they are to SLA-2. Therefore, it is likely that these arose as gene duplications after the separation of primates and Artiodactyla.

The loci encoding the expressed SLA class II antigens include the α- and β-chain genes for the SLA-DR, SLA-DQ, SLA-DM and SLA-DO proteins; genes are linked in tandem on SSC7 (Fig. 6.1a) (Renard et al., 2003; Smith et al., 2005a; Ho et al., 2009a). From the most centromeric SLA-DRA gene in the class II gene cluster, the order of the expressed SLA genes is DRB1, DQA, DQB1, DOB1, DMB, DMA and DOA (Fig. 6.1b). In contrast to the HLA system, there are no loci encoding the DP proteins. There are several class II β-chain pseudogenes in the SLA class II region, including DRB2, DRB3, DRB4, DRB5, DQB1, DOB2 and uDYB. The expressed SLA class II antigens are designated as SLA-DR and SLA-DQ based on their homology to the human molecules. They are expressed primarily on the surface of professional APCs, such as macrophages, B cells and dendritic cells, as well as capillary endothelia (Wilson et al., 1996; Chamorro et al., 2000; Summerfield et al., 2003; Ezquerra et al., 2009; Summerfield, 2009). Unexpectedly, normal swine T cells express SLA class II antigens, with preferential expression on the CD8+ T cell subset (Pescovitz...
et al., 1985; Saalmüller et al., 1991; Saalmüller and Maurer, 1994). SLA class II antigens function mainly in presenting exogenous 10–18 amino acid peptides to CD4+ helper T cells.

Genes involved in the antigen presentation pathway, such as the transporter associated with antigen processing (TAP) genes (TAP1 and TAP2) and proteasomes (PSMB8 and PSMB9), are also located in the class II region between the DOB1 and DMB loci. The overall genomic organization of SLA class II region is well conserved with HLA, but the length of the SLA class II region is much shorter. Phylogenetic analyses also showed that the SLA class II genes demonstrated much stronger sequence homology with their HLA counterparts than they do with each other (Smith et al., 2005b). As a result, the functional SLA class II genes were named after their human counterparts to indicate the orthology between the two systems.

The SLA class III region is centromeric and physically linked to the contiguous class I region. Over 60 loci have been characterized in this region, including many important genes in the immune defence mechanism, such as the tumour necrosis factor (TNF) gene families (TNF, LTA and LTB), the steroid cytochrome P450 21-hydroxylase (CYP21) enzyme and components of the complement cascade (C2, C4A and CFB) (Brule et al., 1996; Peelman et al., 1996; Renard et al., 2006).

**SLA gene and haplotype polymorphisms**

One of the most remarkable features of the MHC genes is the extremely high degree of genetic polymorphism within loci. The human MHC Haplotype Project affirmed that they are the most polymorphic genes in the vertebrate genomes, with 300 total loci, including 122 gene loci with coding substitutions, of which 97 were non-synonymous (Horton et al., 2008). The Immuno Polymorphism Database-MHC (IPD-MHC) web site (http://www.ebi.ac.uk/ipd/mhc/sla/) serves as the repository for maintaining a list of all recognized SLA genes and their allelic sequences. The Nomenclature Committee for Factors of the SLA System regularly updates SLA nomenclature and designates alleles of each locus into groups based on sequence similarity (identification of ‘groupspecific’ polymorphic sequence motifs) (Smith et al., 2005a; Ho et al., 2009a). The allelic group assignments were based primarily on polymorphisms in the exon 2 and 3 sequences for class I alleles and in the exon 2 sequences for class II alleles, given that these regions encode the peptide-binding domains, as well as interacting directly with the immune cell receptors, and are therefore considered to be functionally vital (Ho et al., 2009b, 2010). Given the strong linkage disequilibrium exhibited for SLA loci, haplotypes have been designated for specific combinations of alleles of class I and II genes; these were first defined by high-resolution DNA sequencing (Smith et al., 2005a; Ho et al., 2009a). These high-resolution SLA haplotypes are named with a prefix ‘Hp’, and a number for the class I haplotype followed by a number for the class II haplotype separated by a period (e.g. Hp-1.1).

Based on the IPD-MHC SLA database, a total of 116 SLA classical class I alleles and 13 non-classical class I alleles have been identified to date; the SLA-1, SLA-3 and SLA-2 genes are highly polymorphic (Smith et al., 2005b; Ho et al., 2009b). There are 12 SLA-1 allele groups with a total of 44 alleles; seven SLA-3 allele groups with 26 alleles, and 14 SLA-2 allele groups with 46 alleles. The extreme polymorphisms of the SLA class I genes are, as expected, concentrated in exons 2 and 3 of the coding regions which form the class I protein peptide-binding groove. The non-classical SLA-6 gene appears to be largely monomorphic.

There are a total of 167 SLA class II alleles identified to date (128 β-chain, 39 α-chain alleles), with polymorphisms mainly located in exon 2 of the coding sequences (Smith et al., 2005b; Ho et al., 2009b; IPD-MHC web site). The SLA-DRB1 and SLA-DQB1 loci display a very high degree of polymorphism. There are 14 DRB1 allele groups with a total of 82 alleles, and nine DQB1 allele groups with 44 alleles. The SLA-DQA locus exhibits a moderate degree of polymorphism, with 20 alleles and five allele groups identified to date. As with HLA-DRA, the SLA-DRA locus exhibits a very limited polymorphism, with 13 alleles representing three allele groups, despite the fact that it also encodes part of the domain for binding antigenic peptides. Ando et al. (2001) characterized the DNA sequence of five SLA-DMA alleles, which
showed only a few nucleotide substitutions in exon 3 and exon 4 of their coding regions.

Originally, SLA haplotypes were serologically defined; a total of 72 SLA class I haplotypes (designated H01–H72) were reported, reflecting mostly European commercial pig breeds (Renard et al., 1988; Chardon et al., 2000). To date, a total of 29 SLA class I haplotypes and 21 SLA class II haplotypes have been defined by means of high-resolution DNA sequencing. With PCR-sequence-specific primer (SSP)-based SLA typing methods, a total of 49 class I and 30 class II SLA haplotypes have so far been identified after testing nearly 850 pigs obtained from multiple commercial sources (Ho et al., 2009b, 2010). This apparently low SLA diversity in commercial pigs may be the result of selection and resultant selective breeding for desirable production traits.

Based on earlier data, there is substantial linkage disequilibrium, the non-random association of alleles at adjacent loci, for the SLA complex genes; the overall recombination frequencies were reported to be 0.4–1.2% within the SLA region and 0.05% within the class I region (Vaiman et al., 1978; Pennington et al., 1981; Edfors-Lilja et al., 1993; Smith et al., 1995; Chardon et al., 2000). Most previously documented crossovers mapped to the SLA class III region. However, use of PCR-SSP has resulted in the identification of additional recombinations corresponding to crossover frequencies of 0.56% between the class I and class II regions, and 0.39% within the class I region, reflecting better detection methods (Ho et al., 2009a,b, 2010).

Role of SLA genes in immunity and swine health

Several authors have reviewed the potential of using genetic approaches to improve animal disease resistance (Mallard et al., 1991; Stear et al., 2001; Rothschild, 2003; Gibson and Bishop, 2005; Lewis et al., 2007; Rothschild et al., 2007; Lunney, 2010). Mallard, Willkie and their colleagues have performed a series of experiments to establish populations of pigs that they predicted would be more immunologically active and thus more resistant to infectious diseases; however, the high-immune-response pigs were more susceptible to Mycoplasma hyorhinis infection (Mallard et al., 1989a; Magnusson et al., 1998; Willkie and Mallard, 1999). Edfors-Lilja et al. (1998, 2000) were the first to trace quantitative trait loci (QTLs) regulating normal immune traits, as detailed below. Several QTL-influencing traits, including growth, backfat thickness and carcass composition map to the SLA complex (Rattink et al., 2000; Bidanel et al., 2001; Malek et al., 2001; Milan et al., 2002; Rothschild et al., 2007; Nagamine et al., 2009). Overall, these results could help to guide breeders in selectively increasing the frequency of certain SLA alleles, i.e., those that are known to be associated with enhanced disease resistance or QTL effects.

As management in the pig industry changes, the range of pathogens to which pigs are exposed is altered. Moreover, as consumers demand pork products free of antibiotic contamination, it becomes increasingly important that new heartier, healthier breeding stock be available (Lewis et al., 2007; Rothschild et al., 2007; Lunney, 2010). Disease-resistant pigs, in well-managed facilities, would help to decrease drug usage by producers and increase the health of the nation's food supply. Several groups have attempted to evaluate the relationship between the level and function of circulating immune cells with average daily gain, live and carcass measurements, feed intake and feed conversion (Mallard et al., 1991; Stear et al., 2001; Rothschild, 2003; Gibson and Bishop, 2005). Galina-Pantoja et al. (2006) showed that the CD16+, CD2+/CD16+, CD8+ and SLA-DQ+/- cell subsets appear to be important biomarkers involved with the inherent ability of the pig to efficiently grow and produce better carcass weight in representative commercial environments. Clapperton et al. (2009) estimated the heritabilities of numerous immune traits and measured their relationship with performance within both specific pathogen-free (SPF) and non-SPF environments. They found negative genetic correlations between performance and CD11R1+ cells, monocytes and alpha-1 acid glycoprotein.

The influence of SLA-encoded genes on immune and disease traits is broad. Based on studies using SLA-defined and SLA-inbred lines of pigs it was affirmed that SLA genes determined levels of antibody responses to protein and vaccine antigens (Table 6.3). Cellular
responses to defined antigens showed weak associations with specific SLA haplotypes. Because of the difficulty and expense of performing controlled pathogen challenge studies in SLA-defined pigs, only limited studies have been performed. Lunney and colleagues have established that both primary and secondary responses to the food-borne helminth parasite, *Trichinella spiralis*, are regulated by SLA-associated genes, whereas no such SLA association was found for *Toxoplasma gondii* infections (Table 6.3) (Lunney and Murrell, 1988; Madden et al., 1990, 1993; Dubey et al., 1996). Recent advances in SLA typing enabled J.H. Fritz et al. (2010, unpublished) to identify SLA class I and II alleles associated with the control of PRRSV infections.

This last decade has seen major progress in swine immunology and genetics, and particularly in understanding the SLA complex, its genetic loci and the role of SLAs in normal immunity and in infectious disease and vaccine responses. New technologies now mean that comprehensive probing of the role of SLA alleles and haplotypes should reveal specific antigenic epitopes that stimulate immune and vaccine responses. They should help to identify critical immune cell subsets and the exact SLA loci that facilitate cellular interactions for effective immune responses, and unveil novel immune pathways regulated by SLA genes. Overall, the stage is set for determining the crucial role of SLA genes and proteins in pig health and productivity.

### Table 6.3. SLA gene encoded disease and vaccine responses.

<table>
<thead>
<tr>
<th>Immune parameter</th>
<th>SLA association</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibody response levels</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-lysozyme</td>
<td>Higher Hp-14.0; lower Hp-2.0</td>
<td>Vaiman et al., 1978</td>
</tr>
<tr>
<td></td>
<td>Higher Hp-3.3; lower Hp-4a.4</td>
<td>Lunney et al., 1986</td>
</tr>
<tr>
<td>Anti-model antigen</td>
<td>Higher Hp-4a.4</td>
<td>Lunney et al., 1986; Mallard et al., 1989b</td>
</tr>
<tr>
<td>Anti-sheep red blood cell vaccination for <em>Bordetella bronchiseptica</em></td>
<td>Lower Hp-3.3</td>
<td>Lunney et al., 1986</td>
</tr>
<tr>
<td></td>
<td>Higher Hp-4a.4</td>
<td>Mallard et al., 1989a</td>
</tr>
<tr>
<td></td>
<td>Higher Hp-2.0</td>
<td>Rothschild et al., 1984; Meeker et al., 1987a,b</td>
</tr>
<tr>
<td><strong>Cellular responses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salmonella bacterial phagocytosis</td>
<td>Higher Hp-2.2</td>
<td>Lumsden et al., 1993</td>
</tr>
<tr>
<td>Delayed contact-type hypersensitivity induced by tuberculin protein</td>
<td>Higher Hp-4a.4</td>
<td>Wilkie and Mallard, 1999</td>
</tr>
<tr>
<td>Parasite antigen proliferation interferon induction</td>
<td>Higher Hp-3.3</td>
<td>Lunney and Murrell, 1988</td>
</tr>
<tr>
<td>Interferon induction</td>
<td>None significant</td>
<td>Jordan et al., 1995</td>
</tr>
<tr>
<td>Bacterial phagocytosis</td>
<td>Lower Hp-2.2</td>
<td>Lacey et al., 1989</td>
</tr>
<tr>
<td>Macrophage superoxide production</td>
<td>None with class II</td>
<td>Groves et al., 1993</td>
</tr>
<tr>
<td><strong>Disease responses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma initiation; tumour incidence</td>
<td>Higher Hp-2.2</td>
<td>Tissot et al., 1987, 1989, 1993; Blangero et al., 1996</td>
</tr>
<tr>
<td></td>
<td>QTLs map to SLA</td>
<td>Hruban et al., 1994; Geffrotin et al., 2004; Du et al., 2007</td>
</tr>
<tr>
<td>Primary <em>Trichinella</em> infection</td>
<td>Lower parasite burden in Hp-3.3</td>
<td>Lunney and Murrell, 1988</td>
</tr>
<tr>
<td>Secondary <em>Trichinella</em> infection</td>
<td>Faster anti-parasite in Hp-2.2</td>
<td>Madden et al., 1990, 1993</td>
</tr>
<tr>
<td>Primary <em>Toxoplasma</em> infection</td>
<td>None significant</td>
<td>Dubey et al., 1996</td>
</tr>
<tr>
<td>Primary PRRSV infection</td>
<td>Increased viral levels with Hp 1.0; decreased serum neutralizing antibody with Hp 0.1, 0.23, 0.26</td>
<td>J.H. Fritz et al. (2010, unpublished)</td>
</tr>
</tbody>
</table>
The Genetics of the Porcine Immunoglobulins (Igs)

The Ig heavy chain (IGH) locus isotypes and allotypes

The Ig heavy chain (IGH) locus is located on SSC 7q26 (Chaudhary et al., 1997). Swine possess the same complement of heavy-chain isotypes as other mammals: IgM, IgD, multiple subclasses of IgG, and IgE and IgA. These are listed in Table 6.4 along with features describing the gene products they encode, the known genetic variants, and the source of sequence information for each in GenBank and on the Comparative Immunoglobulin Workshop (CIgW) web site (CIgW, 2010). As in other mammals, IgG is the major serum Ig and IgM is the major BCR (Butler et al., 2009a,c). IgD is transcribed in many lymphoid tissues (McAleer et al., 2005), and presumably acts as a second BCR. As there are no antibodies specific for IgD available, neither its serum concentration nor its prevalence as a BCR is known. Fig. 6.2a illustrates the domain structure of the IgM, IgD and IgG3, the three most 5’ genes of the heavy-chain constant region locus (IG). The Cμ1 and Cδ1 domains are nearly identical and IgD can be transcribed with the first domain of IgM (Cμ1). Unlike mice and humans, but similar to cattle, IgD is associated with a short switch region (Sδ) and, at least in cattle, this appears to be functional (Zhao et al., 2003). Both IgM and IgD have both transmembrane exons, e.g. Cδm1 and Cδm2, and an exon needed for secretion (Cδs; Fig. 6.2a).

There are six expressed IgG heavy chains that are designated IgG1 to IgG6 (Table 6.4). Alleles have been identified for all of them except IgG3. Based on their frequency of transcription, IgG1, IgG3 and IgG5 appear to be the most important (Butler and Wertz, 2006).

Table 6.4. Porcine immunoglobulin isotypes. Modified from Butler et al. (2009b).

<table>
<thead>
<tr>
<th>Isotype or subisotype</th>
<th>Molecular massa</th>
<th>S20,w</th>
<th>Constituent polypeptides</th>
<th>GenBank number</th>
<th>Allotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>H-chainb</td>
<td>H-chainc</td>
<td>L-chainb</td>
</tr>
<tr>
<td>IgM</td>
<td>1100</td>
<td>19.0</td>
<td>69–74</td>
<td>63.8</td>
<td>25</td>
</tr>
<tr>
<td>dIgA</td>
<td>330</td>
<td>9.5</td>
<td>58</td>
<td>51.2</td>
<td>25</td>
</tr>
<tr>
<td>sIgA</td>
<td>380</td>
<td>11.5</td>
<td>58</td>
<td>51.2</td>
<td>25</td>
</tr>
<tr>
<td>IgG1</td>
<td>160–170e</td>
<td>6.6</td>
<td>55</td>
<td>49.6</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG2</td>
<td>NDf</td>
<td>6.6</td>
<td>ND</td>
<td>49.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG3</td>
<td>ND</td>
<td>6.6</td>
<td>ND</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG4</td>
<td>ND</td>
<td>6.6</td>
<td>ND</td>
<td>49.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG5</td>
<td>ND</td>
<td>6.6</td>
<td>ND</td>
<td>49.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG6</td>
<td>ND</td>
<td>6.6</td>
<td>ND</td>
<td>49.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgE</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>60.2</td>
<td>25</td>
</tr>
<tr>
<td>IgD</td>
<td>~160</td>
<td>ND</td>
<td>ND</td>
<td>56</td>
<td>25</td>
</tr>
</tbody>
</table>

a Determined by equilibrium sedimentation by Setcavage and Kim (1976) and supported by less rigorous estimates from size exclusion chromatography. Estimates for sIgA are based on chromatographic behaviour. Values are in kilodaltons (kDa).
b Determined by SDS-PAGE and expressed as kDa.
c Based on deduced amino acid sequence and using a mean HCDR3 length of 42 nucleotides. Differences among Cγ subclass chain length are too small to be resolved by SDS-PAGE. The deduced mass of the kappa and lambda chains is 24kDa, i.e. the same as determined by SDS-PAGE, which is consistent with the absence of glycosylation.
d J-chain and secretory component.
e Individual subclass proteins have never been purified. The value given is only for partially purified IgG1, which is the major IgG transcribed.
f ND = not determined.
Unfortunately there are currently no fully verified monoclonal antibodies (mAbs) that can identify these subclass IgGs, and therefore their concentrations in serum or their role in any particular immune response cannot be measured. IgG is the ‘flagship Ig’ for all mammals; in nearly all eutherian mammals IgG has been diversified after speciation into subclasses, except among very closely related species like domestic sheep, cattle and goats (Kehoe and Capra, 1974; Nguyen, 2001; Butler, 2006; Butler et al., 2009b). Among the artiodactyls of the family Bovidae, diversification appears to have begun in a common ancestor. It is important to appreciate that the names, e.g. IgG1 and IgG2, among mammals do not imply homology. As in other mammals, major differences among swine subclasses are located in the hinge region (Butler et al., 2009b).

Evolutionary divergence of subclasses appears to have been through a combination of gene duplication and genomic gene conversion. The total number of IgG genes in swine may exceed the six subclasses that are expressed, as there is evidence for additional pseudogenes (T. Eguchi-Ogawa et al., unpublished). In addition there are some animals that have deletions of IGHG2 and IGHG6 (Butler et al., 2009b). This suggests that the IGHG portion of the heavy chain locus has been very flexible during evolution.

The allelic variants of IGHG are best known for IGHG1 and IGHG2. There are outbred swine that are homozygous for IgG1* and IgG2*, especially among some Mieshan that are also homozygous for IgA*. There are outbred swine that are homozygous for IgG1* and IgG2*, especially among some Mieshan that are also homozygous for IgA*. The haplotype linkages for the allelic variants of the IgG subclasses have not been established but should be forthcoming with additional mapping studies and from ongoing investigation of JH5± heterozygotes that have only one functional...
haplotype. An interesting feature of IGHG3, which is phylogenetically the ancestral IGHG for swine (Butler et al., 2009b), is its expression in fetal and newborn piglets (Butler and Wertz, 2006). However, once the newborn encounters antigen, IgG3 expression drops precipitously from 50% to <1% (J.E. Butler et al., unpublished). The map location of IGHG3 is the most IGHD proximal IGHG gene. As IgG3 appears in a variety of peripheral lymphoid tissues of the newborn as the major expressed IgG transcript, it may be spliced to the rearranged VDJ (a DNA rearrangement between a selected IGHV, a selected IGHD and a selected IGHJ gene) in the manner of IgD in mice, although a >3 kb switch sequence (Sy3) is present (Fig. 6.2a) (Eguchi-Ogawa et al., 2010).

Genomic map of the constant heavy-chain locus

The genes for IgE (IGHE) and IgA (IGHA), in that order, are found at the 3′ end of the heavy chain constant region locus, as is true for other mammals (T. Eguchi-Ogawa et al., unpublished). IGHE is highly conserved and most homologous to human and horse IGH. Swine IGH is also conserved and most homologous to human and dolphin IGH. Swine IgA occurs in two allelic forms; IgAβ has a mutated splice acceptor site, which results in four amino acid deletion in the hinge (Brown et al., 1995). The expressed gene frequency is much higher for IgAα, although, up to now, the hinge deficiency of IgAβ does not seem to have any deleterious effect. Rather, the gene frequency for the two allotypes appear founder related (Navarro et al., 2000). Both alleles appear to be equally susceptible to cleavage by proteases of Haemophilus parasuis (Mullens et al., unpublished). As in other mammals, IgA is the major mucosal Ig for swine, including in mature milk (Porter, 1969; Klobas and Butler, 1987). The IgA and IgM repertoires were more diverse in ileal than in jejunal piglet Peyer’s patches (PP), reflecting a more diversified microflora in the ileal PP (Levast et al., 2010). IgA responses to natural mucosal infections are well documented, such as with foot-and-mouth disease viruses, whereas killed vaccines preferentially stimulate IgG responses (Pacheco et al., 2010). While verified mAbs are available for the study of IgM and IgA, those for IgE, IgD or the various subclasses of swine IgG are not.

The variable heavy chain locus is composed of a yet undetermined number of IGHV genes, four diversity IGHD segments and five IGHJ segments (Eguchi-Ogawa et al., 2010). Nearly 30 IGHV genes have been identified (Butler et al., 2006a; ClgW, 2010). The 15 3′-most IGHV genes are listed in Table 6.5 and their position is shown in Fig. 6.2b. All porcine IGHV genes belong to the ancestral VH3 family of clan III (Sun et al., 1994). The porcine IGHV genes were originally named according to their order of discovery and referred to in the vernacular as VHA, VHB, etc., because their position in the locus was unknown at the time of their discovery. Since the recent mapping of the locus (Fig. 6.2b) these genes can now be renamed according to the nomenclature used for other well-studied species. We provide both the familiar vernacular (VH1, VH2, etc.) and the international ImMunoGeneTics (IMGT) nomenclature (IMGT, 2010) for these genes in Table 6.5 (Eguchi-Ogawa et al., 2010). Among these are almost all of the genes that comprise the expressed repertoire of fetal and young pigs (Sun et al., 1998; Butler et al., 2006a), as well as a number of pseudogenes.

All porcine IGHV genes have identical recombination signal sequences (RSS), which simplifies studies on B cell lymphogenesis (see below). Various IGHV genes appear as hybrids by sharing complementarity determining region 1 (CDR1) and CDR2 sequences, whereas their framework (FR) regions are virtually identical (Sun et al., 1994; Butler et al., 2006a). Therefore, it was postulated that they evolved by a combination of gene duplication and genomic gene conversion similar to the IgG subclasses of swine, and by the mechanism described by others (Meselson and Radding 1975; Szostak et al., 1983; Butler et al., 2009b). There is also a major segment duplication within the 3′ region of the locus that involves the unit containing VHA′, VHF and VHβ′, with the unit containing VHA, VHT, VHB and VHE (Fig. 6.2b). Repetitive sequences around these genes suggest that the duplication has been recent (Eguchi-Ogawa et al., 2010).
For the last 15 years, only two IGHD and one IGHJ gene segments were known to be expressed (Butler et al., 1996; Sun and Butler, 1996). The recent mapping studies (Fig. 6.2) indicated that only two of the four IGHD segments are functional and only the last IGHJ segment (JH5) is functional (Eguchi-Ogawa et al., 2010). In most cases, those segments not used have non-canonical RSS. Further evidence that swine have only one functional JH segment was demonstrated by disrupting IGHJ5, which results in an IGHJ5−/− piglet that completely lacks B cells (Mendicino et al., 2010). Swine are currently the only species in which disruption of a single IGHV gene segment eliminates the entire humoral immune system and, should it occur naturally, would result in the death of all affected pigs as well as heterozygous JH5−/+ piglets (Mendicino et al., 2010).

Further studies on the germ line IGHV repertoire will require mapping the region upstream from that shown in Fig. 6.2. This may also require the use of bacterial artificial chromosome (BAC) libraries prepared from swine of different breeds to account for the VH genes not accounted for by the current map (Fig. 6.2; Table 6.4). It is conceivable that a number of the IGHV genes that have been reported (Butler et al., 2006b) are alleles of those listed in Table 6.4. The NIH minipigs, which were selected for their SLA haplotypes, are homozygous for IGHV (Navarro et al., 2000), and usually for IgG1*

Table 6.5. Nomenclature for swine Ig DH, JH and 3′ VH Genes.\(^a\)

<table>
<thead>
<tr>
<th>IMGT nomenclature</th>
<th>Vernacular nomenclature</th>
<th>Number of exons</th>
<th>CDS (coding sequence) length (bp)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHV15</td>
<td>VHN</td>
<td>2</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>IGHV14</td>
<td>VHY</td>
<td>2</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>IGHV13p(^b)</td>
<td>P</td>
<td>1</td>
<td>213</td>
<td>Exon 1 and part of exon 2 were collapsed</td>
</tr>
<tr>
<td>IGHV12</td>
<td>VHB(^*)</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV11</td>
<td>VHF</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV10</td>
<td>VHA(^*)</td>
<td>2</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>IGHV9p</td>
<td></td>
<td>1</td>
<td>213</td>
<td>Exon 1 and part of exon 2 were collapsed</td>
</tr>
<tr>
<td>IGHV8</td>
<td>VHE</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV7p</td>
<td></td>
<td>1</td>
<td>213</td>
<td>Exon 1 and part of exon 2 were collapsed</td>
</tr>
<tr>
<td>IGHV6</td>
<td>VHB</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV5</td>
<td>VHT</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV4</td>
<td>VHA</td>
<td>2</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>IGHV3p</td>
<td></td>
<td>1</td>
<td>213</td>
<td>Exon 1 and part of exon 2 were collapsed</td>
</tr>
<tr>
<td>IGHV2</td>
<td>VHG</td>
<td>2</td>
<td>359</td>
<td></td>
</tr>
<tr>
<td>IGHV1p</td>
<td>P</td>
<td>2</td>
<td>352</td>
<td>Stop codon in exon 2</td>
</tr>
<tr>
<td>IGHD1</td>
<td>DHA</td>
<td>1</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>IGHD2</td>
<td>DHB</td>
<td>1</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>IGHD3</td>
<td></td>
<td>1</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>IGHD4</td>
<td></td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>IGHJ1</td>
<td></td>
<td>1</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>IGHJ2</td>
<td></td>
<td>1</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>IGHJ3</td>
<td></td>
<td>1</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>IGHJ4</td>
<td></td>
<td>1</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>IGHJ5</td>
<td>JH5</td>
<td>1</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)VH gene names were classified according to complementarity determining region 1 (CDR1) and CDR2 sequence specificities (Butler et al., 2006b). The vernacular terminology is based on order of discovery, whereas the IMGT (ImMunoGeneTics) system is based on the position in the genome as shown in Fig. 6.2b.

\(^b\) \(p\) = pseudogenes discovered in genomic DNA before mapping studies (see Butler et al., 2006b).
(J.E. Butler and N. Wertz, unpublished); they could be valuable for such extended studies.

Genomic organization of the Ig kappa and lambda light-chain loci

Like other mammals, swine possess two light-chain loci. The Ig lambda light-chain (IGL) locus maps to SSC14q17–q21 whereas the Ig kappa light-chain (IGK) locus maps to 3q12–q14 (Rettenberger et al., 1996; Chaudhary et al., 1997). The IGK locus in swine is organized in the same translocon manner as kappa in other mammals and as the Ig heavy-chain locus (Butler et al., 2004). The kappa locus contains a single IGKC gene and five IGKJ gene segments, which align very well to the IGKJ segments of the human, sheep, rabbit, rat, horse and mouse. These are located 3 kb upstream from IGKC and are separated by ~300 nucleotides from each other. IGHJ1 has a canonical RSS but this is not true for the heptamers of IGKJ2–IGKJ5. Nevertheless, Jκ2 accounts for 90% of all IGKJ usage in newborn piglets. Thus, the exchange of an A for C in the heptamer does not seem to matter. While not extensively studied in animals of different age and breed, it has not been confirmed whether IGKV3–IGKV5 are used and that their less canonical heptamers occur in all swine.

The IGKV genes of swine belong to at least two families: IGKV1 (Vκ1) and IGKV2 (Vκ2). There are >60 IGKV2 genes belonging to at least six subfamilies. The number of IGKV1 genes has not been estimated but they appear to belong to at least two to three subfamilies. Both IGKV1 and IGKV2 genes share 87% sequence similarity with the same families in humans (Butler et al., 2004). Kappa light chains in young and adult pigs appear to account for ~60% of total light chain usage. This is very similar to that in humans. However, at sites of early B cell development there appears to be preference for lambda usage, which probably represents expression of λ5, which is similar to the pre-BCR in mice (Butler et al., 2006b).

The IGL is less well characterized than kappa. There has never been a focused study of this locus. What is known is that IGL is organized as in other mammals with three to four tandem-arrayed IGLJ-IGLC cassettes that are located downstream from an unknown number of Vλ genes. These belong to at least two families, IGLV3 and IGLV8 (Butler et al., 2006a; K. Wells and J.E. Butler, unpublished).

B cell lymphogenesis and development of the porcine VH repertoire

B cell lymphogenesis is initiated when lymphocyte stem cells begin to rearrange their Ig genes. This starts with IGHD to IGHJ rearrangements and may occur simultaneously on both alleles, as it does in well-studied species, and is followed by rearrangement of IGHD-IGHJ to an IGHV gene. The process generates signal joint circles of intervening DNA that includes the RSS. These can be detected by PCR (Fig. 6.3a). Evidence for heavy chain rearrangement is also obtained by PCR, using a primer set spanning a region from FR1 of an IGHV gene to the 3' end of JH, which includes FR4. This part of the process starts in the fetal yolk sac at 20 days of gestation (20 DG). At 30 DG the process becomes evident in the fetal liver, and the primary transcript of the rearranged VDJs are then spliced to the Cμ chain and expressed on the surface, together with a surrogate light chain comprising a λ5 light chain and VpreB to form the pre-BCR. Porcine VpreB is highly homologous to that in mice and humans (Butler et al., 2009a). However, there are currently no molecular or antibody reagents that can distinguish the swine IGL gene products, e.g. IGL5 from other IGLC sequences. Surface IgM, i.e. the initial BCR, has not been detected in fetal liver, so the pre-BCR is probably expressed on only very few cells. In mice, the surrogate light chain is replaced by an authentic Cκ light chain. This remains unknown for swine because of the early predominance of Cλ expression in fetal piglets (Butler et al., 2005c). As mentioned above, this early λ expression could represent λ5, which is part of the pre-BCR, or could indicate that swine rearrange lambda before kappa. Later in development, the ratio of Cλ:Cκ is about 1, similar to that in humans (Butler et al., 2005b). The first expression of IgM on the B cell surface is at 45 DG in fetal liver (Sinkora...
Because the IGL locus genes have not been characterized, exactly which IGLV-IGLJ-IGLC or IGLV-VpreB rearrangements are first used remains unknown. However, when kappa light chains are later expressed, IGGV family genes rearranged to IGKJ2 predominate (Butler et al., 2004). Much more is known about IGHV usage during fetal and neonatal life, and how usage leads to the appearance of the heavy-chain pre-immune repertoire, than is known about the light-chain repertoire. It has been recognized for 15 years that swine use very few VH genes (Sun and Butler, 1996), and this pattern persists throughout most of fetal life (Sun et al., 1998). The major genes that are used are VHA, VHB, VHE, VHF, VHC and VHA* (Table 6.5). With the exception of VHC, all of these are located among the VH genes at the 3′ end of the locus, although VHJ, which is the terminal functional VH gene, is seldom used, while IGHV15 (VHN) is used in 20 DG yolk sac (Fig. 6.2b; Eguchi-Ogawa et al., 2010). Interestingly, IGHV15 (VHN) is expressed very early, but usage then declines in late fetal life. By monitoring the relative usage of the major IGHV genes described above, it is...
possible to construct a repertoire diversification index (RDI), the numerical value indicating the degree of diversity of the BCR. We have shown that there is a tendency for greater diversity at 20 DG, and this becomes contracted at 30 DG and remains so during the remainder of fetal life. However, the RDI increases after birth in antigen-sensitized animals (Butler et al., 2006b; J.E. Butler et al., unpublished). Much of this increase is the result of SHM. We have speculated that once the pre-BCR or BCR appears, e.g. at 30 DG, there is positive B cell selection, perhaps by stromal ligands in a manner similar to what results in the proliferation of pre-B cells when they express a pre-BCR (Melchers, 1995).

As swine use very few IGHV genes and have only two functional IGHD genes and one functional IGHI (IGHJ5), developing a repertoire comprising $10^{10}$ different B cells (Cohn, 2006) is dependent on the structure of CDR3 as well as on SHM. Because of the restricted VH repertoire of swine, we estimated that 95% of the swine repertoire is determined by CDR3 diversity (Butler et al., 2000). CDR3 is the sequence resulting from the junction of FR3 of the VH gene to the invariant tryptophan codon of FR4 in the 3’ portion of the JH segment (Fig. 6.3c). Diversity within CDR3 mostly results from exonuclease activity, which can shorten especially IGHD and IGHI, and by insertion of nucleotides between IGHV, IGHD and IGHI segments through the action of terminal deoxynucleotidyl transferase. The CDR3 region is also the target of SHM, but the extent of SHM is difficult to quantify compared with that for IGHV genes themselves (Fig. 6.3). Figure 6.3b illustrates that CDR3 regions differ in length as well as sequence (Fig. 6.3c), even when rearranged to the same IGHV gene. The origin of the sequence, i.e. from IGHD, IGHI or inserted nucleotides, is indicated in Fig. 6.3c. Among the thousands of VDJ sequences from swine that we have sequenced, we have never found the same sequence unless we accidentally recovered two VDJ from the same B cell. In other words, CDR3 is clone specific. In addition to differences in sequence, most CDR3s are of different length, so using a length analysis method called spectratyping it is possible to monitor the clonality of a particular lymphoid organ. Figure 6.3b shows that the repertoire of IgM appears unselected because the spectrum pattern is Gaussian, whereas the spectrum for IgG3 is especially selected, and that for IgA is selected to a lesser extent. Thus, spectratyping is a convenient way to study the genetics of the B cell repertoire. The importance of CDR3 to antibody specificity was demonstrated by Xu and Davis (2000), who showed that transgenic mice with only one IGVH gene, but with the IGHD and IGHI region of the locus intact, could make antibodies to nearly all antigens.

Modification of the BCR through somatic hypermutation (SHM) in swine

SHM depends on the expression of activation-induced cytidine deaminase (AID), a member of the APOBEC family of cytosine deaminases (Conticello et al., 2005; Chiu and Greene, 2008). This enzyme is primarily expressed in germinal centres following antigen stimulation (Muramatsu et al., 1999). However, there is very little germinal centre formation during fetal life in any mammal. In fetal piglets, the SHM frequency is <20 mutations per kilobase (Table 6.6). This frequency is constant across the various segments of the IGHV genes. This sharply differs from that in piglets exposed to environmental antigen in which mutations accumulate in CDR1 and CDR2. A similar pattern is known for mice (Berek and Milstein, 1987). We know that part of the cause for the rise in repertoire diversification in antigenized piglets is the result of SHM (Butler et al., 2006b).

The Genetics of the Porcine T Cell Receptor Repertoire of Swine

Organization of the TCR loci in swine

Like other mammals, pigs have TCRs that consist of $\alpha/\beta$ and $\gamma/\delta$ heterodimers, TCR $\alpha/\beta$ and TCR $\gamma/\delta$, respectively. The TCR $\delta$ chain locus (TRD) is embedded in the TCR $\alpha$ chain locus (TRA) and located on the middle of the q-arm
of chromosome 7 (7q15.3→q21) (Hiraiwa et al., 2001; Uenishi et al., 2003). The TCR β (TRB) and γ (TRG) loci are mapped on the q-arms of chromosome 18 (18q11.3→q12) and 9 (9q21→q22), respectively (Eguchi-Ogawa et al., 2009). The TRA and TRG chain genes use variable (TRV) and joining (TRJ) segments for genomic recombination to generate functional genes, whereas TRB and TRD use TRV, diversity (TRD) and TRJ segments for recombination. Among pig TR loci, detailed genomic sequences from constant regions to the 3′-region carrying V segments are available in TRA/TRD and TRB. In pig TRA, 62 TRAJ segments were observed, and the sequence of most of the segments was highly conserved to their human and mouse genome counterparts (Uenishi et al., 2003). Humans and pigs commonly have four TRDJ segments on their genomes. Humans and mice have four and three TRDD segments, respectively; however, in pig TRD, seven TRDD segments were observed, although one of the TRDD segments is considered to be non-functional. This suggests that the pig TRD locus has an ability to generate broader diversity of the rearranged TCR gene than the human and mouse loci, particularly in CDR3. Comparison of the genomic region carrying TRDD segments between humans and pigs demonstrates a fourfold duplication in the pig genomic sequence, which to date covers only the 3′ part of the entire TRAV-TRDV region (Fig. 6.4) (Uenishi et al., 2009). For TRB loci, the 3′-proximal regions carrying TRBV and the entire region containing TRBD-TRBJ-TRDC units were completely sequenced. The composition of individual sets of TRBD-TRBJ-TRBC does not have remarkable differences between humans, mice and pigs. However, three TRBD-TRBJ-TRBC sets were observed in the pig genome, whereas the human and mouse genomes have only two (Fig. 6.5); all of the three units in pigs are considered to be functional by observation of locus-specific expression (Eguchi-Ogawa et al., 2009).

### The expressed TCRVβ (TRBV) repertoire of swine

Nineteen families of porcine TRBV genes fitting into seven supergroups have been identified (Butler et al., 2005b). Based on >70% sequence similarity, we identified 17 that are sequence homologues of recognized human TRBV families (Fig. 6.6). Many of the same TRBV genes were also reported by Baron et al. (2001). The porcine gene families were given the same names as their human counterparts, e.g. pTRBV12 corresponds to hTRBV12. One family (pTRBVX) has no human homologue. TRBV families that belong

<table>
<thead>
<tr>
<th>Source</th>
<th>Number</th>
<th>FR1</th>
<th>CDR1</th>
<th>FR2</th>
<th>CDR2</th>
<th>FR3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newborn piglet</td>
<td>7</td>
<td>16.8 ± 10</td>
<td>28 ± 12</td>
<td>0 (NA)</td>
<td>35 ± 28</td>
<td>20 ± 10</td>
<td>17.9 ± 7</td>
</tr>
<tr>
<td>Colonized isolator</td>
<td>10</td>
<td>8.2 ± 11</td>
<td>140 ± 56</td>
<td>2.3 ± 2</td>
<td>104 ± 60</td>
<td>31 ± 24</td>
<td>34.8 ± 13</td>
</tr>
<tr>
<td>Conventional 5 week</td>
<td>15</td>
<td>14.9 ± 7</td>
<td>195 ± 66</td>
<td>12.6 ± 7</td>
<td>181 ± 45</td>
<td>79 ± 56d</td>
<td>68.6 ± 25</td>
</tr>
</tbody>
</table>

*Mean values for each region of the Vγ gene and for the total among different groups were compared by Student’s t tests.*

*Significantly higher than in newborn piglets (P ≤ 0.007–0.0001).*

*Rearred conventionally and infected with helminth parasites.*

*Significantly higher than in newborn and colonized piglets (P < 0.008).*
Fig. 6.4. Dot-plot analysis for comparison between porcine and human germ line sequences encoding the TCR α/δ chain genes (TRA/TRD). The pig genomic sequence carrying TRA/TRD segments (top, horizontal) were reconstructed from the registered data in DDBJ/EMBL/GenBank (AB457789, AB182374, AB182373, AB182372, AB182371 and AB053451) and subjected to masking of genomic interspersed sequences. The human counterpart sequence (left, vertical) was also reconstructed from the publicized data (AE000660–AE000662). Locations of human TCR α/δ gene segments are annotated according to IMGT (ImMunoGeneTics)/LIGM-DB (Ehrenmann et al., 2010). The TRDV1 segments in the human and porcine genomes are indicated by closed triangles (on the vertical and horizontal sequences) and broken lines. The repeated units carrying TRDV1 and TRDD segments are depicted by dark (a) and light grey (b) rectangles, respectively.
to the same supergroup share the same leader sequence: pTRBV4, 5, 7 and 12 were most frequently recovered. TRBV genes of seven families were only recovered from peripheral T cells and not from thymocytes. Only pTRBV2-9 and -15 were recovered in large enough numbers to render this observation statistically significant. No pTRBV family was exclusively

Fig. 6.5. Genomic structure of the porcine TRBD–J–C gene clusters. CDSs (coding sequences) and repetitive sequences are indicated. Intron lengths of the TRBC segments are indicated in parentheses. Characteristic repetitive sequences detected in the region are also indicated. Recombination signals of the TRBJ3-A segments started within TRBJ3-2, suggesting that TRBJ3-A was a pseudogene.

Fig. 6.6. The expressed porcine TRBV genes belong to 19 families. The porcine TRBV families were renamed using the human nomenclature in the IMGT (ImMunoGeneTics) database. The nomenclature used by Baron et al. (2001) is also given. The numbers in parentheses are the number of clones of each family that were recovered in the analysis by Butler et al. (2005b).
recovered from thymocytes. Most frequently used families are multigenic, but pTRBV5 and pTRBV12 may be represented by a single gene (J.E. Butler and N. Wertz, unpublished data).

The expression of porcine TCRVδ (TRDV) gene families

The expression of the five TRDV families (TRDV1–TRDV5) has been studied in swine of various ages (Holtmeier et al., 2002, 2004). Clonal diversity was measured by a combination of spectratyping (see example of method in Fig. 6.3b) followed by recovery and sequence analysis of CDR3 regions. Organ-specific spectratypic patterns suggested compartmentalization of T cells expressing different TRDV gene families in the spleen, lung, intestine and colon. As expected from their limited antigen exposure, the ileal PP and caecum were more oligoclonal in germ-free piglets than in specific pathogen-free piglets (Holtmeier et al., 2002). The TRDV repertoire also showed increasing clonal restriction with age and development, especially in the intestine. Nevertheless, sequence analysis of recovered CDR3 sequences revealed that some of the same T cell clones were widespread (Holtmeier et al., 2002). The compartmentalization in adult swine is diagrammatically shown in Fig. 6.7, and the overall pattern is reviewed in greater detail elsewhere (Butler et al., 2006a).

Quantitative Trait Loci (QTLs) for Immune Response Traits

Numerous studies have been aimed at mapping QTLs for traits related to immune responses (Table 6.7) (Edfors-Lilja et al., 1998, 2000; Wattring et al., 2005; Reiner

One feature of the TCR that has heretofore been less well realized than in studies on BCR expression is the expansion of certain T cell clones bearing so-called invariant receptors. This has also been shown for the Vδ3-Jδ3 receptor in swine (Holtmeier et al., 2004). An invariant TRDV TCR has also been shown in mice (Havran et al., 1991). Expression of clones bearing this receptor in swine dominates the TRDV3 repertoire of all organs at mid-gestation. At the end of gestation, clones bearing this receptor were absent in the thymus but were still present in intestine and spleen. This, as reviewed by Butler et al. (2006a), may indicate a role for the γδ T cells in innate immunity, just like the dominant BCR expression in fetal piglets discussed above. Both phenomena may suggest positive selection of both T and B cells during development, which would be consistent with the presence of a grey area between innate and adaptive immunity (Flajnik and Du Pasquier, 2004). A role for γδ T cells in innate immunity has also been suggested by others (Takamatsu et al., 2006).

Fig. 6.7. Compartmentalization of the γδ TCR repertoire in adult swine. The various patterns indicate shared (same pattern) or unshared (different pattern) repertoires. The vast majority of γδ T cells are compartmentalized. A highly polyclonal repertoire was always present in the spleen. A minority of clones are shared between the intestine and lung. Adapted from Holtmeier et al. (2002) and Butler et al. (2006a).
### Table 6.7. Quantitative trait loci (QTLs) associated with pig health traits.

<table>
<thead>
<tr>
<th>Trait names</th>
<th>Abbreviation</th>
<th>QTL map location as SSC :cM map position (range)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haptoglobin concentration</td>
<td>HAPT</td>
<td>3: (18–60); 5: (80–119); 6: (88–92); 8: 31; 9: (73–100); 12: 10: 191; 11: 58; 12: 31; 37; 14: (7–60); 105; 17: 9; (48–94)</td>
<td>Wimmers et al., 2009</td>
</tr>
<tr>
<td>Haematocrit</td>
<td>HCT</td>
<td>1: (52–81); 76; 2: (27–42); 3: (17–42); 4: 61; 5: 72; 6: 63; 7: (58–90); 63; 66; 8: (52–62); (74–83); (80–107); 10: (67–101); 69; (101–128); 13: 33; (35–70); (70–102); 18: 32</td>
<td>Watrang et al., 2005; Reiner et al., 2007a; Zou et al., 2008</td>
</tr>
<tr>
<td>Platelet count</td>
<td>PLTCT</td>
<td>2: 65; 3: 42; 53; 5: 31; 40; 6: 91; 7: 30; 13: 76; 18: 37</td>
<td>Reiner et al., 2007b; Yang et al., 2009</td>
</tr>
<tr>
<td>Red blood cell count</td>
<td>RBC</td>
<td>2: (27–42); 5: 72; 7: (58–90); 77; 90; 8: 61; 63; 64; (80–107); 13: (35–70); (70–102); 15: (34–81); 18: (31–45)</td>
<td>Reiner et al., 2007a; Zou et al., 2008</td>
</tr>
<tr>
<td>Disease resistance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV antibody titre</td>
<td>ADVAB</td>
<td>2: 87; 6: 84</td>
<td>Wimmers et al., 2009</td>
</tr>
<tr>
<td>Anti-K88 Escherichia coli Ig</td>
<td>K88Ab</td>
<td>5: 64</td>
<td>Edfors-Lilja et al., 1998</td>
</tr>
<tr>
<td>Anti-O149 E. coli IgG</td>
<td>O149Ab</td>
<td>6: 69</td>
<td>Edfors-Lilja et al., 1998</td>
</tr>
<tr>
<td>Melanoma susceptibility</td>
<td>MELAN</td>
<td>1: 49; 78; 85; 88; 2: 10; 17; 18; 23; 26; 3: 42; 4: 4; 84; 6: 5; 17; 108; 140; 7: 4; 49; 73; 8: 22; 58; 60; 56; 9: 42; 47; 10: 42; 54; 12: 96; 13: 56; 71; 81; 14: 34; 46; 98; 15: 86; 16: 45; 17: 36; 40; 45; 53; 18: 19</td>
<td>Du et al., 2007; Gomez-Raya et al., 2007</td>
</tr>
<tr>
<td>PRRSV antibody titre</td>
<td>PRRSVAB</td>
<td>1: 56; 7: (81–102)</td>
<td>Wimmers et al., 2009</td>
</tr>
<tr>
<td>Resistance/susceptibility to pseudorabies</td>
<td>PrV</td>
<td>5: 31; 6: 41; 9: 139</td>
<td>Reiner et al., 2002</td>
</tr>
<tr>
<td>Sarcocystis miescheriana IgG</td>
<td>SMIGG</td>
<td>4: 39; 5: 104; 126; 7: 156; 8: 89; 12: 33; 17: 83; X: 58</td>
<td>Reiner et al., 2007b</td>
</tr>
<tr>
<td>S. miescheriana IgM</td>
<td>SMIGM</td>
<td>5: 58; 12: 96; 13: 19</td>
<td>Reiner et al., 2007b</td>
</tr>
<tr>
<td>E. coli F18 receptor</td>
<td>ECF18R</td>
<td>6: 76</td>
<td>Meijerink et al., 1997</td>
</tr>
<tr>
<td>Immune capacity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Band-formed neutrophil no.</td>
<td>BFNEUT</td>
<td>1: (79–103)</td>
<td>Watrang et al., 2005</td>
</tr>
<tr>
<td>Basophil no.</td>
<td>BASO</td>
<td>8: 126; 13: (35–70)</td>
<td>Reiner et al., 2008</td>
</tr>
<tr>
<td>Complement (C’) C’3c concentration</td>
<td>C3C</td>
<td>1: 139; 2: 23; 25; 29; 101; 3: (60–72); 4: (79–120); 5: 40; 55; 224; 6: (88–93); 7: 60; (102–135); 8: 164; 10: 191; 12: 147; 14: 24; 43; 91; 15: 73; 16: (0–47); 36; (47–93); 75; 79; 18: 36</td>
<td>Phatsara et al., 2007; Wimmers et al., 2009</td>
</tr>
<tr>
<td>CD2+ leucocyte no.</td>
<td>CD2L</td>
<td>1: (43–89)</td>
<td>Watrang et al., 2005</td>
</tr>
<tr>
<td>CD4+ leucocyte no.</td>
<td>CD4L</td>
<td>1: (43–89); 8: (74–83)</td>
<td>Watrang et al., 2005</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>CD8+ leucocyte no.</td>
<td>CD8L</td>
<td>1: (79.4–102.9)</td>
<td>Wattrang et al., 2005</td>
</tr>
<tr>
<td>conA-induced cell proliferation</td>
<td>conA</td>
<td>7: 73; 8: (74–83)</td>
<td>Edfors-Lilja et al., 1998; Wattrang et al., 2005</td>
</tr>
<tr>
<td>Eosinophil no.</td>
<td>EOS</td>
<td>1: 123; 11: 69; 12: 7; 17: (0–32); 22: 26; 41: X: (0–35); 31: 125</td>
<td>Reiner et al., 2007b; 2008</td>
</tr>
<tr>
<td>Haemolytic C’ activity (alternative pathway)</td>
<td>AH50</td>
<td>2: 17; 20; 28; 3: 117; 4: (79–120); 7: 47; 11: 24; 30; 17: 69</td>
<td>Wimmers et al., 2009</td>
</tr>
<tr>
<td>Haemolytic C’ activity (classical pathway)</td>
<td>CH50</td>
<td>2: 92; 4: (79–120); 6: 81; (93–102); 8: (52–62)</td>
<td>Wimmers et al., 2009</td>
</tr>
<tr>
<td>IgM+ leucocyte no.</td>
<td>IGML</td>
<td>1: (53–81)</td>
<td>Wattrang et al., 2005</td>
</tr>
<tr>
<td>Lymphocyte no.</td>
<td>LYMPH</td>
<td>3: 18; 7: (30–58); 64; 8: (38–58); (52–62); −1.3; 11: (0–33); 12: 31; 95.8; 14: 51; 76; 15: 92; 17: (63–94); 18: (0–32)</td>
<td>Wattrang et al., 2005; Reiner et al., 2008; Yang et al., 2009</td>
</tr>
<tr>
<td>MHCII+ leucocyte no.</td>
<td>MHCIIL</td>
<td>8: (74–82.8)</td>
<td>Wattrang et al., 2005</td>
</tr>
<tr>
<td>Monocyte no.</td>
<td>MONO</td>
<td>2: 64; 8: (107–126); 11: (0–33); 12: (33–65)</td>
<td>Reiner et al., 2008</td>
</tr>
<tr>
<td>N1c+ leucocyte no.</td>
<td>N1cL</td>
<td>8: (52–62)</td>
<td>Wattrang et al., 2005</td>
</tr>
<tr>
<td>PHA proliferation</td>
<td>PHA</td>
<td>1: (79–89)</td>
<td>Wattrang et al., 2005</td>
</tr>
<tr>
<td>PWM-induced cell proliferation</td>
<td>PWM</td>
<td>4: 75; 8: (74–83)</td>
<td>Edfors-Lilja et al., 1998; Wattrang et al., 2005</td>
</tr>
<tr>
<td>Segmented neutrophil no.</td>
<td>SNEUT</td>
<td>1: 79; (81–94); 2: (0–27); 4: 8; 7: 67; 8: (38–58); −1.3; 11: (0–33); 12: 96; 16: (9–24); 17: (0–32)</td>
<td>Wattrang et al., 2005; Reiner et al., 2008; Yang et al., 2009</td>
</tr>
<tr>
<td>Spontaneous cell proliferation</td>
<td>SPONT</td>
<td>13: 107</td>
<td>Edfors-Lilja et al., 1998</td>
</tr>
<tr>
<td>Stress-induced alteration in no. of neutrophils</td>
<td>IMM</td>
<td>8: 60</td>
<td>Edfors-Lilja et al., 2000</td>
</tr>
<tr>
<td>Stress-induced leucocyte proliferation</td>
<td>LEUKPRO</td>
<td>2: 78</td>
<td>Edfors-Lilja et al., 2000</td>
</tr>
<tr>
<td>White blood cell counts</td>
<td>WBC</td>
<td>1: 78; (79–94); (94–123); 2: 64; 7: 65; 8: (−1–38); (80–107); 10: (20–39); 12: 38; (80–96); 15: 83</td>
<td>Edfors-Lilja et al., 1998; Wattrang et al., 2005; Reiner et al., 2008; Yang et al., 2009</td>
</tr>
<tr>
<td>Pathogen</td>
<td>Parasite load</td>
<td>2: 64.3; 7: 60.7; 16: 24.3cM</td>
<td>Reiner et al., 2007b</td>
</tr>
</tbody>
</table>
QTLs are generally identified by comparing the linkage (degree of covariation) of polymorphic molecular markers and phenotypic trait measurements. The Pig QTL database, PigQTLdb (2010a), is a relational database that provides a comprehensive tool for QTL repository, comparisons and dynamic linking to comparative structural genome information. Selected QTLs are hypothesized to occur in specific chromosomal regions containing genes that make a significant contribution to the expression of a complex trait. A search of the PigQTLdb for health and immune traits (PigQTLdb, 2010b) produced numerous QTL map locations for blood cell subset frequencies, serum protein and antibody responses (Table 6.7). Studies to affirm these map locations with pig health and resistance/susceptibility to pathogen infections have only just begun, as illustrated by the work of Reiner et al. (2002, 2007b). The next decade will clearly lead to major discoveries, as illustrated by the detailed analyses of melanoma resistance by Geffrotin et al. (2004), Du et al. (2007) and Gomez-Raya et al. (2007).

Summary

The genetic diversity of the loci determining the immune responses of pigs is only beginning to be explored. The availability of the annotated swine genome sequence is expected to stimulate many more advances into understanding how immune genes affect pig health in all phases of production. Future studies will reveal the impact of allele variations on swine health and pathogen resistance/susceptibility. The stage is now set for fundamental studies to explore these areas.

References


Hiraiwa, H., Uenishi, H., Kiuchi, S., Watanabe, M., Takagaki, Y. and Yasue, H. (2001) Assignment of T cell receptor (TCR) alpha-chain gene (A), beta-chain gene (B), gamma-chain gene (G), and delta-chain gene (D) loci on swine chromosomes by in situ hybridization and radiation hybrid mapping. *Cytogenetics and Cell Genetics* 93, 94–99.


Meijerink, E., Fries, R., Vogeli, P., Masabanda, J., Wigger, G., Stricker, C., Neuenschwander, S., Bertschinger, H.U. and Stranzinger, G. (1997) Two alpha(1,2) fucosyltransferase genes on porcine chromosome 6q11 are closely linked to the blood group inhibitor (S) and Escherichia coli F18 receptor (ECF18R) loci. Mammalian Genome 8, 736–741.


PigQTLdb (Pig Quantitative Trait Locus (QTL) Database) (2010a) Available at: http://www.genome.iastate.edu/cgi-bin/QTLdb/SS/index (accessed 25 June 2010).


Since the early days of animal cytogenetics, the pig (Sus scrofa, SSC) has been one of the most widely studied farm animals. This interest has been promoted by the relatively fewer chromosomes in the domestic pig (2n = 38) compared with other domestic species, such as dogs (2n = 78), alpacas (2n = 74), cattle (2n = 60), goats (2n = 60), sheep (2n = 54), buffaloes (2n = 50), horses (2n = 64) and chickens (2n = 78). Furthermore, pig chromosomes are readily distinguishable from each other, and it is relatively easy to identify most of the chromosomes even without the use of special staining methods. Porcine chromosomes have largely been studied for the analysis of karyotypic anomalies in somatic and germ cells, gametes and embryos. Additionally, the pig is believed to be a good model for the human to study the effect of chromosome aberrations in male and female meiosis.

The 1990s witnessed an increasing interest in gene mapping in all farm animals, including pigs. Needless to say that a priori knowledge of the porcine chromosomes significantly contributed to enhancing this cause, making the porcine genome one of the best studied animal genomes by the mid-1990s. The physical and the meiotic linkage maps developed for individual pig chromosomes during the past two decades have played a significant role in understanding the organization and evolution of this genome, and finding genes and markers of economic significance in pigs. The recently launched porcine whole genome (WG) sequencing project does not mark the end of the era of chromosome maps. It is, rather, the opposite: good, integrated maps are prerequisites to validate and support the WG sequence assembly.
Molecular hybridization-based methods, initially developed for physical gene mapping, have become an integral part of cytogenetic analysis, blurring the boundaries between classical cytogenetics and molecular biology. Rapidly evolving array-based genome analysis tools further obscure this picture, bringing chromosome analysis to submicroscopic levels. Pig cytogenetics is no exception; it has stepped into the era of ‘new cytogenetics’ or ‘cytogenomics’ (Speicher and Carter, 2005).

The present chapter aims to address the porcine genome through two main sections. The first of these, Pig Chromosomes, deals with classical and molecular cytogenetics, with the emphasis on the latter. After a brief background to pig chromosomes, the reader will be introduced to the basic concepts of normal and abnormal pig chromosomes, and evolutionary aspect of the pig karyotype. The second main section, Gene Mapping, deals with the generation of genetic linkage and physical gene maps in pigs – from the first mapped locus to high-resolution integrated WG maps. This information, in combination with the WG draft sequence, helps in developing a thorough understanding of the organization of the porcine genome.

**Pig Chromosomes**

**Methods for studying pig chromosomes**

During the past 15 years, an impressive series of technical and conceptual advancements have shaped human and animal cytogenetics, leading to improvement of the resolution and accuracy of chromosome studies (Trask, 2002; Speicher and Carter, 2005; Dobigny and Yang, 2008). Pig cytogenetics today is essentially a combination of conventional chromosome analysis methods, such as chromosome banding, and a variety of sequencing-, PCR-, hybridization- and immunochemistry-based molecular approaches. Most of the molecular methods were originally developed for genome analysis and physical gene mapping, but were quickly adopted by cytogeneticists, and applied in many different fields of porcine chromosome analysis.

**Classical chromosome banding methods and pig chromosome nomenclature**

For the past 40 years, various banding techniques, such as QFQ (quinacrine fluorescence (Q)-banding) (Caspersson et al., 1970), GTG (Giemsa (G)-Trypsin-banding) (Seabright, 1971), RBG (reverse (R)-banding with Giemsa staining) (Dutrillaux and Lejeune, 1971) and RBA (reverse (R)-banding with acridine orange staining) (Dutrillaux, 1973), have been used to identify individual porcine chromosomes and arrange them into a karyotype. The first karyotype was based on Q-banding (Gustavsson et al., 1972), followed by several GTG-banded karyotypes (Gustavsson, 1980) that were consolidated into the first international standard (Ford et al., 1980). The latter was soon revised by adding RBA-banding (Lin et al., 1980; Ronne et al., 1987), schematic drawings of G- and R-banded chromosomes, and a band nomenclature system (Fig. 7.1) (Gustavsson, 1988). It was the first comprehensive chromosome nomenclature among domestic animals and enabled the description of normal karyotypes, aberrant patterns, and established the foundation for physical chromosome maps.

**The nucleolus organizer regions (NORs)**

The NORs are the chromosomal sites of 5.8S, 18S and 28S ribosomal RNA (rRNA) genes and were originally detected by the Ag-I technique or NOR-banding (Goodpasture and Bloom, 1975; Bloom and Goodpasture, 1976). The method uses ammoniacal silver solution and stains the chromosomal sites containing actively transcribed rRNA genes. More recently, NORs have been studied using fluorescence in situ hybridization (FISH) with ribosomal DNA (rDNA) probes that detect all loci of the rRNA genes, regardless of their transcriptional status. In pigs, NORs are located close to the centromeric regions of SSC8 and 10 (Lin et al., 1980; Toga-Piquet et al., 1984; Mellink et al., 1991). Additionally, a narrow intercalary site suggesting a small rRNA gene cluster is present on SSC16 (Bosma et al., 1991a). Another class of rRNA genes – the 5S rRNA genes – has been mapped by in situ hybridization (ISH) to SSC14q22 (Lomholt et al., 1995; Mellink et al., 1996). Numerical
variations and polymorphism of Ag-positive NORs have been demonstrated in pigs at the cellular, individual and population levels (Mellink et al., 1992, 1994). The activity of rDNA is correlated with the concentration of oxidative agents present in the environment, and quantitation of active NORs can be used to measure oxidative stress in porcine cells (Wnuk et al., 2008). The genes encoding for all four classes of rRNA genes have also been mapped.

Fig. 7.1. Schematic drawings of G-banded porcine chromosomes. Source: Gustavsson (Committee for the Standardized Karyotype of the Domestic Pig) (1988).
in other suids/Suiformes – the babirusa and the white-lipped peccary (Zijlstra et al., 1997) confirming the previously recognized chromosomal homologies between the three species (Bosma et al., 1996, 2004).

Centromeric and telomeric heterochromatin

Traditionally, regions containing centromeric and telomeric heterochromatin have been highlighted by the C (centromere)-banding (Sumner, 1972) or THA (terminal)-banding (Dutrillaux, 1973) techniques. While C-banding stains only pig chromosome centromeres, the THA-method stains all centromeres and most of the telomeric regions (Gustavsson, 1983). Additionally, bright interstitial THA-bands are present on SSC2q, 6q and 10p. More recently, heterochromatic regions in the pig genome have been studied using molecular methods such as restriction enzyme digestion (Miller et al., 1993; Adega et al., 2005), FISH (de la Seña et al., 1995) and primed in situ synthesis (PRINS) (Gu et al., 1996; Rogel-Gaillard et al., 1997b; Wnuk et al., 2008). Telomeric (TTAGGG) repeats localize at the ends of all chromosomes, but also interstitially on SSC6q22, 6q, 7q, and at the centromeres of all acrocentric chromosomes (de la Seña et al., 1995; Wnuk et al., 2008). Restriction enzyme analyses show the presence of two distinct centromeric satellite DNA families (Adega et al., 2005), of which the Ac2 family shares some similarity with telomeric sequences (Wnuk et al., 2008). Interestingly, the intercalary (TTAGGG), site at 6q coincides with a bright THA-band (Gustavsson, 1983), while the 6q and 7q sites are at the junction of porcine segments homologous to (the human segments) HSA1 and HSA19, and HSA6 and HSA15, respectively (Frönicke et al., 1996). Overall, porcine constitutive heterochromatin is very heterogeneous, consists of chromosome-specific satellite DNA families and can be divided into 22 heterochromatin subclasses. Of these, 12 are centromeric, four interstitial, five telomeric and one specific to the long arm of the Y chromosome (Adega et al., 2005). Polymorphism in the amount of centromeric heterochromatin has been described both in uniarmed (13–18) and biarmed (1, 2, 8, 11 and 12) chromosomes (Glahn-Luft et al., 1982; Eldridge, 1985; Gustavsson, 1990).

Chromosome analysis by fluorescence in situ hybridization, FISH

ISH is widely used in different branches of biology; it relies on the Watson–Crick base-pairing complementarity principle, and permits the location of DNA markers in their original place or in situ (Raudsepp and Chowdhary, 2008). The two major components of FISH are the probe and the target. In the case of cytogenetics and gene mapping, the targets can be whole cells, mitotic or meiotic chromosome preparations at different stages of the cell cycle, DNA fibres (Trask, 2002; Speicher and Carter, 2005; Raudsepp and Chowdhary, 2008; Rubes et al., 2009) or DNA tiling-arrays (Shaffer and Bejjani, 2006; Sharp, 2009). The probes vary considerably in size and origin, e.g. from a few base pairs of long telomeric or centromeric repeats to composite DNA sequences from the whole chromosome or chromosomal segment. The probes are labelled directly with fluorochromes or indirectly with molecules that need binding by fluorochrome-conjugated antibodies to visualize the signals. Over the past 20 years, FISH methodology has improved tremendously (Nederlof et al., 1989; Lichter, 1997; Trask, 2002). Sophisticated imaging systems and the available number of spectrally distinct fluorochromes allow simultaneous detection of two, three or even multiple probes, and the overall sensitivity of FISH has increased about 10,000 times.

In regard to porcine chromosomes, FISH was initially used for physical gene mapping (see the second part of the chapter). However, during the past 10 years the method has also become an integral part of cytogenetic analysis (Ducos et al., 2007, 2008; Rubes et al., 2009). The availability of large insert clones from pig genomic libraries (Rubes et al., 2009) and the development of flow-sorted and microdissected chromosome-specific probes (see below) have been instrumental in the use of FISH in both porcine cytogenetics and gene mapping.

Generation of chromosome-specific probes by flow sorting.

Among various domesticated species, the pig is the most attractive animal for flow sorting individual chromosomes, primarily due to small chromosomal number, distinct shape and size, and variation in GC- and AT-base pair ratios. Flow sorting of porcine chromosomes
was first carried out in the 1980s (Grunwald et al., 1986, 1989; Matsson et al., 1986; Gelfrotin et al., 1987), and fairly distinct separation of the chromosomes from normal and translocated cell lines was obtained. However, it took a few more years and a dual laser fluorescence activated cell sorter (FACS) to resolve the pig chromosomes into 19–20 definite peaks (Dixon et al., 1992; Schmitz et al., 1992). Chromosome painting (Lichter et al., 1988; Pinkel et al., 1988) was used to demonstrate that five of the 20 clusters observed in the bivariate flow karyotype of a male pig corresponded to chromosomes 1, 13, 18, X and Y (Langford et al., 1992). This breakthrough was significant for the unambiguous correlation of individual peaks to particular chromosomes. Later, two independent studies (Langford et al., 1993; Yerle et al., 1993) established complete correspondence of each of the cluster in the porcine bivariate flow karyotype with a specific chromosome, thus providing a ‘standard’ flow karyotype for the pig.

Whole chromosome paints (WCPs) obtained through flow cytometry have found good use in molecular cytogenetic analyses of various chromosomal abnormalities (Table 7.1).

**Table 7.1.** The use of pig chromosome-specific painting probes for the study of chromosomal abnormalities (in chronological order).

<table>
<thead>
<tr>
<th>Aberration</th>
<th>Painting probes</th>
<th>Probe origin</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rcp(7;15)</td>
<td>SSC7, 15</td>
<td>Flow sorted</td>
<td>Konfortova et al., 1995</td>
</tr>
<tr>
<td>rcp(6;13); rcp(11;16); rcp(6;16); rcp(13;17); rcp(6;14); rcp(3;5); rcp(2;14); rcp(15;17)</td>
<td>SSC2, 3, 5, 6, SSC11, 13, 14, SSC15, 16, 17</td>
<td>Flow sorted</td>
<td>Pinton et al., 1998</td>
</tr>
<tr>
<td>39,XXY</td>
<td>HSAX</td>
<td>Oncor, Coatasome®</td>
<td>Makinen et al., 1998</td>
</tr>
<tr>
<td>rcp(4;12); rcp(1;7); rcp(1;6)</td>
<td>SSC1, 4, 6, 7, 12</td>
<td>Flow sorted</td>
<td>Pinton, A. et al., 2000</td>
</tr>
<tr>
<td>rcp(4;6); rcp(2;6); rcp(5;17); rcp(5;8); rcp(15;17); rcp(7;8); rcp(5;8); rcp(3;15)</td>
<td>SSC2,3, 4, 5, 6, SSC7, 8, 15, 17</td>
<td>Flow sorted</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>rcp(7;18)</td>
<td>All SSC chromosomes</td>
<td>Microdissected</td>
<td>Kubickova et al., 2002</td>
</tr>
<tr>
<td>SSC4 pericentric inversion</td>
<td>SSC4p, 4q</td>
<td>Microdissected</td>
<td>Pinton et al., 2003</td>
</tr>
<tr>
<td>Structural aberrations rcp(3;15) and rcp(12;14)</td>
<td>SSC1,13</td>
<td>Microdissected</td>
<td>Rezacova et al., 2003</td>
</tr>
<tr>
<td>rcp(Y;14)</td>
<td>SSC14, SSCY</td>
<td>Flow sorted</td>
<td>Pinton et al., 2008</td>
</tr>
<tr>
<td>rob(13;17)</td>
<td>SSC13, 17</td>
<td>Flow sorted</td>
<td>Pinton et al., 2009</td>
</tr>
</tbody>
</table>

rcp, reciprocal translocation; SSC, Sus scrofa; HSA, Homo sapiens; rob, Robertsonian translocation.
rearrangements, such as minute marker chromosomes and minute deletions and translocations. In pigs, painting probes produced by microdissection have been used to study translocations (Kubickova et al., 2002; Rezacova et al., 2003; Pinton et al., 2005) and intrachromosomal rearrangements (Pinton et al., 2003) (Table 7.1).

**SPERM-FISH.** Sperm-FISH is carried out on decondensed sperm heads and has become a state-of-the-art technique to analyse the chromosomal constitution of sperm. In pigs, sperm-FISH has been used to validate the purity of flow cytometrically sorted boar sperm (Kawarasaki et al., 2000; Parrilla et al., 2003) but also to estimate the rate of aneuploidies in normal individuals (Rubes et al., 1999), and analyse meiotic segregation in translocation and inversion carriers (Pinton et al., 2004; Massip et al., 2008, 2009; Bonnet-Garnier et al., 2009).

**COMPARATIVE GENOMIC HYBRIDIZATION (CGH).** CGH was originally designed to overcome the difficulties associated with preparing high-quality metaphase spreads from some types of cells, such as various solid tumours. The technique was developed by a team lead by Ollie and Anna Kallioniemi, Dan Pinkel and Joe Gray (Kallioniemi et al., 1992) and uses competitive hybridization of two differently labelled (red and green) DNA probes (e.g. DNA from a normal cell and from a cancer cell) to normal metaphase chromosomes. The ratio of red-to-green fluorescence is measured along each chromosome, and regions with deletions and amplifications are determined. In porcine cytogenetics, chromosome-CGH has been used in the detection of the loss of SSC13q36–q39 in swine cutaneous melanoma (Apiou et al., 2004) and in investigating chromosome aneuploidies in early embryos (Hornak et al., 2009).

A new and promising refinement of the chromosome-CGH technology is array-CGH, in which metaphase chromosomes are replaced by an array of thousands of bacterial artificial chromosome (BAC) clones or short oligonucleotide sequences that form a tiling path over the entire genome or the genomic region of interest (Trask, 2002; Albertson and Pinkel, 2003; Speicher and Carter, 2005). It is anticipated that array-CGH will considerably improve the detection of disease-related complex chromosomal rearrangements (Shaffer and Bejiani, 2006; Lee et al., 2007; Sharp, 2009). It is also believed that array-CGH will help to uncover the extent of naturally occurring structural variations in genomes between normal individuals (Redon et al., 2006). Recently, array-CGH was for the first time used in pigs to get a snapshot of copy number variation (CNV) in the pig genome (Fadista et al., 2008). A custom tiling oligonucleotide array with a median probe spacing of 409 bp was designed for four porcine chromosomes, namely SSC4, 7, 14 and 17. Array-CGH analysis was carried out on 12 Duroc boar founders and one unrelated Hampshire boar, and 37 CNV regions were identified across the four chromosomes. Notably, some CNVs overlapped with known segmental duplications.

**Immunolocalization of chromosomal proteins**

The combined use of fluorescently labelled antibodies for the synaptonemal complex proteins SCP1 and SCP3, the mismatch repair protein MLH1 present at late recombination nodules (Barlow and Hulten, 1998), and the γH2AX protein that is associated with asynapsed chromatin and meiotic silencing (Fernandez-Capetillo et al., 2003) has considerably improved knowledge about the behaviour of meiotic chromosomes in normal cells, and in cells with chromosomal aberrations in humans, mice and chickens (reviewed by Villagomez and Pinton, 2008). The pig is so far the only domestic animal in which fluorescence immunocytochemistry has been used to study chromosomally abnormal spermatocytes. In one study, the meiotic behaviour of rob(13;17) was analysed (Pinton et al., 2009), while another showed the extent of chromatin silencing in a case of Y;14 translocation (Pinton et al., 2008).

**Chromosomal abnormalities – an overview**

As observed in other species, chromosome aberrations have also been found in pigs. These aberrations can be classified into two major categories: numerical and structural. Chromosome abnormalities associated with abnormal sexual
development, and cytogenetic anomalies found in porcine germ cells, gametes and embryos are discussed separately at the end of this section.

**Numerical aberrations**

Chromosome euploidy, which is characterized by variation of the whole haploid complement, has been found only in porcine sperm, ova and zygotes (discussed below). In contrast, aneuploidy is associated with individual chromosomes and can involve either the autosomes or the sex chromosomes.

**SEX CHROMOSOME ANEUPLOIDY.** This is relatively infrequent in pigs and only a few cases have been described. Among these are 37, XO X-monosomy (Nes, 1968; Lojda, 1975), 39, XXY and 40, XXXY Kleinfelter syndrome (Breeuwsma, 1968; Hancock and Daker, 1981; Gustavsson, 1984; Makinen et al., 1998), and mosaic karyotypes such as 39, XXY/40, XXXY (Breeuwsma, 1970; Hancock and Daker, 1981) and 37, X/38, XY/39, XY (Quilter et al., 2003). Individuals with sex chromosome aneuploidy may phenotypically appear male, female or intersex, but always demonstrate some abnormalities of the gonads or external genitalia.

**AUTOSOMAL ANEUPLOIDY.** Autosomal aneuploidy appears to be a non-existent phenomenon in live-born piglets. So far, only animals with mixoploid conditions, such as 37, XY−18/38, XY/39, XY+18, 38, XY/37, XY−18 (Vogt et al., 1974) and 38, XY/39, XX+14 ( Bölch et al., 1985), have been described. However, the condition exists at embryonic stages. Monosomy of SSC11 (Smith and Marlowe, 1971) and double trisomy of SSC17 and 18 (Ruzicska, 1968) were identified using traditional cytogenetic methods. Double trisomy of SSC1 and 10 at a blastomere stage was analysed using FISH with porcine chromosome specific probes (Zudova et al., 2003).

**Structural aberrations**

Structural chromosomal rearrangements are common in pigs and can be as frequently as 0.47% (Duocos et al., 2007). Many of these abnormalities, mainly reciprocal translocations, were discovered by I. Gustavsson and co-workers during the 1970s and 1980s. Since then, the world centre of porcine cytogenetics has shifted to INRA (Institut National de la Recherche Agronomique) in France, to the Cytogenetic Laboratory at the National Veterinary School of Toulouse. Thanks to nationally funded programmes, the laboratory carries out chromosome analyses for about 1500–2000 animals per year, analysing mainly young purebred boars for artificial insemination (AI) centres (Duocos et al., 2002, 2007, 2008). Over the period of 2002–2006 the laboratory has identified over 66 new reciprocal and Robertsonian translocations and eight inversions (Duocos et al., 2007, 2008). Systematic cytogenetic screening programmes of AI boars are also ongoing in Poland, the Netherlands and Hungary, and occasionally in Finland and Portugal, thus providing an additional input to the discovery of structural chromosomal abnormalities in pigs (Duocos et al., 2008).

**RECIPROCAL (RCP) TRANSLOCATIONS.** There are over 130 different types of rcp translocations known today (Table 7.2), and these are the most frequently reported chromosomal rearrangements in pigs. Analysis of the data indicates that the distribution of participating chromosomes is not random. Chromosomes 1, 7, 14 and 15 are involved most frequently, and SSC10, 12, 18, X and Y the least frequently. Furthermore, some chromosomal bands are more prone to breakages than others; for example, band 1q21 participates in rcp translocations with five different chromosomes, and band 7q24 is more frequently involved than six other break-points on the same chromosome (Table 7.2).

Irrespective of the chromosomes involved, rcp translocations result in the production of a variety of balanced and unbalanced gametes (McClintock, 1945; Ford and Clegg, 1969; King, 1980; Pinton et al., 2005; Villagomez and Pinton, 2008; Villagomez et al., 2008) which is considered to be the main reason for reduced fertility in translocation carriers (Gustavsson, 1990; Pinton, A. et al., 2000; Villagomez et al., 2008). Litters of such animals are 25–50% smaller, mainly because of the early loss of zygotes with chromosomal imbalances (Akesson and Henricson, 1972; Hagelton et al., 1976; King et al., 1981; Popescu and Boscher, 1982; Gustavsson,
Table 7.2. Comprehensive summary of reciprocal (rcp) translocations detected in pigs, showing the participating chromosomes, location of break-points, breeds involved and references.

<table>
<thead>
<tr>
<th>No.</th>
<th>Rcp translocation</th>
<th>Breed*</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1;3)(p;q)</td>
<td>LW, RU</td>
<td>Konovalov et al., 1987</td>
</tr>
<tr>
<td>2</td>
<td>(1;4)(q27;q21)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>3</td>
<td>(1;5)(p21;q21)</td>
<td>Polish</td>
<td>Danielak-Czech et al., 1997</td>
</tr>
<tr>
<td>4</td>
<td>(1;6)(p11;q11)</td>
<td>LW, SWE</td>
<td>Yang et al., 1992</td>
</tr>
<tr>
<td>5</td>
<td>(1;6)(p11;q35)</td>
<td>LW</td>
<td>Locniskar et al., 1976</td>
</tr>
<tr>
<td>6</td>
<td>(1;6)(q12;q22)</td>
<td>GA</td>
<td>Ducos et al., 1997a, 1998a</td>
</tr>
<tr>
<td>7</td>
<td>(1;7)(q17;q13)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>8</td>
<td>(1;7)(q2.13;q24)</td>
<td>LR, SWE</td>
<td>Gustavsson et al., 1988</td>
</tr>
<tr>
<td>9</td>
<td>(1;8)(p13;q27)</td>
<td>Y, SWE</td>
<td>Gustavsson et al., 1982</td>
</tr>
<tr>
<td>10</td>
<td>(1;9)(p;p)</td>
<td>LW</td>
<td>Ducos et al., 1997a, 1998a</td>
</tr>
<tr>
<td>11</td>
<td>(1;10)(q2.11;p15)</td>
<td>ni</td>
<td>Ravaarimanana et al., 1992</td>
</tr>
<tr>
<td>12</td>
<td>(1;11)(p23;q15)</td>
<td>LR, FIN, BL</td>
<td>Kuokkanen and Makenin, 1988; Tzocheva, 1994</td>
</tr>
<tr>
<td>13</td>
<td>(1;11)(q;q)</td>
<td>LR, SWE</td>
<td>Hansen-Melander and Melander, 1970</td>
</tr>
<tr>
<td>14</td>
<td>(1;11)(q11;q11)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>15</td>
<td>(1;11)(q24;p13)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>16</td>
<td>(1;13)(q27;q41)</td>
<td>DU</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>17</td>
<td>(1;14)(p25;q15)</td>
<td>Y, SWE</td>
<td>Gustavsson, 1984</td>
</tr>
<tr>
<td>18</td>
<td>(1;14)(q17;q21)</td>
<td>LW</td>
<td>Tarocco et al., 1987</td>
</tr>
<tr>
<td>19</td>
<td>(1;14)(q2.12;q22)</td>
<td>ni</td>
<td>Zhang et al., 1992</td>
</tr>
<tr>
<td>20</td>
<td>(1;14)(q23;q21)</td>
<td>SML</td>
<td>Golish et al., 1982</td>
</tr>
<tr>
<td>21</td>
<td>(1;15)(p25;q13)</td>
<td>LR, FIN</td>
<td>Kuokkanen and Makenin, 1988</td>
</tr>
<tr>
<td>22</td>
<td>(1;15)(q17;q22)</td>
<td>LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>23</td>
<td>(1;15)(q27;q26)</td>
<td>LW</td>
<td>Popescu et al., 1988</td>
</tr>
<tr>
<td>24</td>
<td>(1;16)(q11;q11)</td>
<td>LR</td>
<td>Förster et al., 1981</td>
</tr>
<tr>
<td>25</td>
<td>(1;17)(p11;q11)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>26</td>
<td>(1;17)(q21;q11)</td>
<td>Y, SWE</td>
<td>Gustavsson, 1984</td>
</tr>
<tr>
<td>27</td>
<td>(1;18)(q;q)</td>
<td>H</td>
<td>Villagomez et al., 1991</td>
</tr>
<tr>
<td>28</td>
<td>(2;4)(p17;q11)</td>
<td>ni</td>
<td>Gustavsson et al., 1982</td>
</tr>
<tr>
<td>29</td>
<td>(2;6)(p17;q27)</td>
<td>Y</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>30</td>
<td>(2;8)(p11;p13)</td>
<td>LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>31</td>
<td>(2;9)(q13;q24)</td>
<td>LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>32</td>
<td>(2;9;14)(q23;q22;q25)</td>
<td>ni</td>
<td>Makenin et al., 1997</td>
</tr>
<tr>
<td>33</td>
<td>(2;14)(p14;q23)</td>
<td>ni</td>
<td>Villagomez, 1993</td>
</tr>
<tr>
<td>34</td>
<td>(2;14)(p15;q26)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>35</td>
<td>(2;14)(q13;q27)</td>
<td>LF</td>
<td>Ducos et al., 1998b</td>
</tr>
<tr>
<td>36</td>
<td>(2;14)(q21;q24)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>37</td>
<td>(2;15)(p13;q24)</td>
<td>LR, FIN</td>
<td>Makenin et al., 1987</td>
</tr>
<tr>
<td>38</td>
<td>(2;15)(q28;q24)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>39</td>
<td>(2;16)(q28;q21)</td>
<td>SE</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>40</td>
<td>(2;17)(p12;q14)</td>
<td>DU</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>41</td>
<td>(3;5)(p13;q23)</td>
<td>LF</td>
<td>Ducos et al., 1998b</td>
</tr>
<tr>
<td>42</td>
<td>(3;6)(p14;q21)</td>
<td>ni</td>
<td>Villagomez et al., 2008</td>
</tr>
<tr>
<td>43</td>
<td>(3;7)(p13;q21)</td>
<td>LW</td>
<td>Gabriel-Robez et al., 1988</td>
</tr>
<tr>
<td>44</td>
<td>(3;7)(p;q)</td>
<td>Indian</td>
<td>Ducos et al., 1997a</td>
</tr>
<tr>
<td>45</td>
<td>(3;8)(q25;p21)</td>
<td>ni, LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>46</td>
<td>(3;11)(q13;p11)</td>
<td>LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>47</td>
<td>(3;13)(p15;q31)</td>
<td>LW</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>48</td>
<td>(3;15)(q27;q13)</td>
<td>LW</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>49</td>
<td>(3;16)(q23;q22)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>50</td>
<td>(4;5)(p13;q21)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>51</td>
<td>(4;6)(q21;p14)</td>
<td>P</td>
<td>Ducos et al., 2002</td>
</tr>
</tbody>
</table>

Continued
Table 7.2. Continued.

<table>
<thead>
<tr>
<th>No.</th>
<th>Rcp translocation</th>
<th>Breed$^a$</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>(4;6)(q21;q28)</td>
<td>LW</td>
<td>Ducos et al., 1998b</td>
</tr>
<tr>
<td>53</td>
<td>(4;12)(q21;q13)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>54</td>
<td>(4;13)(p15;q41)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>55</td>
<td>(4;13)(q25;q41)</td>
<td>LR, FIN</td>
<td>Makinen and Remes, 1986</td>
</tr>
<tr>
<td>56</td>
<td>(4;14)(p11;q11)</td>
<td>LW × LF</td>
<td>Popescu et al., 1984</td>
</tr>
<tr>
<td>57</td>
<td>(4;14)(q;q)</td>
<td>Indian</td>
<td>Ducos et al., 1997a</td>
</tr>
<tr>
<td>58</td>
<td>(4;15)(q;q)</td>
<td>P</td>
<td>Popescu et al., 1988</td>
</tr>
<tr>
<td>59</td>
<td>(4;15)(q25;q11)</td>
<td>LF</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>60</td>
<td>(4;16)(q25;q21)</td>
<td>LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>61</td>
<td>(5;7)(p23;p11)</td>
<td>SE</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>62</td>
<td>(5;8)(p11;p23)</td>
<td>SE</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>63</td>
<td>(5;8)(p12;q21)</td>
<td>LW</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>64</td>
<td>(5;8)(q12;q27)</td>
<td>Y, SWE</td>
<td>Gustavsson, 1984</td>
</tr>
<tr>
<td>65</td>
<td>(5;9)(p11;p24)</td>
<td>LF</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>66</td>
<td>(5;9)(p21;p13)</td>
<td>LF</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>67</td>
<td>(5;14)(q11;q)</td>
<td>H × P</td>
<td>Popescu and Tixier, 1984</td>
</tr>
<tr>
<td>68</td>
<td>(5;14)(q21;q12)</td>
<td>DU</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>69</td>
<td>(5;15)(q25;q25)</td>
<td>LR</td>
<td>Parkanyi et al., 1992</td>
</tr>
<tr>
<td>70</td>
<td>(5;17)(p12;q13)</td>
<td>Y</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>71</td>
<td>(6;8)(q33;q26)</td>
<td>GA × MS</td>
<td>Bonneau et al., 1991</td>
</tr>
<tr>
<td>72</td>
<td>(6;8)(p15;q27)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>73</td>
<td>(6;13)(p15;q41)</td>
<td>LF</td>
<td>Ducos et al., 1998b</td>
</tr>
<tr>
<td>74</td>
<td>(6;13)(p13;q49)</td>
<td>LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>75</td>
<td>(6;14)(p11;q11)</td>
<td>LW × Essex</td>
<td>Madan et al., 1978</td>
</tr>
<tr>
<td>76</td>
<td>(6;14)(q27;q21)</td>
<td>LW × P</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>77</td>
<td>(6;15)(p;q)</td>
<td>LR, FIN</td>
<td>Bouters et al., 1974</td>
</tr>
<tr>
<td>78</td>
<td>(6;15)(p15;q13)</td>
<td>P × LW</td>
<td>Bonneau et al., 1991</td>
</tr>
<tr>
<td>79</td>
<td>(6;16)(q11;q11)</td>
<td>SML</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>80</td>
<td>(7;8)(q13;q27)</td>
<td>ni</td>
<td>Ravaoarimanana et al., 1992</td>
</tr>
<tr>
<td>81</td>
<td>(7;8)(q24;p21)</td>
<td>3/4LF,1/4MS</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>82</td>
<td>(7;9)(q11;q26)</td>
<td>LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>83</td>
<td>(7;9)(q15;q15)</td>
<td>LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>84</td>
<td>(7;11)(q21;q11)</td>
<td>Y, SWE</td>
<td>Gustavsson et al., 1982</td>
</tr>
<tr>
<td>85</td>
<td>(7;12)(q11;p15)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>86</td>
<td>(7;12)(q24;q15)</td>
<td>Y, FIN</td>
<td>Kuokkanen and Makinen, 1987</td>
</tr>
<tr>
<td>87</td>
<td>(7;13)(p13;q21)</td>
<td>H</td>
<td>Gustavsson et al., 1988</td>
</tr>
<tr>
<td>88</td>
<td>(7;13)(p13;q46)</td>
<td>Polish</td>
<td>Danielak-Czech et al., 1997</td>
</tr>
<tr>
<td>89</td>
<td>(7;14)(q15;q27)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>90</td>
<td>(7;14)(q26;q25)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>91</td>
<td>(7;15)(q24;q12)</td>
<td>LW</td>
<td>Popescu et al., 1984; Konfortova et al., 1995</td>
</tr>
<tr>
<td>92</td>
<td>(7;15)(q24;q26)</td>
<td>ni</td>
<td>Makinen et al., 1997</td>
</tr>
<tr>
<td>93</td>
<td>(7;17)(q26;q11)</td>
<td>H</td>
<td>Villagomez et al., 1995a</td>
</tr>
<tr>
<td>94</td>
<td>(8;10)(p11;q13)</td>
<td>ni</td>
<td>Makinen et al., 1999</td>
</tr>
<tr>
<td>95</td>
<td>(8;12)(p11;p11)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>96</td>
<td>(8;13)(q27;q36)</td>
<td>ni</td>
<td>Ravaoarimanana et al., 1992</td>
</tr>
<tr>
<td>97</td>
<td>(8;14)(p23;q27)</td>
<td>ni</td>
<td>Ravaoarimanana et al., 1992</td>
</tr>
<tr>
<td>98</td>
<td>(8;14)(p21;q25)</td>
<td>Polish</td>
<td>Danielak-Czech et al., 1997</td>
</tr>
<tr>
<td>99</td>
<td>(9;11)(p24;q11)</td>
<td>Y, SWE</td>
<td>Gustavsson et al., 1982</td>
</tr>
<tr>
<td>100</td>
<td>(9;11)(q14;p13)</td>
<td>MS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>101</td>
<td>(9;14)(p14;q23)</td>
<td>Polish</td>
<td>Rejduch et al., 2003</td>
</tr>
<tr>
<td>102</td>
<td>(9;14)(p24;q15)</td>
<td>LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>103</td>
<td>(9;15)(p24;q13)</td>
<td>LW × LF</td>
<td>Ducos et al., 1998a</td>
</tr>
</tbody>
</table>
Table 7.2. Continued.

<table>
<thead>
<tr>
<th>No.</th>
<th>Rcp translocation</th>
<th>Breed</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>(9:17) (p24;q23)</td>
<td>X</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>105</td>
<td>(10:11)(q16;q13)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>106</td>
<td>(10:13)(q11;q11)</td>
<td>Polish</td>
<td>Danielak-Czech et al., 1996</td>
</tr>
<tr>
<td>107</td>
<td>(10:13)(q13;q22)</td>
<td>LS</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>108</td>
<td>(10:13)(q16;q21)</td>
<td>LW × P × DU × H</td>
<td>Danielak-Czech and Slota, 2007</td>
</tr>
<tr>
<td>109</td>
<td>(10:17)(q11;q21)</td>
<td>ni, LW</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>110</td>
<td>(11:13)(q;q)</td>
<td>LW</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>111</td>
<td>(11:15)(p15;q13)</td>
<td>LR, FIN</td>
<td>Henricson and Backstrom, 1964</td>
</tr>
<tr>
<td>112</td>
<td>(11:16)(p14;q14)</td>
<td>LW × P</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>113</td>
<td>(11:17)(p13;q21)</td>
<td>SE</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>114</td>
<td>(12:13)(q13;q11)</td>
<td>Minisib</td>
<td>Astachova et al., 1988</td>
</tr>
<tr>
<td>115</td>
<td>(12:14)(q13;q15)</td>
<td>DU</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>116</td>
<td>(12:14)(q15;q13)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>117</td>
<td>(12:15)(p;q)</td>
<td>LW, RU</td>
<td>Konovalov et al., 1987</td>
</tr>
<tr>
<td>118</td>
<td>(13:14)(q21;q27)</td>
<td>Y, SWE</td>
<td>Hagelton et al., 1976</td>
</tr>
<tr>
<td>119</td>
<td>(13:15)(q31;q26)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>120</td>
<td>(13:16)(q41;q21)</td>
<td>ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>121</td>
<td>(13:17)(q41;q11)</td>
<td>LF</td>
<td>Ducos et al., 1998b</td>
</tr>
<tr>
<td>122</td>
<td>(14:15)(q28;q13)</td>
<td>SE, ni</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>123</td>
<td>(14:15)(q29;q24)</td>
<td>ni, H</td>
<td>Golish et al., 1982; Gustavsson and Jonsson, 1992</td>
</tr>
<tr>
<td>124</td>
<td>(14:16)(q13;q21)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>125</td>
<td>(15:16)(q26;q21)</td>
<td>Y</td>
<td>Gustavsson et al., 1988</td>
</tr>
<tr>
<td>126</td>
<td>(15:17)(q13;q21)</td>
<td>LW</td>
<td>Ducos et al., 1998a</td>
</tr>
<tr>
<td>127</td>
<td>(15:17)(q24;q21)</td>
<td>LF</td>
<td>Ducos et al., 2002</td>
</tr>
<tr>
<td>128</td>
<td>(16:17)(q23;q21)</td>
<td>LR × DU</td>
<td>Popescu and Bosch, 1986; Astachova et al., 1991</td>
</tr>
<tr>
<td>129</td>
<td>(17:18)(q21;q11)</td>
<td>P</td>
<td>Ducos et al., 2007</td>
</tr>
<tr>
<td>130</td>
<td>(X;13)(q24;q21)</td>
<td>LR × Viet, H</td>
<td>Gustavsson et al., 1989</td>
</tr>
<tr>
<td>131</td>
<td>(X;14)(p;q)</td>
<td>ni</td>
<td>Singh et al., 1994; Neal et al., 1998</td>
</tr>
<tr>
<td>132</td>
<td>(Y;14)(q10;q11)</td>
<td>DU</td>
<td>Ducos et al., 2007</td>
</tr>
</tbody>
</table>

*Breed abbreviations: BL, Belgian Landrace; DU, Duroc; FIN, Finnish; GA, Gascon; H, Hampshire; LF, French Landrace; LR, Landrace; LS, Synthetic line; LW, Large White; MS, Meishan; ni, not indicated; P, Piétrain; RU, Russian; SE, Sino–European; SML, Synthetic male line; SWE, Swedish; Viet, local breed from Vietnam; Y, Yorkshire; X, crossbred.

1983; Gustavsson and Settgren, 1983). Notably, the phenotype and semen characteristics of translocation carriers appear normal, though in some cases degenerative changes in the testicles have been found (Chandley et al., 1972; Gustavsson et al., 1989; Villagomez, 1993).

Genome sequences and comparative maps of mammals have provided deeper insight into the molecular nature of chromosomal break-points (Kemkemer et al., 2009; Larkin et al., 2009). Some break-points, such as evolutionary break-point regions (EBRs), are evolutionarily conserved and ‘reused’ across species. The EBRs are frequently associated with structural variations, SNPs (single nucleotide polymorphisms) and retrotransposons (Kemkemer et al., 2009; Larkin et al., 2009), but also with some common fragile sites and cancer break-points (Ruíz-Herrera et al., 2006). Correlation between the known translocation break-points in pigs with mammalian EBRs is yet to be determined, though the human–pig Zoo-FISH map (Rettenberger et al., 1995; Frönicke et al., 1996; Goureau et al., 1996) does not show any clear correspondence between the rcp translocation and evolutionary synteny break-points. However, correlation has been found between the translocation break-points and fragile sites in pigs (Riggs et al., 1993; Ronne, 1995).
ROBERTSONIAN (ROB) TRANSLOCATIONS. Robertsonian (rob) or centric-fusion translocations are relatively rare in pigs, and involve four of the six acrocentric porcine chromosomes, corresponding to rob(13/17), rob(14/15) and rob(14/17) (Ducos et al., 2007). Among these, rob(13/17) is most frequent and has been described in normal pigs (Miyake et al., 1977; Alonso and Cantu, 1982; Pinton et al., 2009), intersex animals (Masuda et al., 1975) and malformed piglets (Miyake et al., 1977). Like the carriers of rcp translocations, animals with rob translocations demonstrate varying degrees of decrease in litter size (Schwerin et al., 1986; McFeely et al., 1988). Notably, the rate of unbalanced gametes in the case of rob(13/17) is higher in females (28.91%) than in males (3.21%) (Pinton et al., 2009). Besides affecting fertility in domestic pigs, rob translocations have played an important role in chromosome evolution among Suidae species (see Table 7.4, pp. 148–149).

DUPICATIONS AND DELETIONS. Duplications and deletions of whole chromosomes or chromosome arms, mainly caused by reciprocal translocations and inversions, are rarely observed in live pigs or piglets. This is probably because embryos fertilized by such games will suffer from lethal overdoses or underdoses of many genes. The few cases described so far are: deletion of a whole arm of a medium-sized bi-armed chromosome in 10-day-old embryos of cytogenetically normal parents (McFeely, 1966), and monosomy of SSC17 and double trisomy in piglets from an rcp translocation carrier (Villagomez et al., 1995a). All piglets, except one, died within a day of birth. The survivor with the double trisomy had normal body conformation and testis size, though 17.5% of the spermatozoa of this boar demonstrated acrosomal defects (Villagomez et al., 1995b).

INVERSIONS. Inversions have been described for only six porcine chromosomes, namely SSC1, 2, 4, 6, 8 and 9 (Ducos et al., 1997b, 2008; Massip et al., 2009) (Table 7.3). Recently, the first study of the meiotic segregation pattern of inversions in pigs was carried out using FISH with telomeric probes on condensed sperm nuclei (Massip et al., 2009). The estimated proportion of recombinant gametes was very low for all inversions studied, and no correlation was found between the size of the inverted fragment and the proportion of recombinant gametes. The authors conclude that inversions have very little impact on the reproductive performance in carrier pigs, which explains why so few cases have been detected so far.

FRAGILE SITES (FS). FS are heritable chromosomal loci prone to breakage under in vitro induction, and have been relatively well studied in pigs (Riggs and Ronne, 2009). So far, 11 aphidicolin-induced (Riggs et al., 1993), 24 folate-sensitive (Yang and Long, 1993) and 25 BrdU (bromodeoxyuridine)-induced (Ronne, 1995) FS have been identified, and a composite karyotype showing the location of 60 different FS has been generated (Ronne, 1995). As discussed above,

Table 7.3. Summary of pericentric and paracentric chromosome inversions found in pigs.

<table>
<thead>
<tr>
<th>Chromosomes involved</th>
<th>Inversion type</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>inv(1)(p21;q2.10)</td>
<td>Pericentric</td>
<td>Massip et al., 2009</td>
</tr>
<tr>
<td>inv(1)(p22;q11)</td>
<td>Pericentric</td>
<td>Danielak-Czech et al., 1996</td>
</tr>
<tr>
<td>inv(1)(p24;q29)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008; Massip et al., 2009</td>
</tr>
<tr>
<td>inv(1)(q12;q24)</td>
<td>Paracentric</td>
<td>Massip et al., 2009</td>
</tr>
<tr>
<td>inv(1)(q18;q24)</td>
<td>Paracentric</td>
<td>Ducos et al., 2008</td>
</tr>
<tr>
<td>inv(2)(p11;q11)</td>
<td>Pericentric</td>
<td>Massip et al., 2009</td>
</tr>
<tr>
<td>inv(2)(p11;q21)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008; Massip et al., 2009</td>
</tr>
<tr>
<td>inv(2)(p13;q12)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008</td>
</tr>
<tr>
<td>inv(2)(q13;q25)</td>
<td>Paracentric</td>
<td>Ducos et al., 2008; Massip et al., 2009</td>
</tr>
<tr>
<td>inv(4)(p14;q23)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008; Pinton et al., 2003</td>
</tr>
<tr>
<td>inv(6)(p14;q12)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008</td>
</tr>
<tr>
<td>inv(8)(p11;q25)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008</td>
</tr>
<tr>
<td>inv(8)(p21;q11)</td>
<td>Pericentric</td>
<td>Ducos et al., 2008</td>
</tr>
<tr>
<td>inv(9)(p12;p22)</td>
<td>Paracentric</td>
<td>Pinton et al., 2002</td>
</tr>
</tbody>
</table>
some FS coincide with reciprocal translocation break-points (Riggs et al., 1993; Ronne, 1995; Riggs and Ronne, 2009). Comparative analysis has identified eight putative orthologous FS between humans and pigs on SSC1p21, 1p25, 2q23, 11q12, 16q21, 18p21, Xp21 and Xq21 (Riggs and Ronne, 2009). Of these, the FS on SSC11q12 (HSA13q21) has been conserved in mammalian evolution and is shared with corresponding sites in river buffaloes, cattle, horses and rabbits (Riggs and Ronne, 2009). Interestingly, all eight orthologous FS are located in the middle of human–pig synteny segments (Frönicke et al., 1996) and are not associated with evolutionary synteny break-points.

Chromosome abnormalities associated with abnormal sexual development and infertility

INTERSEXUALITY AND SEX REVERSAL. These are relatively common conditions in pigs, with a frequency ranging from 0.1 to 0.6% in porcine populations (Pailhoux et al., 1994; Pinton et al., 2002). Such animals are usually infertile (Bösch et al., 1985), can have testes, ovaries or ovotestes (Gustavsson, 1990), and most have a normal female 38,XX karyotype (Miyake, 1973; Hunter, 1996; Hunter and Greve, 1996; Pailhoux et al., 1997, 2001a,b; Villagomez et al., 2009) with no SRY gene or any other Y chromosome markers (Thomsen and Poulsen, 1993; Switonski et al., 2002). Three intersex cases, however, have been associated with structural rearrangements on one homologue of SSC9p. One animal had a break (Tambasco et al., 1990) and the other two a paracentric inversion inv(9)(p12;p22) (Pinton et al., 2002). Using FISH with SSC9p markers, the inversion break-point was narrowed down to 76.5 cR (Pinton et al., 2002). The genetic causes of the abnormal phenotype, though, remain unknown, because the homologous region on HSA11q23 in humans contains no genes associated with intersex or sex reversal. While the genetic background to 38,XX intersexuality/sex reversal in pigs remains obscure, there is evidence that the condition may be inherited (Sittmann et al., 1980; Pailhoux et al., 1997). This is in line with the studies of exotic Vanuatu sacred intersex pigs where the trait is of maternal inheritance (Lum et al., 2006).

CHIMERISM. Chimerism in blood leucocytes or in the whole body is either rare or has not been detected in pigs. This is probably because all males and the majority of females with XX/XY leucocchimerism are phenotypically normal (Bruere et al., 1968; Somlev et al., 1970; Toyama, 1974; Christensen and Nielsen, 1980; Clarkson et al., 1995; Padula, 2005). Some chimeric females, however, have inguinal hernia and ovarian aplasia (Bosma et al., 1975). A few cases of XX/XY (Basrur and Kanagawa, 1971) and XX/XXY (Toyama, 1974) whole body chimerism have also been described.

Cytogenetics of gametes and embryos

In domestic pigs, where fertility is the trait of particular interest, a considerable amount of cytogenetic research is focused on the analysis of chromosomes in the prophase of male meiosis and in sperm, oocytes and embryos (King, 2008; Villagomez and Pinton, 2008). Porcine gamete and embryo cytogenetics has gained particular importance with the development of oocyte in vitro maturation and fertilization, and somatic cell nuclear cloning. Improved molecular methods, such as fluorescence protein immunostaining and FISH, have considerably improved the accuracy and quality of these studies in pigs (Pinton et al., 2008, 2009).

SYNAPTONEMAL COMPLEX (SC) ANALYSIS. SC analysis of the sperm of boars with translocations has been instrumental in understanding how structurally abnormal chromosomes pair in meiotic prophase, and how unbalanced gametes are produced (Gustavsson et al., 1988, 1989; Villagomez, 1993; Koykul et al., 2000; Villagomez and Pinton, 2008; Villagomez et al., 2008; Pinton et al., 2009). An additional outcome of this work has been the preparation of an SC karyotype of boar spermatocytes (Villagomez, 1993) (Fig. 7.2). An unusual, but relatively frequently observed, feature of rcp translocation carrier pigs is the phenomenon of early heterosynapsis, in which SCs are formed between non-homologous chromosomes or chromosome segments without previous homosynapsis at the early stages of pachytene (Villagomez and Pinton, 2008; Villagomez et al., 2008). Early heterosynapsis is also common in sex chromosome pairing in boar
were recently analysed in rcp(3;15)(q27;q13) and rcp(12;14)(q13;q21) in boars using sperm-FISH with painting probes for SSC1, 20, 11, 13, 18, X and Y (Bonnet-Garnier et al., 2009). Except for SSC1 in the case of rcp(3;15), no other ICEs were detected. Overall, it is suggested that as balanced constitutional rcp translocations are the most common structural chromosomal rearrangements in humans and pigs, the pig is a good model to investigate the consequences of such rearrangements on meiotic segregation in both male and female humans (Pinton et al., 2005; Massip et al., 2008; Bonnet-Garnier et al., 2009).

CYTOGENETIC ANALYSIS OF MATURE SPERM. This has been carried out using the zona-free hamster oocyte assay (Bird and Houghton, 1990) and, more recently, sperm-FISH (see above) with pig chromosome specific painting probes (Rubes et al., 1999; Kawarasaki et al., 2000; Parrilla et al., 2003; Pinton et al., 2004, 2005; Massip et al., 2008). Notably, the frequency of structural aberrations in porcine sperm is about 7.1%, and exceeds that of numerical abnormalities, which is about 4.4% (Bird and Houghton, 1990).

CYTOGENETIC ANALYSIS OF PIG OOCYTES AND WHOLE PRE-IMPLANTATION EMBRYOS. The cytogenetic analysis of pig oocytes and whole pre-implantation embryos dates back almost 40 years to when the techniques for the preparation of chromosomes from these cell types were developed (McFeely, 1967; McGaughey and Polge, 1971). Despite the early start, data on the occurrence of chromosome abnormalities in porcine oocytes and embryos are limited (Sosnowski et al., 2003; Zudova et al., 2003; Lechniak et al., 2007; Malekinejad et al., 2007; Boulanger et al., 2008). The subject, however, is of importance because the quality of in vitro produced embryos is low and this is partially attributed to the increased frequency of chromosome abnormalities in in vitro matured oocytes (Nagai et al., 2006). Recent studies show that the frequency of diploidy and aneuploidy in secondary oocytes is correlated with the duration of maturation and the age of the animals. For example, over 15% of oocytes are diploid after maturation for 40 h, while only 9% are diploid when in vitro...
maturation lasts 30–36 h (Sosnowski et al., 2003). Further, oocytes of gilts have over eight times more aneuploidies (10.8%) than the oocytes of sows (1.3%) (Lechniak et al., 2007). The chromosome most frequently involved in aneuploidies is SSC10, suggesting that, as in humans, non-disjunction more frequently involves small chromosomes.

The percentage of chromosomally abnormal in vivo produced morulae/blastocysts in pigs is about 11% (Zudova et al., 2003), while the overall incidence of embryos with abnormal chromosomal make-up is 5% (McFeely, 1967; King, 2008). Recent CGH analysis detected aneuploidies in 14.3% of 77 pig embryos studied (Hornak et al., 2009). The chromosomes most frequently involved were SSC8, 11, 12, 13, 17 and X, while aneuploidies of SSC 2, 9 and 18 were rare. Interestingly, aneuploidy of SSC10, which is frequently detected in oocytes (Lechniak et al., 2007), was not found in embryos. The studies also suggest that there are specific stages of development that are more sensitive to chromosomal imbalance than others. For example, all abnormalities related to rcp(13;14) and X-autosome translocation are eliminated during the first 21 days of development (Neal et al., 1998; King, 2008).

Finally, porcine sperm, oocytes and zygotes are the only cells where numerical variation of the whole haploid complement has been found (Hancock, 1959; Thibault, 1959; Bomsel-Helmreich, 1961; McFeely, 1966; Hunter, 1967; Moon et al., 1975; Dolch and Chrisman, 1981; McCauley et al., 2003; Zudova et al., 2003). The described cases involve: (i) the presence of more than two pronuclei in the ovum; (ii) the occurrence of heteroploid, triploid or mosaic embryos; (iii) the presence of XXXXY or XXXY sex chromosomes in tetraploid zygotes; and (iv) the detection of a few tetraploid, triploid or diploid/triploid 10-day-old blastocysts. Factors such as polyspermy, polygyny or suppression of the first cleavage division are among the reasons attributed to these abnormal chromosome numbers.

**Cytogenetics of cloned pigs**

Compared with the limited number of cytogenetic studies of porcine oocytes and embryos, even less is known about the chromosomes of cloned pigs. The need for such analysis, however, exists because chromosomal abnormalities have been found in 60–100% of the cells that were genetically modified to be used for somatic cell nuclear transfer (SCNT) (Mir et al., 2003). Therefore, part of the inefficiency of nuclear cloning, in which only 10% of embryos survive to term (Heyman et al., 2002; Campbell et al., 2007), can be attributed to chromosome aberrations (King, 2008). Telomere shortening, as shown in cattle and sheep (Shiels and Jardine, 2003; Alexander et al., 2007), is another genetic consequence of cloning by SCNT. Interestingly, just the opposite tendency has been observed in pigs. The telomeres of cloned animals are either the same size or even longer than the telomeres in the donor cells or in age-matched controls (Jeon et al., 2005; Kurome et al., 2008; Tian et al., 2009). It has been proposed that the restoration or enhancement of telomere length in cloned pigs is due to higher telomerase activity in porcine SCNT blastocysts than in nuclear donor cells and in vitro fertilization derived blastocysts (Jeon et al., 2005).

**Comparative cytogenetics and pig karyotype evolution**

The advent of banding techniques during the early 1970s facilitated chromosome identification but also provided the possibility of comparing chromosomes across closely and distantly related species. Comparison of the karyotypes of 15 living Suidae species (Table 7.4) supports the idea that karyotype evolution within this genus has been significantly influenced by centric fusions (Gustavsson, 1990; Adega et al., 2006). Furthermore, centric fusion polymorphisms in populations of wild/feral pigs have been observed in Asia, Europe (Tikhonov and Troshina, 1974, 1975) and North America (our unpublished data; Table 7.4). Interestingly, in different geographical areas, different types of polymorphisms exist. In Kyrgyzstani boars the centric fusion involves SSC15 and 16, while in European (Tikhonov and Troshina, 1978; Troshina et al., 1985) and North American animals (our unpublished data) it is predominantly rob(15;17). The translocation reduces the chromosome number to 37 in heterozygotes (Fig. 7.3) and to 36 in homozygotes.
Table 7.4. A summary of the current status of cytogenetics in the family Suidae. The evolutionary systematics of the family is derived from Lucchini et al. (2005), Robins et al. (2006) and Chen et al. (2007).

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Common name</th>
<th>Chromosome no. (2n)</th>
<th>Comments</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babyrousa (babirusas)</td>
<td>Babyrousa babyrussa</td>
<td>Babirusa</td>
<td>38</td>
<td>11 autosomes and X identical to <em>Sus scrofa</em>; 5 autosomes with no direct equivalents; Y is acro with distinct p arm</td>
<td>Bosma, 1980; Bosma and De Haan, 1981; Bosma et al., 1996; O’Brien, 2006</td>
</tr>
<tr>
<td></td>
<td>Babyrousa babyrussa celebensis</td>
<td>Sulawesi babirusa</td>
<td>38</td>
<td>Similar to babirusa</td>
<td>O’Brien, 2006</td>
</tr>
<tr>
<td>Phacochoerus (warthogs)</td>
<td>Phacochoerus aethiopicus</td>
<td>Desert warthog</td>
<td>34</td>
<td>14 autosomes and sex chromosomes similar to <em>S. scrofa</em>; Robertsonian translocation of chromosomes 13–16 and 15–17</td>
<td>Bosma, 1978; Melander and Hansen-Melander, 1980</td>
</tr>
<tr>
<td></td>
<td>Phacochoerus africanus sundervallii</td>
<td>Southern warthog</td>
<td>34</td>
<td>Similar to desert warthog</td>
<td>O’Brien, 2006</td>
</tr>
<tr>
<td>Potamochoerus (African forest pigs)</td>
<td>Potamochoerus porcus</td>
<td>Red river hog</td>
<td>34</td>
<td>12 autosomes meta/submeta; 4 acros; X the largest and Y the smallest submeta</td>
<td>Melander and Hansen-Melander, 1980; Bosma et al., 1991a; O’Brien, 2006</td>
</tr>
<tr>
<td></td>
<td>Potamochoerus larvatus</td>
<td>African bush pig</td>
<td>34</td>
<td>Similar to red river hog</td>
<td>O’Brien, 2006</td>
</tr>
<tr>
<td>Hylochoerus</td>
<td>Hylochoerus meinertzhageni</td>
<td>Giant forest hog</td>
<td>32</td>
<td>All autosomes and X are meta/submeta; Y is not analysed</td>
<td>Melander and Hansen-Melander, 1980</td>
</tr>
<tr>
<td>Sus (Eurasian swine)</td>
<td>Sus scrofa</td>
<td>European wild pig</td>
<td>36–38</td>
<td>Rob(15/17) polymorphism</td>
<td>McFee et al., 1966; Gropp et al., 1969; Rittmannsperger, 1971; Gustavsson et al., 1973; Tikhonov and Troshina, 1975; Bosma, 1976; Popescu et al., 1980; O’Brien, 2006</td>
</tr>
<tr>
<td>Species</td>
<td>Description</td>
<td>Chromosome Number</td>
<td>Remarks</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><em>S. scrofa</em></td>
<td>Asian wild pig</td>
<td>36–38</td>
<td>Rob(16/17) polymorphism</td>
<td>Tikhonov and Troshina, 1975; Troshina et al., 1985; Liu et al., 2003</td>
<td></td>
</tr>
<tr>
<td><em>S. s. domestica</em></td>
<td>Domestic pig</td>
<td>38</td>
<td>2n = 38, the same in all breeds</td>
<td>Bosma et al., 1991a,b</td>
<td></td>
</tr>
<tr>
<td><em>S. s. leucomystax</em></td>
<td>Japanese wild pig</td>
<td>38</td>
<td>Karyotype identical to domestic pig</td>
<td>Muramoto et al., 1965; Okamoto et al., 1981; Bosma et al., 1991a,b</td>
<td></td>
</tr>
<tr>
<td><em>Sus porcula salvanius</em></td>
<td>Pygmy hog</td>
<td>38</td>
<td>Karyotype similar to <em>S. scrofa</em>, centromeric regions of acros have an extra band</td>
<td>Bosma et al., 1983</td>
<td></td>
</tr>
<tr>
<td><em>Sus verrucosus</em></td>
<td>Javan warty pig</td>
<td>38</td>
<td>Karyotype similar to <em>S. scrofa</em> except for chromosomes 10 and Y</td>
<td>Bosma et al., 1991a</td>
<td></td>
</tr>
<tr>
<td><em>Sus barbatus</em></td>
<td>Bearded pig</td>
<td>38</td>
<td>Karyotype similar to <em>S. scrofa</em></td>
<td>Bosma et al., 1991a; O’Brien, 2006</td>
<td></td>
</tr>
<tr>
<td><em>Sus celebensis</em></td>
<td>Sulawesi warty pig</td>
<td>38</td>
<td>Karyotype similar to <em>S. scrofa</em> except Y</td>
<td>Bosma et al., 1991a</td>
<td></td>
</tr>
<tr>
<td><em>Sus cebifrons</em></td>
<td>Visayan warty pig</td>
<td>34</td>
<td>14 meta/submeta, 2 acros</td>
<td>O’Brien, 2006</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 7.3. Giemsa-stained karyotype (2n = 37) of a female feral pig (from Virginia, USA) carrying a heterozygous translocation rob(15/17, arrow). Source: authors’ unpublished results.

No information is as yet available for the chromosomes of two babirusa and three pig species, namely Bola Batu babirusa (*Babyrousa bolabatuensis*), Malenge babirusa (*Babyrousa togeanensis*), Heude’s pig (*Sus bucculentus*), Philippine warty pig (*Sus philippensis*) and Oliver’s warty pig (*Sus philippensis oliveri*).

Finally, it is worth mentioning that compared with human, mouse and several other mammalian species, almost no G-band similarity has been detected between human and pig chromosomes (Grafodatskii and Biltueva, 1987). The only region of putative homology was found between HSA2 and the distal part of SSC1q. This, however, as shown later by Zoo-FISH (Frönicke *et al.*, 1996; Goureau *et al.*, 1996), turned out to be a false homology.

Gene Mapping

Gene mapping studies in pigs date back to almost five decades ago, when Andresen (Andresen, 1963) initiated the basic study of pig blood groups with the initial linkage maps. The formal beginning of physical gene mapping in pigs falls into the early 1980s, and was inspired by the introduction of somatic cell hybrid technology (Goss and Harris, 1975). The 1990s added radiation hybrid panels and ISH to the genome mapping ‘toolkit’, which tremendously increased the quantity and quality of the maps generated during the last two decades. The development of resources such as cDNA (complementary DNA) and BAC genomic libraries (Rubes *et al.*, 2009) has essentially facilitated the development of markers for the construction
of cytogenetic, radiation hybrid (Faraut et al., 2009) and high-resolution BAC contig (contiguous) maps (Rogatcheva et al., 2008). Today, over 10,000 loci have been mapped in the pig genome, using both linkage mapping and different physical mapping approaches (Rogatcheva et al., 2008; Faraut et al., 2009). Of these, more than a half are aligned with the human genome sequence and establish a comparative framework between the two genomes. Most importantly, integrated physical, meiotic and comparative maps are critical for the assembly and annotation of the ultimate map of the porcine genome – the whole genome sequence.

**Meiotic or genetic linkage maps**

Since the first linkage map, built by A. Sturtevant in T. Morgan’s laboratory nearly 100 years ago, the construction of meiotic or linkage maps has become an essential genetic procedure (Archibald and Haley, 1998; Moran and James, 2005). Genetic linkage was initially revealed as a deviation from Mendel’s predictions on independent assortment. Genes that are located close to each other on the same chromosome do not assort independently in meiosis, which is explained by the linkage. The exchanges or crossovers between homologous chromosomes, which occur at meiosis during the formation of the gametes, break the linkage with a certain frequency. The proportion of recombinant genotypes is a measure of the crossing-over frequency. In general, the further apart two loci are on a chromosome the greater the chance that a crossover event will have taken place between them, and so the greater will be the proportion of recombinants. Hence, the recombination rate can be used for measuring the distance between two loci on a chromosome. There are two important requirements for the basic linkage mapping: large pedigrees, in which the relationships are known, and the availability of polymorphic genetic loci. Both these requirements were satisfied for the pig beginning in the early 1990s, and modern genomic tools now provide practically endless numbers of the polymorphic loci (Chapter 5). Physical distances between loci on DNA remain constant, and can be expressed in the number of nucleotides or other common metrics. In contrast, linkage between two genes or markers always varies depending on the type of cross, genotype, region of a chromosome, sex and other factors. Despite this well known 'volatility' in measuring recombination distances between loci, linkage maps remain a unique tool in genetic research and selection even in the post-genomic era. While physical/genomic maps allow the highest possible accuracy (Chapter 8), linkage maps provide valuable functional information which can be used widely, as described later.

During the 1980s and 1990s, significant efforts were made to develop linkage maps for humans and other species, including the pig. The difficulties in building a linkage map of the pig are usually compounded by a lack of knowledge of the relative position of alleles on homologous chromosomes, known as the phase. The major solution to this problem was calculating the likelihood ratio, which takes into account alternative phases. This procedure can be quite complex, particularly with large and complicated pedigrees. Fortunately several computer programs were developed, including **LINKAGE** (Lathrop and Lalouel, 1988) and **CRI-MAP** (Green et al.–1990), which are capable of resolving these problems in most cases. The theoretical solution of these problems and the computer programs were major advances, which eventually led to the construction of multi-locus linkage maps. For the pig, such linkage maps were built by the mid-1990s, and initially included nearly all the microsatellites available at that time (for details see Archibald and Haley, 1998). The latest porcine linkage map includes over 5000 loci, hundreds of which are known genes (www.thearkdb.org).

The measurement of the total length of sex-averaged linkage maps of all porcine autosomes plus the X chromosome made by Rohrer et al. (1996) was 2286.3 cM. The estimate of the complete genome length according to these authors is about 2470 cM, which is rather close to the observed values. Similar estimates were independently made by Marklund et al. (1996). Based on these data, Archibald and Haley (1998) assumed that the complete length of porcine linkage maps is around 2500 cM. This is roughly equivalent to ∼50 meiotic chiasmata
over the porcine genome, as one chiasma approximately corresponds to 50 cM. The latest available measurement of porcine linkage maps gives a sex-averaged estimate of 1711.8 cM (Vingborg et al., 2009). However, the lengths of several linkage groups in this study were significantly lower than the minimal possible value of 50 cM. As in many other mammalian species, there is a significant difference in recombination rate between males and females. According to measurements by Marklund et al. (1996), the average ratio of female to male recombination was estimated at 1.4:1, but this parameter varied between chromosomes as well as between regions within chromosomes. A similar ratio of 1.6:1 was observed by Vingborg et al. (2009). The true causes of distinct recombination rates in females and males, as well as differences between chromosome regions, remain unclear (Moran and James, 2005).

Linkage maps, besides their significant theoretical value in several fields of genetics, are essential for locating quantitative trait loci (QTLs), and can be used in marker-assisted selection (MAS, see Chapter 16). During the last 10–15 years, the underlying genetic architecture of critically important porcine traits such as growth, litter size, disease resistance, meat quality and behavioural characteristics has become less hidden. Further study of QTLs and their interactions will continue to be of primary interest. However, a link between phenotype and genotype for quantitative traits is usually not very strong, as the genes involved in the development of these traits do not have large effects; there may also be a significant influence of environmental factors and developmental randomness (Ruvinsky, 2009).

Nevertheless, tracking the inheritance of markers in populations whose performance is recorded should allow some of the QTLs to be identified and the architecture of the genetic control of production traits to be at least partially determined. The general principle of this approach is simple; as soon as significant associations between the inheritance of a particular chromosomal region (as determined by marker inheritance) and trait variation are detected in a sufficiently large population, this suggests the existence of a gene or genes affecting the traits in question. Efforts of numerous research groups, and particularly from Iowa State University, led to the creation of a QTL database for different agricultural animals, including the pig (http://www.genome.iastate.edu/QTLdb/notes.php). Once a QTL has been mapped to an interval between two arbitrary markers, there is a need to identify markers which are as close to that QTL as possible. These tightly linked markers are very rarely involved in meiotic recombination and will continue to frame the QTL for a long time. As outlined in this and the following chapters, syntenic or linkage relationships over short distances (<3 cM) are often conserved across species, and the pig has a typical mammalian genome in this regard. Now that the sequence of the pig genome is known, the genes in the QTL region can be examined for causative mutations.

**Somatic cell hybrid (SCH) panels and synteny mapping**

Somatic cell genetics (Kucherlapati and Ruddle, 1975; Ruddle, 1981; Kao, 1983; Faraut et al., 2009) has been the key to the development of the hybrid cell panels commonly used for generating synteny maps. Over the years, very little has changed in terms of the approaches used to develop these panels in different species.

As in most mammalian species, SCH genetics in pigs began with the mapping of enzyme genes, first to the X chromosome (Förster, 1980; Förster et al., 1980; Leong et al., 1983) and then to the autosomes (Ryttman et al., 1986, 1988). The introduction of PCR revolutionized porcine SCH analysis, and, in 1998, synteny mapping of more than 228 loci was reported (Winternational et al., 1996; Yerle et al., 1996; Fridolfsson et al., 1997). In the following few years, this figure increased more than threefold (Lahbib-Mansais et al., 1999, 2000; Davoli et al., 2000). The development of expressed sequence tags (ESTs) from porcine cDNA libraries and of sequence tagged sites (STSs) from BAC end sequences provided hundreds of new markers for synteny mapping in the early 2000s (Maak et al., 2001; Tosser-Klopp et al., 2001; Cirera et al., 2003; Robic et al., 2003; Demeure et al., 2005). The ‘popularity’ of porcine
Synteny mapping in the era when much more efficient and precise physical mapping approaches are available is probably because two porcine SCH panels (Rettenberger et al., 1994a,b; Robic et al., 1996; Yerle et al., 1996) are so well characterized that they enable the accurate assignment of markers either to a chromosomal arm or even to specific bands. The third panel (Zijlstra et al., 1996) is also fairly useful, although it is not completely informative for some chromosome pairs. Overall, while SCH mapping is gradually coming to an end in most domestic species, the method is still occasionally used in pigs. For example, the genome-wide distribution of porcine endogenous retroviruses was recently determined using several mapping approaches, including SCH analysis (Jung et al., 2010).

Radiation hybrid (RH) panels and RH mapping

Radiation hybrid (RH) mapping (see Chowdhary and Raudsepp, 2005; Faraut et al., 2009) is essentially an SCH technique with the difference that, before the fusion of cell lines, the cells of the species of interest (donor) are exposed to high (lethal) doses of X-ray irradiation that cause the fragmentation of chromosomes. The dosage may range from as low as 3000 rad (Gyapay et al., 1996; McCarthy et al., 1997) to as high as 50,000 rad (Lunetta et al., 1996). The higher the radiation dose, the higher is the resolution and mapping power of the RH panel, because more breaks and smaller chromosomal fragments are produced. Higher radiation doses also mean that more markers need to be genotyped in order to construct RH maps. It has been calculated that, in a typical mammalian genome, an irradiation dose of 3000 rad should allow the mapping of about 3000 markers, while about 12,000 markers can be mapped to unique positions on a 12,000-rad RH panel (Chowdhary and Raudsepp, 2005; Faraut et al., 2009). Traditionally, RH genotyping is carried out by PCR, and any type of markers, regardless of their polymorphism status, can be assigned to RH maps. The data are analysed with dedicated software programs, of which CONCORDE (Agarwala et al., 2000) and CARThAGene (de Givry et al., 2005) have been recently used most.

To date, four different whole genome (WG) RH panels have been constructed in the pig (Faraut et al., 2009): a commercial 3000-rad panel, T43RH (produced in Peter Goodfellow’s laboratory, commercially available through Research Genetics); a 5000-rad pig–mouse panel, SSRH (Hamasima et al., 2003); a 7000-rad pig–hamster panel ImpRH (Yerle et al., 1998); and a 12,000-rad pig–hamster panel IMNpRH2 (Yerle et al., 2002). During the past decade, these panels have been extensively used for mapping individual chromosomes and chromosomal regions, and for developing WG RH maps. The T43RH panel has led to the development of RH and comparative maps for SSC2 (Rattink et al., 2001), SSC6 (Cao et al., 2004) and porcine sex chromosomes (McCoard et al., 2002; Quilter et al., 2002). The panel was recently used to map six Toll-like receptor-related genes and generate RH and comparative maps for SSC8 and 13 (Jann et al., 2009). The 5000-rad SSRH panel allowed the construction of comparative maps for SSC8 and HSA4 (Jiang et al., 2002) and two WG RH maps. The first comprised 298 ESTs from porcine backfat tissue (Mikawa et al., 2004), and the other contained 4016 ESTs and microsatellite markers (Hamasima et al., 2008).

The most intensely exploited RH panel, however, has been the 7000-rad IMpRH. During the golden years of RH mapping at the end of the 1990s and in the early 2000s (Faraut et al., 2009), this panel allowed the mapping of large numbers of ESTs (Lahbib-Mansais et al., 1999, 2000, 2003) and orthologues of human genes (Lahbib-Mansais et al., 2000, 2003), resulting in high-resolution WG comparative maps (Rink et al., 2002, 2006). The comparative maps revealed 40 major breaks in synteny with the human genome and identified SSC2, 5, 6, 7, 12 and 14 as ‘gene rich’ and SSC11 and X as ‘gene deserts’ (Rink et al., 2006). The IMpRH panel has contributed to mapping regions containing QTLs and disease genes. For example, detailed RH maps were developed over the RN locus on SSC15q25 (Robic et al., 1999) and the ETEC F4ac receptor locus on SSC13q41 (Ren...
et al., 2009). High-resolution RH and comparative maps were also constructed for individual chromosomes, namely SSC2q and SSC16 (Shimogiri et al., 2006), SSC1 and SSC 7 (Demeure et al., 2005), SSC17 (Lahbib-Mansais et al., 2005), or chromosomal regions, namely SSC1q (Yasue et al., 2008). The panel has been instrumental in integrating the linkage and physical maps of individual porcine chromosomes, namely SSC13 (Van Poucke et al., 2001), SSC4q (Stratil et al., 2001), SSC15q21 (Robic et al., 2001) and SSC1q (Sarker et al., 2001), or the whole genome (Hawken et al., 1999).

The highest resolution porcine RH panel is the 12,000-rad IMNpRH2 (Yerle et al., 2002), which was generated to complement the 7000-rad IMpRH (Yerle et al., 1998) for fine mapping QTLs and major genes, and to support the WG BAC fingerprint map and the WG sequence assembly. Comparison of the IMpRH and IMNpRH2 framework maps for the RN region on SSC15 showed that the resolution of IMNpRH2 exceeds that of the IMpRH almost three times (Yerle et al., 2002). Since then, the IMNpRH2 has been used alone or in combination with the IMpRH to generate high-resolution maps for the SSC7 QTL region (Demars et al., 2006), refine conserved synteny between SSC12 and HSA17 (Liu et al., 2005) and orient BAC/PAC contigs over SSC6q1.2 (Martins-Wess et al., 2003a,b). Development of STS markers from BAC end sequences (BES) and their genotyping on IMpRH and/or IMNpRH2 have essentially increased the number of markers and the resolution of maps (Kiuchi et al., 2002; Meyers et al., 2005). This has led to the fine integration of RH maps with meiotic, BAC fingerprint and comparative maps for SSC2p and SSC9p (Liu et al., 2008), SSC10 (Ma et al., 2009) and SSC13 (Rogatcheva et al., 2008). Notably, the RH map of SSC13 is solely based on BES markers and has a comparative anchor at every 1.22 Mb relative to the human genome sequence. Most importantly, RH mapping of over 2000 BES markers (Meyers et al., 2005) has assisted the assembly of the WG BAC fingerprint map (Humphray et al., 2007) and the selection of a BAC minimal tiling path for the sequencing of the pig genome.

Public release of the first draft of the swine genome sequence (University of Illinois at Urbana-Champaign, 2009) will not reduce the need for accurate and high-resolution RH and integrated maps because such maps are indispensable for sequence assembly precision (Lewin et al., 2009). However, increasing amounts of sequence data and the construction of new genome analysis tools, such as SNP arrays, may introduce dramatic changes in mapping strategies (Faraut et al., 2009). The availability of sequences simplifies and expedites production of mapping markers. This creates a need for higher resolution RH panels, with an increased dose of irradiation and more clones. Alternatively, the RH panels can be genotyped on the pig 60K SNP beadchip (Ramos et al., 2009) or the Illumina 50K SNP chip (Porcine SNP Chip Consortium), which allows simultaneous genotyping of 50,000–60,000 markers on each RH clone. As a consequence, very large numbers of markers can be genotyped on an RH panel in a short time, and the construction of maps will need improved statistical analysis methods. Overall, despite the rapid development of technology and changes in the mapping paradigm, it is not likely that the need for fine and accurate RH maps in swine will disappear too soon.

Large-insert clone libraries and clone-based mapping

An essential resource for the construction of fine-scale physical genome maps is large-fragment genomic libraries cloned into yeast artificial chromosome (YAC), P1 bacteriophage, P1-derived artificial chromosome (PAC) or bacterial artificial chromosome (BAC) vectors. The first such libraries in pigs were constructed using YAC vectors (Leeb et al., 1995; Alexander et al., 1997; Rogel-Gaillard et al., 1997a), followed by a PAC library (Al-Bayati et al., 1999). It was soon realized that, although YACs can carry inserts larger than 1000kb, they also show a high degree of chimerism and are thus not the best for mapping or sequencing. PACs, in contrast, can take in about 100–300kb inserts and are stable, but are more difficult to handle than BACs. Thereafter, BACs have been the vectors of
choice to construct genomic libraries for the pig. Currently, there are seven WG BAC libraries (Table 7.5), most of which have been extensively used for positional cloning, QTL-, RH-, FISH- and clone-based mapping. Four libraries, namely CHORI-242, RPCI-44, INRA and PigEBAC, are incorporated into the integrated physical map of the pig genome (Humphray et al., 2007), which has been the backbone for the pig genome sequencing project (Schook et al., 2005). The most recent library, CHORI-247, has the highest genome coverage (15X) and was constructed to get a good representation of the pig Y chromosome sequences (B.P. Chowdhary, unpublished data).

**BAC-based chromosome and regional maps**

Since the first porcine BAC libraries became available in 1999–2000, BAC clones have found intensive use in the development of contig maps over shorter regions in the pig genome, of which SSC6, 7, 13, and 15 have been of particular interest. The first BAC contigs were constructed over the ∼2.5 Mb region around the ‘acid meat’ RN locus on SSC15q25 (Jeon et al., 2001; Robic et al., 2001). The contig allowed the identification and fine mapping of tens of genes in the region, and showed that the homologous segment on HSA2q35 shares a high degree of linkage conservation with the pig. Several BAC maps have been constructed in SSC7 for the porcine MHC (major histocompatibility complex) (Barbosa et al., 2004; Ando et al., 2005; Demars et al., 2006) and for regions containing fat-related QTLs (Sato et al., 2006; Tanaka et al., 2006). Another chromosome of interest is SSC6, because it harbours the RYR1 locus for stress susceptibility and a putative QTL for muscle growth (Martins-Wess et al., 2002). Two BAC/PAC contigs, one 1.2 Mb (Martins-Wess et al., 2002) and another 5.5 Mb (Martins-Wess et al., 2003a,b), were developed to span the RYR1 locus at SSC6q21. The two contigs have been very useful in characterizing this gene-rich region and comparing it with HSA19q13.1 (Martins-Wess et al., 2003b). Sequencing BACs from a contig on SSC6q28–q31 that spans several fat-related QTLs provided important information for comparative mapping and positional candidate genes (Lee et al., 2006). Similarly, BAC contigs have helped positional candidate gene cloning and refined the map in the region associated with neonatal diarrhoea on SSC13q31–q32 (Van Poucke et al., 2005; Joller et al., 2009). Besides these chromosomes, a contig comprising 51 BAC clones was developed over SSC2p-q13, a region of maternally imprinted QTL for backfat thickness (Rattink et al., 2001). Fine regional BAC maps have also been instrumental in the discovery of a large duplication in the KIT gene associated with dominant white coat colour in pigs (Giuffra et al., 2002), and for the discovery of a mutation in ASIP locus responsible for the black-and-tan pigmentation in Mangalitsa pigs (Drogemuller et al., 2006).

**Whole genome BAC fingerprinted contig (FPC) and comparative maps**

The construction of porcine BAC libraries and clone-based mapping have played a vital role in the international Porcine Genome Sequencing Project (http://www.sanger.ac.uk/Projects/S_sscrofa/). Recent sequencing projects for dogs (Lindblad-Toh et al., 2005), cattle (Elsik et al., 2009) and horses (Wade et al., 2009) mainly used the WG shotgun method, in which BACs, BAC contig maps and BES helped only to validate sequence assembly and to anchor sequence scaffolds to the chromosomes. In contrast, pig genome sequencing is exclusively clone-based and relies on the BAC fingerprinted contig (FPC) map (Humphray et al., 2007), the BES-based WG RH map (Meyers et al., 2005) and the comparative maps produced by aligning pig BES with the sequence map of the human genome (Meyers et al., 2005; Humphray et al., 2007; Rogatcheva et al., 2008).

The BAC FPC map was constructed by digesting BAC DNA with HindIII followed by high-throughput fingerprinting of over 260,000 BAC clones from four different BAC libraries (Table 7.5). Altogether, 267,884 fingerprints were assembled into 524 contigs, and a highly continuous BAC contig map with 15.3x coverage across the 2.7 Gb of porcine genome was constructed. The map spans 98% of the euchromatin of the 18 pig autosomes and the X and the Y chromosomes (Humphray et al., 2007). Incorporation of the Y chromosome is an outstanding difference from species such as cattle (Elsik et al., 2009) and horses.
Cytogenetic mapping using FISH

Cytogenetic maps of the pig genome

The pig was the first farm animal to which the ISH technique was applied for gene mapping. Twenty-six years ago, the porcine MHC locus was assigned to the pericentromeric region of SSC7 (Geffrotin et al., 1984; Rabin et al., 1985; Echard et al., 1986). Thereafter, a few groups initiated organized ISH experiments to expand the porcine physical gene map. By 1993, 68 genes had already been localized using the radioactive method (Chowdhary, 1998). One of the major advantages of radioactive ISH was the ability to use small heterologous (mainly human) probes because, at the onset of organized genome analysis in pigs, very few cloned porcine sequences were available.

The early 1990s witnessed the introduction of non-isotopic techniques and, since then, cytogenetic mapping in pigs has been conducted using FISH. One of the first principal uses of FISH in domestic animals was to align all individual syntenic and linkage groups to specific chromosomes. In pigs, this ‘mission’ was accomplished in about 5 years (Ellegren et al., 1993, 1994; Robic et al., 1996, 1997; Rohrer et al., 1996). An important ‘booster’ for FISH mapping was that finally there were available porcine specific large-insert clones – first cosmids (Ellegren et al., 1993; Alexander et al., 1996) and, later, YACs and BACs (see previous section). In a few studies, with the aim
of constructing comparative maps, attempts were made to FISH map small cDNA sequences (Chaudhary et al., 1997; Thomsen et al., 1998). Owing to the relatively weak signals produced, this approach did not find much use later.

The most intensive FISH mapping period in porcine genomics was at the end of the 1990s when, in just a few years, the number of cytogenetically assigned loci tripled – from 154 in 1995 (Yerle et al., 1995) to 436 in 1997 (Yerle et al., 1997). Among the 436 loci, 160 represented specific genes and the remaining 276 were Type II (polymorphic) loci. Since then, the use of FISH for porcine gene mapping has gradually decreased. In the past 10 years, only about 500 new loci have been added to the cytogenetic map. Among these, the majority are single or a few loci assignments to map disease or QTL-related genes, for example, mapping candidate genes for human obesity (Nowacka-Woszuk et al., 2008) or porcine pregnancy-associated genes (Majewska et al., 2010). Only a few studies have carried out GW mapping of larger numbers (over 50) of loci (Lopez-Corrales et al., 1999; Pinton, P. et al., 2000; Anistoroaei et al., 2004). Several FISH projects have focused on specific chromosomes, namely SSC13 (Sun, H.F. et al., 1999; Van Poucke et al., 1999, 2003; Stratil et al., 2001), SSC11 (Sun, H.S. et al., 1999) or chromosomal regions, such as the RN region on SSC15 (Tornsten et al., 1998). The FISH mapping of 19 genes to the pig X and Y chromosomes (Quilter et al., 2002) has been so far, the only detailed study of the comparative organization of porcine sex chromosomes. Compared with the horse and dog, in which cytogenetic anchors have been an instrumental part in the construction of WG RH and integrated maps (Breen et al., 2004; Raudsepp et al., 2008), the approach has found much less use in pigs (Cirera et al., 2003; Lahbib-Mansais et al., 2003). This is probably because RH mapping of thousands of BES markers (Meyers et al., 2005) and their alignment with BAC FPCs (Humphray et al., 2007) and human genome sequences (Rogatcheva et al., 2008) have already provided sufficient numbers of anchors for most of the porcine chromosomes.

High-resolution cytogenetic maps

Fibre-FISH. Fibre-FISH is an essential improvement in the resolution of chromosome mapping in which FISH is conducted on preparations of extended chromatin fibres. DNA probes are typically mapped to metaphase chromosomes at a resolution of about 5Mb (Raudsepp and Chowdhary, 2008). In contrast, fibre-FISH enables the distinction of probes separated by 1–2kb, and can be used for positional cloning, determining the transcriptional orientation of clones, or the detection of minor chromosomal rearrangements (Laan et al., 1996). In domestic animals, the technique was first used in pigs to study the organization of the MHC (Sjöberg et al., 1997) and to order subclones from the porcine erythropoietin gene (Liu et al., 1998). The studies delineated the order of individual clones, and also estimated the physical distances (kb) between them.

DNA-combing. DNA-combing is an approach that technically resembles fibre-FISH (Conti et al., 2001; Labit et al., 2008), but differs in the source and preparation of target DNA. The target may include any cloned DNA that is applied in solution, on a silanated glass slide. Gradual evaporation of the solution under a glass coverslip stretches, straightens and fixes the DNA molecules, thus making them available as targets for hybridization. The technique is best suited to constructing a contig map of a set of subclones originating from a BAC or YAC clone. The only study hitherto carried out in pigs, or in any other farm animal, relates to ordering a set of plasmid subclones derived from a lambda clone containing the porcine EPO gene (Liu et al., 1998). Distinct hybridization signals on combed DNA fibres enabled the ordering of the three subclones, and the results were in complete agreement with those obtained by fibre-FISH in the same study.

Physical mapping using flow-sorted and microdissected chromosomes

Relatively few physical maps have been constructed using porcine flow-sorted and microdissected chromosomes. In one study, chromosomal segments for SSC1p, 1q26–q2.13, 2q11–q14,
4q12–q25, 13q12–q31, 13q32–q43 and 16q21–q23 and for whole chromosomes SSC13 and SSC15 were microdissected (Chaudhary et al., 1998a). After verifying DNA origin by FISH, the microdissected material was used to construct regional libraries and to isolate microsatellite markers. Microdissected or flow-sorted chromosome-specific libraries have also been constructed for SSC1q (Sarker et al., 2001), SSC6 (Ambady et al., 1997; Grimm et al., 1997; Zhao et al., 1999), SSC8 (Wang et al., 2000a,b), SSC11 (Riquet et al., 1995), SSC13 (Davies et al., 1994; Wang et al., 2001) and SSC18 (Ellegren and Basu, 1995). The approach has produced microsatellite markers to integrate linkage or RH maps with the precise cytogenetic location.

### Comparative chromosome maps – Zoo-FISH

An important breakthrough in the construction of physical chromosome maps was the invention of cross-species FISH, also referred to as Zoo-FISH (Scherthan et al., 1994). The procedure is basically the same as for regular FISH except for some modifications, of which the most critical are the use of heterologous probes with over two times higher DNA concentration and prolonging the hybridization time (Chowdhary and Raudsepp, 2001). In a few studies, the probes originated from large insert libraries. For example, human PAC clones containing human LCAT gene cluster were used to determine the homologous region on SSC6p13 (Frengen et al., 1997). The majority of the Zoo-FISH studies, however, have used composite chromosome-specific painting probes to determine chromosomal homologies between species. The method is, therefore, also called comparative chromosome painting (Scherthan et al., 1994).

The primary goal of Zoo-FISH was, and still is, to transfer gene map information from ‘map rich’ species to ‘map poor’ species, and to study karyotype and chromosome evolution (Chowdhary et al., 1998; Chowdhary and Raudsepp, 2001; Ferguson-Smith and Trifonov, 2007). With respect to the porcine genome, Zoo-FISH experiments have been conducted between closely related species, namely the pig and other suids/Suiformes or cetartiodactyls, and between distantly related species, namely the pig and primates or carnivores (Table 7.6).

**Human–pig Zoo-FISH**

The pig genome was the first among farm animals that was examined with the whole set of human chromosome paints (Rettenberger et al., 1995; Frönicke et al., 1996). Both studies essentially arrived at the same conclusion and proposed that there are 47 segments of conserved synteny between the human and the pig.

<table>
<thead>
<tr>
<th>Table 7.6.</th>
<th>Zoo-FISH (fluorescence in situ hybridization) between pig and other mammalian species. Unidirectional painting is indicated by →, and bidirectional by ↔.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
<td><strong>Reference(s)</strong></td>
</tr>
<tr>
<td>Between suids</td>
<td>Pig → babirousa Bosma et al., 1996</td>
</tr>
<tr>
<td>Between Suiformes</td>
<td>Pig → collared and white-lipped peccary Bosma et al., 2004 Pig → Arizona collared peccary Adega et al., 2006</td>
</tr>
<tr>
<td>Between cetartiodactyls</td>
<td>Pig → cattle Schmitz et al., 1996 Sheep → pig Frönicke and Wienberg, 2001 Dromedary camel → pig Balmus et al., 2007</td>
</tr>
<tr>
<td>Between distantly related mammals</td>
<td>Human → pig Rettenberger et al., 1995; Frönicke et al., 1996; Chaudhary et al., 1998b Human ↔ pig Goureau et al., 1996 Dog → pig Biltueva et al., 2004</td>
</tr>
</tbody>
</table>
pig (Plate 2). These initial results were further refined by human–pig whole genome reciprocal chromosome painting (Goureau et al., 1996; Milan et al., 1996), and by Zoo-FISH using arm-specific paints for HSA2, 5, 6 and 16 (Chaudhary et al., 1998b). The main significance of these early studies was their contribution to the development of the porcine WG gene map.

Later on, FISH mapping of single genes (Pinton, P. et al., 2000) and the construction of high-resolution comparative RH and BAC contig maps (Rink et al., 2002; Meyers et al., 2005; Humphray et al., 2007; Rogatcheva et al., 2008) essentially confirmed the findings of Zoo-FISH, but also refined and improved it. Rink and colleagues (Rink et al., 2002) were able to add three new conserved synteny groups to the human–pig comparative map, and provisionally assigned a few others. They also identified at least 60 break-points and 90 micro-rearrangements between the human and the porcine genomes. Today, 51 conserved synteny groups and 173 conserved segments are known between the two genomes (Meyers et al., 2005), although the most comprehensive comparison will soon be available at the genome sequence level.

Zoo-FISH between pig and other domestic species

Zoo-FISH has been used to delineate segmental homologies and integrate gene maps between pigs and cattle (Schmitz et al., 1996), pigs, sheep and cattle (Frönckke and Wienberg, 2001), pigs and dogs (Biltueva et al., 2004) and, most recently, between pigs, humans, cattle and camels (Balmus et al., 2007) (Table 7.6). Notably, a high number of evolutionary rearrangements, involving fissions, fusions and inversions, have been found between the karyotypes of pig and species of Bovidae (Frönckke and Wienberg, 2001). For example, pigs and sheep share 62 conserved synteny groups, which is more than the number between pigs and humans (Meyers et al., 2005), and unusually high for a species belonging to the same mammalian order. Differences between ruminant and suid genomes have been attributed to extensive karyotype rearrangements in the Suidae family, characterized by multiple inversions, fissions and fusions (Frönckke and Wienberg, 2001). In contrast, only 53 autosomal conserved segments were found between camels and pigs (Balmus et al., 2007). It is, however, of no surprise that almost 100 homology blocks were found between the pig and dog karyotypes (Biltueva et al., 2004). The two species are evolutionarily distant and both have undergone relatively rapid karyotype evolution.

Zoo-FISH has also helped in the understanding of chromosome evolution within suid and Suiformes species. Pig chromosome painting probes have helped to refine chromosomal homologies with babirusas (Bosma et al., 1996) and two different peccary species (Bosma et al., 2004; Adega et al., 2006).

Concluding Remarks

During the ‘golden’ years of pig genome mapping in the 1990s, it seemed that the primary role of porcine cytogenetics is to be a launch pad for physical mapping, while the need for classical chromosome analysis would gradually disappear. Therefore, it is pleasant to see that pig cytogenetics has successfully survived the ‘mapping era’ and is entering the ‘sequencing era’, armed with molecular tools borrowed from physical gene mapping. Furthermore, there are clear signs that porcine cytogenetics is ready to take the next step – combining the genome sequence data with array technology and bringing chromosome analysis to submicroscopic levels. Thus, application of novel molecular approaches, together with pressing commercial needs, is expected to continue porcine chromosome analysis as an important part of agricultural programmes in many countries.

Gene mapping, in contrast, is in a transitional phase. With the availability of the WG BAC FPC map (Humphray et al., 2007), VC maps (Bio::Neos, 2010) and the draft genome sequence, the need for other maps will reduce. However, as experienced in humans and other sequenced species, the need for high-resolution meiotic and physical maps will not disappear completely. This idea was elegantly formulated in the phrase ‘every genome
sequence needs a good map’ (Lewin et al., 2009), meaning that high-resolution gene and comparative maps remain indispensable to validate sequence assembly. This need is even more pronounced for the assembly of millions of short reads produced by next and next-next generation sequencing technologies. Overall, integrated meiotic and physical maps continue to offer a well-tested resource for the analysis of genome architecture, function and evolutionary history in many species, including the pig.

Acknowledgements

We thank Dr Nandina Paria and graduate student Felipe Avila for technical help with the references.

References


Chaudhary, R., Raudsepp, T., Guan, X.Y., Zhang, H. and Chowdhary, B.P. (1998b) Zoo-FISH with microdissected arm specific paints for HSA2,5,6,16, and 19 refines known homology with pig and horse chromosomes. Mammalian Genome 9, 44–49.


Rettenberger, G., Fries, R., Engel, W., Scheit, K.H., Dolf, G. and Hameister, H. (1994b) Establishment of a partially informative porcine somatic cell hybrid panel and assignment of the loci for transition protein 2 (TNP2) and proteasome 1 (PRM1) to chromosome 3 and polyubiquitin (UBC) to chromosome 14. *Genomics* 21, 558–566.


comparison with a physical map of the porcine homologous region 15q25. *Mammalian Genome* 12, 380–386.


Introduction

The systematic characterization and mapping of the porcine genome started in the late 1980s and early 1990s following the emergence of genome research stimulated by the Human Genome Organisation (HUGO) with its goal to map and sequence the complete human genome. The EU-funded Pig Gene Mapping Project (PiGMaP) constituted the first internationally coordinated effort to map the porcine genome, and, with this, the pig was the first livestock species whose scientific community organized to completely map the pig genome (Haley et al., 1990). The initial efforts of the PiGMaP project focused on the development of genetic markers in the pig (Davies et al., 1994; Coppieters et al., 1995; Groenen et al., 1995), and the establishment of a genetic linkage map (Archibald et al., 1995) and a cytogenetic map (Echard et al., 1992; Yerle et al., 1995). In September 2003, the sequencing of the complete genome of the pig formally started with the establishment of the Swine Genome Sequencing Consortium (SGSC) by representatives from academia, government and industry (Schook et al., 2005). The necessary foundation towards this goal had been set up in the previous decade through the development of detailed linkage maps (Ellegren et al., 1994; Rohrer et al., 1994, 1996; Archibald et al., 1995) and physical maps based on a variety of methods such as somatic cell hybrids (Rettenberger et al., 1994, 1996; Yerle et al., 1996), in situ hybridization (Frönicke et al., 1996; Goureau et al., 1996; Chowdhary et al., 1998) and whole-genome radiation hybrid (RH) mapping (Yerle et al., 1998, 2002; Hawken et al., 1999). In this chapter, the focus is on comparative genome maps, genomic resources, genome variation and sequencing of the porcine genome. Other maps will be briefly discussed when relevant to these subjects. Because systematic genome mapping in the pig started with the generation of comprehensive linkage
maps, and because the bacterial artificial chromosome (BAC) map of the porcine genome has provided the template for the generation and assembly of the high-quality anchored sequence of the porcine genome (Schook et al., 2005; Humphray et al., 2007), we will also discuss these maps in more detail in this chapter.

Porcine Expressed Sequence Tag (EST) Sequencing and Clustering

Partial sequencing of expressed sequences is an efficient and economical method to rapidly acquire information about the gene content of an organism. Pioneered in the early 1990s by Craig Venter and co-workers (Adams et al., 1991), it has become an important genome resource in functional genomics (expression studies), as well as an invaluable tool for the annotation of the genome sequence and the construction of gene models. The generation of pig ESTs was initiated by several groups (Tuggle and Schmitz, 1994; Wintere et al., 1996; Tosser-Klopp et al., 1997); in particular, the Sino–Danish Pig Genome Project has boosted the number of porcine ESTs by sequencing over one million porcine ESTs derived from 97 different cDNA (complementary DNA) libraries (Gorodkin et al., 2007). The current (27 August 2010) number of pig ESTs in the Expressed Sequence Tags database (dbEST) at NCBI (the US National Center for Biotechnology Information) is 1,621,000 (http://www.ncbi.nlm.nih.gov/dbEST/index.html) with the most recent Unigene Build (No. 39, 7 August 2010; http://www.ncbi.nlm.nih.gov/unigene/estproview) comprising 53,600 different Unigene clusters. Similarly, within the Gene Index Project (Lee et al., 2005), the Computational Biology and Functional Genomics Laboratory at the Dana Farber Cancer Institute has used the porcine EST and mRNA sequences to produce a gene index of transcripts found in the pig. Within this project, more stringent clustering parameters were used, resulting in higher numbers of clusters than obtained with Unigene. Based on 1,387,573 EST and mRNA sequences, the clustering resulted in a total of 110,744 different clusters and 128,917 singleton ESTs (Release 14, March 2010).

Expression Profiling of the Porcine Transcriptome

Although there are numerous methods for the study of the expression of specific genes, for studying gene expression on a genome-wide scale (i.e. for studying all transcripts within a sample – e.g. cell, tissue, etc. – simultaneously), in essence only two fundamentally different methodologies are utilized. The most widely used technique is based on hybridization of the transcripts against an array of probes representing all the genes (Schena et al., 1995). The alternative method is to sequence a representative part of all the transcripts present within the sample being analysed and count the number of times a transcript is observed (Velculescu et al., 1995; Mardis, 2008). The first large-scale porcine expression profiling experiments were pioneered using human microarrays (Medhora et al., 2002; Moody et al., 2002), rapidly followed by the design and use of porcine-specific microarrays (Bai et al., 2003; Nobis et al., 2003). Since that time, the number of porcine-specific microarrays, including commercially manufactured arrays, designed and used has increased dramatically (see Table 8.1 for microarray data deposited at the Gene Expression Omnibus of the NCBI). The number of genes and transcripts investigated in these different studies varied dramatically, from a few hundred to thousand. However, even the studies in which the expression of a few thousand porcine genes was assayed suffered from being incomplete, as there are estimated to be 20,000–25,000 protein-coding genes in the pig genome. Moreover, there are multiple differently spliced transcripts for many of these genes. For comparison, the human genome contains about 23,500 protein-coding genes and encodes more than 140,000 different gene transcripts. Thus, it is only recently, after the completion of a draft sequence of the porcine genome, that it has become possible to design more comprehensive microarrays representing the majority of the porcine genes. For a recent more extensive review of the microarray studies performed in the pig, readers are referred to the paper by Tuggle et al. (2007).
Table 8.1. Microarray experiments deposited in Gene Expression Omnibus (GEO). The table provides information on the collection of 36 different microarray platforms used for the pig and deposited in the GEO, which is available at http://www.ncbi.nlm.nih.gov/geo/. The GEO is a public functional genomics data repository for microarray data. As of 16 March 2010, the GEO contains submissions for 148 porcine microarray expression data sets based on 38 different GEO platforms (GPLs). The table only shows the 36 platforms related to gene expression microarrays.

<table>
<thead>
<tr>
<th>GEO accession no.</th>
<th>Approximate no. of genes (max.)</th>
<th>Probe type</th>
<th>No. of samples</th>
<th>Array description</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPL336</td>
<td>870 cDNA</td>
<td>2</td>
<td>Porcine Brain Library array</td>
<td>Steven Paul Suchyta</td>
<td></td>
</tr>
<tr>
<td>GPL518</td>
<td>1,272 cDNA</td>
<td>6</td>
<td>UIUC Porcine muscle plus</td>
<td>Yewon Cheon</td>
<td></td>
</tr>
<tr>
<td>GPL1209</td>
<td>1,021 cDNA</td>
<td>16</td>
<td>Porcine 1000 embryo gene array</td>
<td>Christopher K. Tuggle</td>
<td></td>
</tr>
<tr>
<td>GPL1270</td>
<td>10 SAGE:10:NlaIII:Sus scrofa</td>
<td></td>
<td></td>
<td>GEO</td>
<td></td>
</tr>
<tr>
<td>GPL1624</td>
<td>2,423 cDNA</td>
<td>15</td>
<td>PorkChip 2,600 cDNA array</td>
<td>Kendra A. Hyland</td>
<td></td>
</tr>
<tr>
<td>GPL1881</td>
<td>12,302 oligo</td>
<td>200</td>
<td>Qiagen-NRSP-8 porcine oligo array</td>
<td>Christopher K. Tuggle</td>
<td></td>
</tr>
<tr>
<td>GPL2731</td>
<td>3,456 cDNA</td>
<td>118</td>
<td>Spotting_muscle_21OCT03</td>
<td>Laurence Liaubet</td>
<td></td>
</tr>
<tr>
<td>GPL3461</td>
<td>10,665 oligo</td>
<td>104</td>
<td>Duke Operon Porcine 10.5K Oligo Array</td>
<td>Heathier Anne Himburg</td>
<td></td>
</tr>
<tr>
<td>GPL3533</td>
<td>20,201 oligo</td>
<td>399</td>
<td>[Porcine] Affymetrix Porcine Genome Array</td>
<td>Affymetrix, Inc.</td>
<td></td>
</tr>
<tr>
<td>GPL3585</td>
<td>26,877 cDNA</td>
<td>10</td>
<td>DIAS_PIG_55K2_v1</td>
<td>Jakob Hedegaard</td>
<td></td>
</tr>
<tr>
<td>GPL3594</td>
<td>5,375 cDNA</td>
<td>6</td>
<td>DIAS_PIG_27K_v2</td>
<td>Jakob Hedegaard</td>
<td></td>
</tr>
<tr>
<td>GPL3608</td>
<td>26,877 cDNA</td>
<td>138</td>
<td>DIAS_PIG_55K3_v1</td>
<td>Jakob Hedegaard</td>
<td></td>
</tr>
<tr>
<td>GPL3707</td>
<td>10,665 oligo</td>
<td>30</td>
<td>Pig_Array_Ready Oligo set v1.0</td>
<td>Bhupinder Juneja</td>
<td></td>
</tr>
<tr>
<td>GPL3729</td>
<td>9,216 cDNA</td>
<td>84</td>
<td>AGENAE_PigGeneric2_9216</td>
<td>Karine Hugot</td>
<td></td>
</tr>
<tr>
<td>GPL3764</td>
<td>192 oligo</td>
<td>36</td>
<td>Porcine oligo micro array version 3</td>
<td>Shila Mortensen</td>
<td></td>
</tr>
<tr>
<td>GPL3970</td>
<td>4,608 cDNA</td>
<td>24</td>
<td>scag_scai Sus scrofa 4.6K triplicate array</td>
<td>Gwenola</td>
<td></td>
</tr>
<tr>
<td>GPL3971</td>
<td>1,152 cDNA</td>
<td>28</td>
<td>scag_scai Sus scrofa 1.2K mono array</td>
<td>Gwenola</td>
<td></td>
</tr>
<tr>
<td>GPL3978</td>
<td>2,854 cDNA</td>
<td>46</td>
<td>INRA Sus scrofa 4K</td>
<td>Agnes Bonnet</td>
<td></td>
</tr>
<tr>
<td>GPL4061</td>
<td>26,877 cDNA</td>
<td>10</td>
<td>SAGE:17:NlaIII:Sus scrofa</td>
<td>GEO</td>
<td></td>
</tr>
<tr>
<td>GPL4262</td>
<td>2 cDNA</td>
<td></td>
<td>SAGE:10:Sau3A:Sus scrofa</td>
<td>GEO</td>
<td></td>
</tr>
<tr>
<td>GPL4872</td>
<td>9,729 oligo</td>
<td>0</td>
<td>SBTM Microarray Laboratory Operon Pig v1.0</td>
<td>Vincent VanBuren</td>
<td></td>
</tr>
<tr>
<td>GPL4930</td>
<td>9,556 oligo</td>
<td>32</td>
<td>Intestinal epithelial crypts and villi in conventional relative to germ-free pig</td>
<td>H. Rex Gaskins</td>
<td></td>
</tr>
<tr>
<td>GPL5171</td>
<td>656 oligo</td>
<td>38</td>
<td>Pork Quality Operon 70-mer oligo array</td>
<td>Mingzhou Li</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 8.1. Continued.

<table>
<thead>
<tr>
<th>GEO accession no.</th>
<th>Approximate no. of genes (max.)</th>
<th>Probe type</th>
<th>No. of samples</th>
<th>Array description</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPL5340</td>
<td>9,944</td>
<td>cDNA</td>
<td>6</td>
<td>Porcine testis cDNA microarray 060717</td>
<td>Wen-chuan Lee</td>
</tr>
<tr>
<td>GPL5374</td>
<td>11,500</td>
<td>cDNA</td>
<td>8</td>
<td>NLI_SSC_11.5K_cDNA_V1</td>
<td>Jin Zhang</td>
</tr>
<tr>
<td>GPL5468</td>
<td>9,290</td>
<td>oligo</td>
<td>0</td>
<td>DIPROVAL - OPERON Sus scrofa AROS V1.0</td>
<td>Roberta Davoli</td>
</tr>
<tr>
<td>GPL5622</td>
<td>1,699</td>
<td>cDNA</td>
<td>48</td>
<td>SLA_Prv porcine DNA/cDNA microarray</td>
<td>Laurence Flori</td>
</tr>
<tr>
<td>GPL5948</td>
<td>Unknown (non-sequenced cDNAs)</td>
<td>cDNA</td>
<td>16</td>
<td>ASG Porcine jejunum spleen cDNA array</td>
<td>Gabriele Gross</td>
</tr>
<tr>
<td>GPL5972</td>
<td>5,375</td>
<td>cDNA</td>
<td>0</td>
<td>DJF Pig oligo 27K1 ver1</td>
<td>Jakob Hedegaard</td>
</tr>
<tr>
<td>GPL6173</td>
<td>26,877</td>
<td>cDNA</td>
<td>134</td>
<td>DJF Pig 55K v1</td>
<td>Jakob Hedegaard</td>
</tr>
<tr>
<td>GPL6472</td>
<td>23,256</td>
<td>oligo</td>
<td>66</td>
<td>Affymetrix GeneChip Porcine Genome Array probe-level</td>
<td>Nicholas Eldon Hardison</td>
</tr>
<tr>
<td>GPL6553</td>
<td>Genomic CNV</td>
<td>oligo</td>
<td>24</td>
<td>Nimblegen 385K pig array CGH</td>
<td>Jakob Hedegaard</td>
</tr>
<tr>
<td>GPL6849</td>
<td>&gt;200</td>
<td>oligo</td>
<td></td>
<td>Porcine oligonucleotide microarray version 4 (POM4)</td>
<td>Kerstin Skovgaard</td>
</tr>
<tr>
<td>GPL7151</td>
<td>17,100</td>
<td>oligo</td>
<td>0</td>
<td>SLA/Immune Response/NRSP8 Pig 70 mers Oligonucleotides 3.8K + 13.3K v1</td>
<td>Karine Hugot</td>
</tr>
<tr>
<td>GPL7435</td>
<td>19,486</td>
<td>oligo</td>
<td>16</td>
<td>Swine Protein-Annotated Oligonucleotide Microarray</td>
<td>Catherine W. Ernst</td>
</tr>
<tr>
<td>GPL7576</td>
<td>&gt;200</td>
<td>oligo</td>
<td>0</td>
<td>Porcine oligonucleotide microarray version 4 (POM4) (Condensed version)</td>
<td>Jayda Siggers</td>
</tr>
</tbody>
</table>

cDNA, complementary DNA; oligo, oligonucleotide.

The alternative method of global gene expression analysis, direct sequencing and numeration of the transcripts, circumvents the bias of only measuring those genes that have previously been identified and sequenced. The first methodology that used sequencing and counting of short tags derived from mRNA to analyse gene expression was called serial analysis of gene expression or SAGE (Velcullescu et al., 1995). Zuelke and co-workers (Zuelke et al., 2003; Blomberg and Zuelke, 2004; Miles et al., 2008) were the first to apply this technology to pigs for their study of gene expression during porcine embryonic development. In these studies, between 80,000 and 100,000 SAGE tags were
sequenced, which represented 20,000–23,000 putative porcine transcripts. However, because at that time the complete sequence of the porcine genome was not yet available, the number of different genes represented by these tags was not known. More recently, SAGE has also been used for the identification of porcine long non-coding RNA (ncRNA) (Ren et al., 2009). Although the SAGE technology circumvents the problem of the absence of sequence information for many porcine genes, and in principle allows an unbiased and sensitive analysis of gene expression, sequencing costs using traditional Sanger capillary sequencing prohibited extensive large-scale studies using this approach.

The opportunities to pursue the approach of assaying gene expression by comprehensive transcript sequencing have been changed dramatically by the recent development of so-called next-generation sequence technology. In particular, the next-generation sequencing technologies that generate millions of short-sequence reads, such as the Illumina GA, ABI SoliD and Helicos sequence technologies, are increasingly being used to study gene expression (Mardis, 2008). Next-generation sequencing of porcine mRNA (referred to as RNAseq) has been used to increase the identification of porcine transcripts and to provide further data to obtain the correct gene models for the genes in the porcine genome. To this end, a number of studies have used tissues from individuals that represent clones of the female pig (TJ Tabasco) whose genomic DNA was used for the genome assembly (see below). A large number of different tissues (Table 8.2) derived from 21 cloned individuals is available at the University of Illinois (contact L.B. Schook). RNAseq analyses were initially performed for tissues known to exhibit high levels of transcriptome complexity, i.e. the brain (J. Beever, personal communication; Uenishi et al., 2009) and the placenta (M.A.M. Groenen, L.B. Schook, O. Madsen and R.P.M.A. Crooijmans, unpublished results). In addition, RNAseq analyses have also been conducted for a number of tissues from pigs of other breeds, including testis tissues from a male wild boar (M.A.M. Groenen, L.B. Schook, O. Madsen and R.P.M.A. Crooijmans, unpublished results) and muscle and liver tissues from a Danish Landrace pig (Hornshøj et al., 2009). As a

Table 8.2. Tissue samples of clones from TJ Tabasco, a Duroc sow from Illinois. Clones were derived from TJ Tabasco using somatic cell nuclear cloning (SCNC). Ear notch fibroblasts were collected and used as a source of nuclear material for SCNC. The resulting embryos were collected and used to generate fetal fibroblasts for use in future studies and to support potential genetic modifications such as knockout or knock-in studies. Fetal fibroblasts were also used to generate TJ Tabasco clones through SCNC that were collected at various stages of fetal development. These staged fetuses have also been used to create full length cDNA (complementary DNA) libraries for RefSeq (Reference Sequence) studies.

<table>
<thead>
<tr>
<th>Origin of tissue</th>
<th>Number</th>
<th>Origin of tissue</th>
<th>Number</th>
<th>Origin of tissue</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipose</td>
<td>2</td>
<td>Adrenal glands</td>
<td>3</td>
<td>Ovary</td>
<td>9</td>
</tr>
<tr>
<td>Bladder</td>
<td>3</td>
<td>Bone</td>
<td>1</td>
<td>Heart</td>
<td>18</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>1</td>
<td>Brain</td>
<td>11</td>
<td>Hypothalamus</td>
<td>6</td>
</tr>
<tr>
<td>Brain</td>
<td>11</td>
<td>Bronchial nodes</td>
<td>1</td>
<td>Ileum</td>
<td>4</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>6</td>
<td>Cerebral cortex</td>
<td>3</td>
<td>Intestine</td>
<td>10</td>
</tr>
<tr>
<td>Colon</td>
<td>7</td>
<td>Colon fecal</td>
<td>1</td>
<td>Jejunum</td>
<td>1</td>
</tr>
<tr>
<td>Colon fecal</td>
<td>1</td>
<td>Dental pulp</td>
<td>1</td>
<td>Kidney</td>
<td>18</td>
</tr>
<tr>
<td>Duodenum</td>
<td>1</td>
<td>Duodenum</td>
<td>1</td>
<td>Liver</td>
<td>18</td>
</tr>
<tr>
<td>Ear notch</td>
<td>10</td>
<td>Ear notch</td>
<td>10</td>
<td>Lymph nodes</td>
<td>7</td>
</tr>
<tr>
<td>Eye</td>
<td>5</td>
<td>Eye</td>
<td>5</td>
<td>Mammary (gland)</td>
<td>5</td>
</tr>
<tr>
<td>Fat</td>
<td>5</td>
<td>Fat</td>
<td>5</td>
<td>Medulla</td>
<td>2</td>
</tr>
<tr>
<td>Fetal liver</td>
<td>1</td>
<td>Fetal liver</td>
<td>1</td>
<td>Muscle</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 8.2. Tissue samples of clones from TJ Tabasco, a Duroc sow from Illinois. Clones were derived from TJ Tabasco using somatic cell nuclear cloning (SCNC). Ear notch fibroblasts were collected and used as a source of nuclear material for SCNC. The resulting embryos were collected and used to generate fetal fibroblasts for use in future studies and to support potential genetic modifications such as knockout or knock-in studies. Fetal fibroblasts were also used to generate TJ Tabasco clones through SCNC that were collected at various stages of fetal development. These staged fetuses have also been used to create full length cDNA (complementary DNA) libraries for RefSeq (Reference Sequence) studies.
further tool for future functional genomics studies, and to improve gene models in the pig, several groups initiated sequencing of cDNA libraries that had been enriched for full-length cDNAs. These include a normalized full-length cDNA library constructed and sequenced from a pool of 11 different tissues (kidney, liver, lymph node, cerebellum, placenta, colon, hypothalamus, frontal lobe, spleen, small intestine and lung; M.A.M. Groenen, L.B. Schook and R.P.M.A. Crooijmans, unpublished results) and the sequencing of clones derived from 28 full-length-enriched cDNA libraries from 25 different porcine tissue and cell lines, including brain, ovary, colon and hypothalamus from clones of TJ Tabasco (Uenishi et al., 2009).

The availability of the sequence of full-length transcripts of the porcine genome will greatly facilitate the correct identification of the transcription start sites (TSS) of the porcine genes. This is not only extremely important to obtain correct gene models, but also for the precise localization of the porcine promoters. The present algorithms designed to predict regulatory elements within promoters have often proven unsatisfactory to a large extent because they assume correct identification of the TSS of the genes being compared, something that often is not the case. Although full-length cDNA sequences already provide the necessary information for correct assignment of TSS, this is further enhanced by the use of cap analysis gene expression (CAGE; Shiraki et al., 2003).

As for RNAseq, the combined use of CAGE and next-generation sequencing (also referred to as deepCAGE) adds a further dimension to the methodology, and in particular enables the identification of less frequently used alternative promoters and tissue specific promoters (de Hoon and Hayashizaki, 2008). In pigs, deepCage has been done on placenta, testis (M.A.M. Groenen, L.B. Schook and R.P.M.A. Crooijmans, unpublished results) and macrophages (D.A. Hume and A.L. Archibald, unpublished results).

Non-coding and Regulatory RNAs in the Porcine Genome

Transcripts that do not encode proteins are referred to as non-protein-coding or non-coding RNAs (ncRNAs). A key question arising from the observation of widespread transcription is whether these transcripts are biologically functional. Increasingly, several specific classes of ncRNAs have been shown to be involved in a wide spectrum of regulatory functions, and an increasing number of such ncRNAs are being discovered in the genomes of metazoans (Mattick, 2009). The best known and most studied class of ncRNAs are the microRNAs (miRNAs), which have been shown to be involved in the regulation of many genes. In addition, numerous other classes of short RNAs, such as Piwi protein-interacting RNA (piRNA) and small nucleolar RNA (snoRNA), and RNAs derived from the XIST locus on the X chromosome (xiRNAs), have been described (Filipowicz et al., 2008). Likewise, long regulatory intergenic ncRNAs are increasingly being studied, although it is not yet clear to what extent these are functional (Louro et al., 2008). Wernersson et al. (2005) analysed genomic sequence data representing an estimated 50% of the porcine genome for the presence of conserved miRNA sequences. By comparison with the sequences present in the miRNA hairpin database (Griffiths-Jones, 2004), a total of 51 mature miRNA sequences could be identified. To identify novel pig ncRNAs, rather than only ncRNAs that are conserved in other species, Seemann et al. (2007) constructed a bioinformatics pipeline, EST2ncRNA, and searched within the 1 million porcine ESTs for potential functional ncRNAs. Within the 48,000 EST contigs (contiguous sets of overlapping DNA segments) and 73,000 singleton ESTs, they identified 1399 different ncRNA predictions. Based on 92 different non-normalized cDNA libraries, the highest number of ncRNA predictions was derived from developmental and neuronal tissues. This high number does not appear to be caused by the complexity of the libraries, as only a small number of ncRNAs were observed in the testis, a tissue normally considered to be among the tissues with the highest number of expressed genes.

In a preliminary analysis of chromosomes 7 and 14, at the time when sequence coverage of these two chromosomes exceeded 95%, over 850 potential miRNAs were identified, as well as an additional 3000 putative ncRNAs (J. Gorodkin and M. Fredholm, personal
communication). Although the false discovery rate of the ncRNA prediction programs used in this analysis (RNAz and RNAmicro) is relatively high, this nevertheless provides a good indication regarding the abundance of such sequences in the porcine genome. The fact that these two chromosomes together comprise 231 Mb of sequence indicates that the porcine genome would be predicted to contain over 37,000 ncRNAs and over 1000 miRNAs.

**Porcine Linkage Maps**

The first coordinated efforts to better understand the pig genome focused on the generation of linkage maps based on polymorphic DNA markers. The major contributors to this effort have been international collaborative projects based in Europe – the PiGMaP consortium (Archibald et al., 1995) and the related Nordic collaboration (Ellegren et al., 1994; Marklund et al., 1996), and the efforts of the USDA Meat Animal Research Center (Rohrer et al., 1994, 1996). These combined efforts resulted in the placement of over 1500 polymorphic genetic markers on the porcine linkage map. However, integration of all the linkage information from the different studies into a single consensus map was not very practical and has never been attempted. The majority of the markers on the pig linkage map are microsatellite markers, short sequences comprising 1–4 bp direct repeats of at least eight copies. Because of the abundance of such sequences in the genomes of vertebrates and many other eukaryotes, these have been the markers of choice for the construction of comprehensive genome-wide linkage maps during the 1990s. In the years following the publication of the first porcine linkage maps, the growth of such linkage maps has slowed. Nevertheless, the number of markers added to these maps has steadily increased through the further mapping of additional new microsatellites and other types of mostly anonymous polymorphic DNA markers, including amplified fragment length polymorphisms (AFLPs) (Rothschild, 2004) and single nucleotide polymorphisms (SNPs) (Vingborg et al., 2009). Currently, over 5000 loci including several hundred genes are located on the different maps (www.thearkdb.org). Gradually, efforts to increase the number of markers on the porcine map have shifted towards the use of physical maps such as the RH (radiation hybrid panel) maps (Yerle et al., 1998, 2002; Hawken et al., 1999) and the BAC maps (discussed in the next section).

**A Highly Continuous BAC Map of the Porcine Genome**

Physical maps based on bacterial artificial chromosomes (BACs) have provided the essential framework for the majority of eukaryotic genomes that have been sequenced to date (Green, 2001). The stability of these clones, their size and their relative ease of use in a standard molecular biology laboratory have been key to their successful application for physical mapping. Five different porcine BAC libraries are available (Table 8.3) providing an estimated 38x coverage of the porcine genome.

<table>
<thead>
<tr>
<th>Library</th>
<th>Fingerprinted clones</th>
<th>Genome coverage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHORI-242</td>
<td>103,758</td>
<td>6.7</td>
<td><a href="http://bacpac.chori.org/library.php?id=124">http://bacpac.chori.org/library.php?id=124</a></td>
</tr>
<tr>
<td>RPCI-44</td>
<td>61,281</td>
<td>3.8</td>
<td>Fahrenkrug et al., 2001</td>
</tr>
<tr>
<td>PigE</td>
<td>73,866</td>
<td>4.2</td>
<td>Anderson et al., 2000</td>
</tr>
<tr>
<td>INRA</td>
<td>28,467</td>
<td>1.5</td>
<td>Rogel-Gaillard et al., 1999</td>
</tr>
<tr>
<td>KPN</td>
<td>361</td>
<td>0.02</td>
<td>Jeon et al., 2003</td>
</tr>
<tr>
<td>Other</td>
<td>151</td>
<td>0.01</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>267,884</td>
<td>16.2</td>
<td>Humphray et al., 2007</td>
</tr>
</tbody>
</table>
(Rogel-Gaillard et al., 1999; Anderson et al., 2000; Suzuki et al., 2000; Fahrenkrug et al., 2001; Jeon et al., 2003). The development of a porcine BAC contig physical map by means of fingerprinting the individual clones (Schein et al., 2004) from two BAC libraries (RPCI-44 and CHORI-242) both produced by Pieter J. de Jong, one made at the Roslin Institute (Anderson et al., 2000), and one produced at INRA (Institut National de la Recherche Agronomique) (Rogel–Gaillard et al., 1999), was undertaken through a coordinated international effort as a precursor to the pig genome sequencing project. The total number of BACs fingerprinted was 267,884, representing 16.2× depth of the porcine genome based on its estimated size of 2.6–2.7 Gb (Schmitz et al., 1992; Rogatcheva et al., 2008). The fingerprinted BACs were assembled into 172 contigs covering an estimated 98% of the porcine genome (Fig. 8.1) (Humphray et al., 2007).

This BAC fingerprint map constitutes the most highly contiguous BAC map of any mammalian genome constructed so far, with one single complete chromosome (Ssc13) represented by a single contig. The map is accessible through the Sanger web site (http://www.sanger.ac.uk/Projects/S_scrofa/mapping.shtml).

To enable the integration of the BAC and genome sequence maps, a total of 620,089 BAC end sequences (BES) were generated from 335,463 BACs with an average Q20 length of 635 bp. These sequences comprised approximately 15% of the porcine genome and enabled the construction of a detailed human–porcine comparative map (Humphray et al., 2007; and discussed below).

Genome Sequencing and Sequence Assembly

The first region of the pig genome subjected to systematic sequencing was the major histocompatibility complex (MHC) region around the centromere on chromosome 7 (SSC7). Initial sequencing concentrated on regions containing the classical MHC class I genes SLA1, 2, 3, 4, 5, 9 and 11 (Renard et al., 2001). A contig consisting of 15 BACs derived from the INRA BAC library (Rogel-Gaillard et al., 1999) was sequenced, resulting in a contiguous sequence of 307 kb in which 11 genes were identified. Subsequent sequencing of those regions flanking this contig resulted in a further 670 kb of porcine MHC region sequence and the identification of an additional 36 genes (Shigenari et al., 2004, Ando et al., 2005). The sequence of the complete porcine MHC region on both sides of the centromere on chromosome 7 containing the class I, II and III MHC genes was published in 2006 (Renard et al., 2006). This 2.4 Mb sized region, excluding the centromere, contains 151 genes, of which 123 could be identified as orthologous to human MHC genes.

The Sino-Danish pig sequencing consortium, although primarily focusing on sequencing large numbers of ESTs (Gorodkin et al., 2007), generated 3.84 million shotgun sequences derived from five different pig breeds: Hampshire, Yorkshire, Landrace, Duroc and Erhualian (Wernersson et al., 2005). The number of sequences per breed varied from 257,000 for the Chinese Erhualian breed to 1.2 million for the Yorkshire (Large White) breed. The 3.84 million sequences represent an estimated 0.66× coverage of the porcine genome. The low coverage and high diversity of the animal material used has prevented any meaningful assembly of the sequences, and the data are primarily a resource for SNP discovery (discussed below).

Based on these sequences, the repetitive sequence content of the porcine genome was estimated to be around 34% (Table 8.4), which is similar to that of the mouse but lower than that of other mammalian genomes, such as the human, dog and cow. This is consistent with the smaller size of the porcine genome of 2.56 billion bp as estimated from build 9 (July 2009, discussed below). There seems to be a clear correlation between the estimated genome size and the repeat content (Table 8.4), with the exception of the dog genome, which, from the mammals whose genome has been sequenced, is the smallest genome but has a repeat content similar to that observed in humans. Like the cow genome, the pig genome seems to have a relatively low number of LTR (long terminal repeat) elements compared with all the other mammals, as well as a relatively low number of
long interspersed elements (LINEs), a feature that the pig shares with the mouse.

The integrated highly contiguous physical map of the pig genome (Humphray et al., 2007) was used as a template for sequencing the porcine genome. Sequencing was primarily undertaken at the Wellcome Trust Sanger Institute at Hinxton, UK, with a hierarchical shotgun sequencing approach using the BAC clones from the BAC map. In order to minimize the number of BACs required to cover the complete porcine genome, BACs were identified in a series of iterative rounds. Briefly, pairs of BACs selected from the minimal tile path (i.e. the path through the overlapping clones in the physical map that represents greatest genome

**Fig. 8.1.** Bacterial artificial chromosome (BAC) map of the porcine genome. The positions of the BAC contigs (contiguous sets of overlapping DNA segments) are indicated by vertical bars bars adjacent to the karyotype of the individual porcine chromosomes.
The current assembly (at the time of writing) of the draft pig genome sequence (build 9; Sscrofa9) is accessible in the Ensembl genome browser (http://www.ensembl.org/Sus_scrofa/Info/Index). This assembly was established from the BAC clone derived sequences as available in April 2009, and covers about 89% of the pig genome. The Ensembl team established the first gene build for the pig as follows: (i) starting with 9277 pig proteins from RefSeq (the NCBI Reference Sequence database, excluding all the predicted models) and UniProt (Universal Protein Resource) sequences of which 7144 aligned uniquely; (ii) and also starting with 19,384 pig cDNA sequences (after predicted cDNAs were removed), of which 11,930 cDNAs met the criteria (i.e. aligned with identity ≥97% and coverage ≥90%); (iii) from 1,532,435 pig ESTs of which 898,859 ESTs passed the same score cut-off (i.e. aligned with identity ≥97% and coverage ≥90%); and (iv) ~130,000 additional proteins, mostly from other mammals, and ~20,000 human Ensembl models, of which around 50% aligned with >90% coverage. All
the gene predictions were merged, giving priority to pig-specific proteins to give a final gene set of 17,493 genes and 520 pseudogenes.

The next assembly (Sscrofa10), which will form the basis for the publication of a draft pig genome sequence, will incorporate not only sequence data from BAC clones that extend the coverage of the genome, but also whole genome shotgun sequence data generated by the Korea National Livestock Research Institute. The porcine genome will be further improved by the incorporation of whole genome shotgun reads representing 24-fold genome coverage and derived from TJ Tabasco, generated using next-generation sequencing performed by the Beijing Genome Institute (BGI) using the Illumina GA sequencing platform, and consisting of 44-bp paired end reads totalling 66.6 Gb of sequence data (Jun Wang, personal communication).

**Comparative Genomics**

As the number of genes mapped across the genomes of different species increased in the late 1980s, it quickly became apparent that the homologues of genes that co-localized on the same chromosome in one species were often also co-localized in other species; this phenomenon was referred to as ‘conserved synteny’ (Nadeau, 1989). Conserved syntenies were defined as homologous segments in different organisms composed of at least two pairs of homologous genes located on the same chromosome, regardless of gene order. As the number of mapped genes increased further and, in particular, after the characterization of the complete genome sequence of multiple species, the definitions ‘conserved synteny block’ and ‘conserved synteny segment’ were often used (Waterston et al., 2002), although this nomenclature has not been used uniformly in genome sequencing papers published in the last decade. Within this chapter, the term ‘conserved block’ is used for regions that are on the same chromosome between species (e.g. pig chromosome 8 and human chromosome 4) and that, at the resolution used, are not interrupted by regions homologous to other chromosomes. Where the homologous sequences and/or genes are in the same order in the two species, they are referred to as ‘conserved segments’.

Conserved synteny between the porcine and other mammalian genomes, in particular that of humans, has already been used for almost 20 years to predict the location of genes and to identify candidate genes for important traits in the pig. The first example where this approach was used successfully was the identification of the RYR1 gene as the gene for the halothane locus on porcine chromosome 6 (MacLennan et al., 1990; Fujii et al., 1991; Otsu et al., 1991). Other well-known examples where comparative mapping was successfully used to identify the candidate gene for the trait under investigation in the pig include the identification of a mutation in the PRKAG3 gene (RN locus) responsible for the excess glycogen content in pig skeletal muscle (Milan et al., 2000), and the identification of an SNP in the IGF2 gene as the causal variation underlying an imprinted quantitative trait locus (QTL) for backfat and muscle growth on porcine chromosome 2 (Van Laere et al., 2003).

The development of a porcine–human comparative map accelerated with the increased efforts to map genes and ESTs (Fridolfsson et al., 1997; Wintero et al., 1998; Rink et al., 2002) on the porcine linkage maps (Ellegren et al., 1993; Johansson et al., 1995) and RH maps (Hawken et al., 1999; Robic et al., 1999; Lahbib-Mansais et al., 2000). The first comprehensive comparative maps between the porcine and human genomes were obtained by bidirectional chromosome painting by means of fluorescent in situ hybridization using individual flow-sorted chromosomes (Rettenberger et al., 1995; Goureau et al., 1996). These results revealed the presence of at least 37 conserved synteny blocks, which was somewhat lower than observed for the bovine–human comparative maps (Hayes, 1995; Solinas-Toldo et al., 1995). Although orthologous genes mapped in both humans and pigs showed that several of these blocks consisted of multiple segments, the mapping resolution available at that time did not permit estimates regarding the number of conserved synteny segments between the human and porcine genomes. The first high-resolution
A porcine–human comparative map that was able to identify conserved synteny segments within these larger conserved synteny blocks was derived from the RH mapping of 1058 ESTs (Rink et al., 2002). Using this approach, Rink et al. (2002) were able to identify at least 60 evolutionary break-points and 90 micro-rearrangements between the genomes of humans and pigs. The availability of a high-resolution physical map based on fingerprinted BACs (Humphray et al., 2007), and in particular the availability of the end sequences (BES) of many of these BACs, allowed the development of even higher resolution human–porcine comparative maps (Meyers et al., 2005; Humphray et al., 2007). Meyers et al. (2005) used these resources to add 2068 BES to the RH map, thus further refining the resolution of the comparative map; they were able to identify 51 conserved synteny groups and 173 conserved synteny segments between the genomes of humans and pigs. Using the definition of conserved synteny blocks presented here, the total number of conserved synteny blocks reported in that study is 65. Completion of the porcine genome sequence will further increase the resolution of the comparative map between the human and the pig. Comparison of the human genome sequence with the currently available pig genome sequence (build 9; Sscrofa9), which covers approximately 89% of the porcine genome, has been further extended by searching Sscrofa9 with 10 kb segments of the human genome sequence using the algorithm blat (blast-like alignment tool; Kent, 2002). The resulting comparative map (Plate 3) reveals additional evolutionary break-points as well as an additional number of (small) conserved synteny blocks not observed at previous resolutions, bringing the total number of conserved synteny blocks to 70 and the number of conserved synteny segments to 194. Because comparing the next pig genome assembly (Sscrofa10) against other maps, including the linkage, RH and comparative maps, will form part of the quality checks on the draft pig genome sequence before its publication, the putative evolutionary break-points revealed by these analyses will be regions that merit careful checking. However, the modest number of conserved synteny segments suggests that the current assembly of the genome sequence data (i.e. build 9; Sscrofa9) is a good assembly.

**Variation in the Porcine Genome**

Establishing the complete sequence of the genome of any given species is extremely important as it allows the analysis of the complete gene content of that organism, thus facilitating the dissection of the molecular basis of all aspects of the functioning of that particular species. Comparing the gene content and the evolution of genes and gene families between closely as well as distantly related species provides further insights into understanding the molecular instructions that contribute to the development and functioning of a given organism. Equally important is the characterization of the natural variation between the genomes of different individuals within a species. Genetic variation is central to the variation observed in traits within any given population, as well as a key that allows specific individuals to adapt to changes in the environment and eventually supports the emergence of new species. Furthermore, in pig breeding, the available genetic variation within the different pig populations has allowed the development of specific breeds and lines, each with specific characteristics (traits), and supplies the raw material from which further improvement in productivity, health and welfare can be built by the breeding industry. In this respect, the establishment of the genome sequence of a single individual is just the start, but it provides the necessary framework and reference to further examine the organization of the genome of a large number of individuals.

Variation within the genome involves changes of single nucleotides (SNPs), variation of repetitive sequences, e.g. at mini- and micro-satellites, and even variation in the numbers of regulatory sequences and genes (copy number variation or CNV). Some of this variation has already been used in the past to develop polymorphic markers to construct the necessary linkage maps or to study specific genetic variation (Chapters 2–5). Probably the first large-scale identification of genetic
variation in the pig at a genomic scale was the development of large numbers of microsatellite markers that were used to construct linkage maps of all the pig chromosomes in the 1990s (Rohrer et al., 1994, 1996; Archibald et al., 1995). Although useful as genetic markers, microsatellites, in general, do not contribute greatly to phenotypic variation. The majority of microsatellite markers developed in the pig were based on the (CA)_n motif, the most frequent type of microsatellite found in vertebrate genomes. The frequency of (CA)_n microsatellites was found to be similar in the porcine and human genomes, with approximately one every 40 kb for (CA)_n microsatellite loci with more than 12 repeats (van Wijk et al., 2007). These results were confirmed by a count based on genome build 9, representing 89% of the genome (Table 8.5).

Over the last 5–10 years, as a direct result of the improved automation of SNP genotyping and the abundance of this type of marker, the emphasis in genetic studies has quickly shifted towards the identification and use of SNPs. Furthermore, SNP variation is thought to underlie most of the observed phenotypic variation, providing an even stronger stimulus for the discovery of this type of variation. Numerous studies have focused on the identification of SNPs in specific genes, particularly in relation to candidate gene approaches for the analysis of quantitative traits (Rothschild et al., 2007). The first studies to systematically identify SNPs at a large scale used sequencing of PCR amplified fragments and focused on porcine genes (Fahrenkrug et al., 2002), and a QTL region identified on the short arm of chromosome 2 (Jungerius et al., 2003). The estimated SNP frequencies in the haploid genome based on these studies were one SNP every 609 bp (Fahrenkrug et al., 2002) and one SNP every 357 bp (Jungerius et al., 2003), respectively. The SNP frequency of one SNP every 357 bp derived from random genomic sequences is twofold higher than that found in the genome of Bos taurus (The Bovine HapMap Consortium, 2009) and twofold lower than in the chicken genome (Wong et al., 2004).

The large-scale EST sequence data have also been used for SNP mining (Uenishi et al., 2004, 2007; Panitz et al., 2007; Vingborg et al.,

### Table 8.5. Dinucleotide microsatellite count per chromosome based on pig genome assembly build 9 (Sscrofa9). Summary of distribution of TG, TA and TC type microsatellite repeats in the porcine genome. The TG type of repeat underlies the majority of the microsatellite markers used for linkage mapping in the pig.

<table>
<thead>
<tr>
<th>Chromosome no.</th>
<th>Chromosome size (bp)</th>
<th>TG&gt;12</th>
<th>TC&gt;12</th>
<th>TA&gt;12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>295,534,758</td>
<td>6,342</td>
<td>712</td>
<td>1,977</td>
</tr>
<tr>
<td>2</td>
<td>140,138,545</td>
<td>3,123</td>
<td>344</td>
<td>829</td>
</tr>
<tr>
<td>3</td>
<td>123,604,833</td>
<td>3,071</td>
<td>377</td>
<td>621</td>
</tr>
<tr>
<td>4</td>
<td>136,259,999</td>
<td>3,240</td>
<td>388</td>
<td>893</td>
</tr>
<tr>
<td>5</td>
<td>100,522,023</td>
<td>2,351</td>
<td>277</td>
<td>611</td>
</tr>
<tr>
<td>6</td>
<td>123,310,224</td>
<td>3,021</td>
<td>368</td>
<td>643</td>
</tr>
<tr>
<td>7</td>
<td>136,414,115</td>
<td>3,266</td>
<td>403</td>
<td>803</td>
</tr>
<tr>
<td>8</td>
<td>119,990,724</td>
<td>2,692</td>
<td>301</td>
<td>809</td>
</tr>
<tr>
<td>9</td>
<td>132,473,644</td>
<td>3,094</td>
<td>334</td>
<td>702</td>
</tr>
<tr>
<td>10</td>
<td>66,741,983</td>
<td>1,664</td>
<td>195</td>
<td>346</td>
</tr>
<tr>
<td>11</td>
<td>79,819,449</td>
<td>1,847</td>
<td>170</td>
<td>433</td>
</tr>
<tr>
<td>12</td>
<td>57,436,398</td>
<td>1,340</td>
<td>146</td>
<td>232</td>
</tr>
<tr>
<td>13</td>
<td>145,240,356</td>
<td>3,187</td>
<td>363</td>
<td>929</td>
</tr>
<tr>
<td>14</td>
<td>148,515,193</td>
<td>3,611</td>
<td>441</td>
<td>1,028</td>
</tr>
<tr>
<td>15</td>
<td>134,546,158</td>
<td>3,029</td>
<td>358</td>
<td>843</td>
</tr>
<tr>
<td>16</td>
<td>77,440,712</td>
<td>1,873</td>
<td>190</td>
<td>435</td>
</tr>
<tr>
<td>17</td>
<td>64,400,393</td>
<td>1,575</td>
<td>212</td>
<td>426</td>
</tr>
<tr>
<td>18</td>
<td>54,314,914</td>
<td>1,289</td>
<td>107</td>
<td>278</td>
</tr>
<tr>
<td>19</td>
<td>125,876,345</td>
<td>3,066</td>
<td>452</td>
<td>1,144</td>
</tr>
<tr>
<td>Total</td>
<td>2,136,704,421</td>
<td>52,681</td>
<td>6,138</td>
<td>13,982</td>
</tr>
</tbody>
</table>
The majority of these SNPs, however, have not been deposited into NCBI’s SNP database (dbSNP), although they are available through a number of EST specific databases. Panitz et al. (2007) described the identification of 7900 candidate SNPs using a data set of over 0.8 million ESTs (Gorodkin et al., 2007). Around 3900 of these SNPs were included on a 7K Illumina iSelect beadchip (C. Bendixen and A. Archibald, personal communication), and genotyping results on a wide variety of breeds indicated a validation frequency of 83%. Similar validation frequencies were observed for SNPs derived from the PEDE (Pig Expression Data Explorer) and TGI (The Gene Index Project at the Dana Farber Cancer Institute) databases (M.A.M. Groenen, unpublished results). SNP identification based on a comparison of all the available porcine genomic sequences in GenBank (the US National Institutes of Health genetic sequences database, available at NCBI) resulted in the identification of 6374 SNPs, for which both variants were observed at least twice (Kerstens et al., 2009). Similar to the putative SNPs derived from the EST sequence data, the conversion rate (i.e. the proportion of the putative SNPs that could be validated and shown to be truly polymorphic) of these SNPs was 82%.

As EST data are known to also contain mitochondrial DNA (mtDNA) sequences, this resource was examined for any variation in mtDNA-derived sequences (Scheibye-Alsing et al., 2008). Although the authors described 374 putative SNPs, their validation results indicate that the majority were false positives, and that the number of true, reliable mtSNPs with high conversion rates that were identified was 112. Conversion rates for SNPs obtained from PCR-based re-sequencing efforts based on BES (A. Archibald, unpublished results) or from the sequence comparison of high-quality sequenced genome sequences (Amaral et al., 2008) were generally significantly higher (>95%). More recently, several large SNP discovery projects were initiated, each of which used a different sequence methodology and strategy. Denis Milan and co-workers (INRA, France) used traditional Sanger sequencing to generate a total of 1 million sequences of Large White, Landrace, Piétrain, Iberic, Götttingen, Meishan and Wild Boar. In total, over 55,000 high-quality SNPs were identified with a conversion rate above 95% (D. Milan, personal communication). Using next-generation sequencing on a Roche GS-FLX sequencer, Wiedmann et al. (2008) identified more than 100,000 SNPs with a conversion rate of more than 91%. The other two studies used Illumina’s GA next-generation sequencing technology to sequence reduced representation libraries (Van Tassell et al., 2008) prepared from pools of different individuals (Amaral et al., 2009; Ramos et al., 2009). This resulted in over 390,000 SNPs, including estimates of the minor allele frequencies for these SNPs. Currently, 541,144 of the SNPs discovered in the pig have been submitted to dbSNP, representing around half a million unique SNPs. This resource of SNPs (Fig. 8.2) was used for the design of a porcine 60K Illumina Beadchip (Ramos et al., 2009), which is being used extensively worldwide within industry and academia for genomic selection, whole genome association studies, as well as within a large international porcine HapMap study. Recently, sequencing the complete genome of an individual boar to 26× using Illumina GAII next-generation sequencing resulted in an additional 2–3 million SNPs (Zhan et al., 2009). Furthermore, further sequencing of other breeds, as well as of individual pigs from different breeds, is expected to result in a resource of over 10 million porcine SNPs in the very near future.

Although, SNPs have attracted most of the attention in respect of the genetic variation underlying phenotypic variation, over the past 4–5 years it has become apparent that vertebrate genomes exhibit another type of variation. In addition to changes of a single base pair (SNPs), it is now clear that vertebrate genomes harbour a large number of structural variants (SV), including inversions, translocations, deletions and insertions. To date, such SVs have been most extensively studied in the human genome (for a recent review see Frazer et al., 2009) and the focus is often in particular on insertions and deletions (generally referred to as copy number variation, or CNV). In humans, it is estimated that up to 30% of the genome is affected by this type of variation. Currently, more than 38,000 SVs have been identified in the human genome (http://projects.tcag.ca/variation). An initial CNV analysis of porcine chromosomes 4, 7, 14 and 17 (Fadista et al., 2008) using array
comparative genome hybridization with a probe spacing of 409 bp identified 37 CNVs with a size range of 2–62 kbp. Although these results clearly show the abundance of SVs in the porcine genome, a true comparison of the frequency of SVs in the pig with that in man has to await more systematic studies addressing the complete porcine genome.

The Future of Porcine Genomics

Our knowledge of the structure, function and variability of complex genomes, and the tools to further analyse genomes seem to be changing at an ever increasing pace. While we are writing this chapter, pig genome assembly build 10 (Sscrofa10), representing around 98% of the porcine genome, is already around the corner, which will further improve our knowledge of this important and fascinating mammal. Furthermore, it is likely that through the recent developments of next-generation sequencing technologies, within the next couple of years we will obtain the sequence of tens to hundreds of individual pigs from different breeds, thus providing further detailed insight into the genetics of this species. Many farm animals have seen their genomes sequenced before that of the pig (Hillier et al., 2004; The Bovine Genome Sequencing and Analysis Consortium et al., 2009; Wade et al., 2009). What makes the porcine genome sequencing project distinctive from recent genome projects is the fact that sequencing has been based on a directed approach of sequencing BACs ordered in a highly contiguous physical map. The resulting high-quality genome sequence allows for a detailed analysis of segmental duplications, rearrangements and SVs to an extent not feasible in many genomes whose sequences are based on whole-genome shotgun sequence approaches. With this resource at hand in 2010, the further comparison of the genomes of additional pigs from a variety of breeds will provide a treasure trove not only to understand the genetic basis of important quantitative traits, but also to study further aspects related to speciation, domestication and selection.
Note Added in Proof

In August 2010 the Swine Genome Sequencing Consortium announced the completion of a draft reference genome sequence for the pig (Sscrofa 10) and has published its plans for the analysis and publication of the sequence (Archibald et al., 2010).

References


of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next
generation sequencing technology. PloS ONE 8, e6524.

structure, chromosomal localization and expression profile of a porcine long non-coding RNA isolated from

pig major histocompatibility region containing the classical class I genes. Immunogenetics 53, 490–500.

Renard, C., Hart, E., Sehra, H., Beasley, H., Coggill, P., Howe, K., Harrow, J., Gilbert, J., Sims, S., Rogers,
sequence and analysis of the swine major histocompatibility complex. Genomics 88, 96–110.


conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics 26,
372–378.

type I loci by somatic cell hybrid analysis detects new regions of conserved synteny between human
and pig. Mammalian Genome 7, 275–279.

geno me. Mammalian Genome 13, 578–587.

radiation hybrid map of the RN region in pigs demonstrates conserved gene order compared with the
human and mouse genomes. Mammalian Genome 10, 565–568.

Rogatcheva, M.B., Chen, K., Larkin, D.M., Meyers, S.N., Marron, B.M., He, W., Schook, L.B. and Beever,
J.E. (2008) Piggy-BACing the human genome I: constructing a porcine BAC physical map through
c omparative genomics. Animal Biotechnology 19, 28–42.

BAC library: application to the characterization and mapping of porcine type C endoviral elements.
Cytogenetics and Cell Genetics 85, 205–211.


map of the porcine genome. Genome Research 6, 371–391.

Rothschild, M.F. (2004) Porcine genomics delivers new tools and results: this little piggy did more than just
go to market. Genetics Research 83, 1–6.

Biological Science 3, 192–197.

quantification by bivariate flow karyotyping and karyotype interpretation. Cytometry 13, 703–710.

Schook, L.B., Beever, J.E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., Milan, D., Rohrer, G. and
sequencing the pig genome. Comparative and Functional Genomics 6, 251–255.

tures in porcine EST data and related mammals. BMC Genomics 10, 316.

Nucleotide sequencing analysis of the swine 433-kb genomic segment located between the non-
classical and classical SLA class I gene clusters. Immunogenetics 55, 695–705.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawai, H., Kodzius, R., Watahiki, A., Nakamura,
M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P. and
starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of the USA* 100, 15776–15781.


Introduction

McGlone et al. (1998) wrote the first version of ‘Behaviour Genetics of the Domestic Pig’, which covered feeding, sexual, social, stress and abnormal behaviours. The aim of this chapter will be to update those areas and introduce new areas of behavioural genetics and the link to swine well-being, behavioural traits that may be useful for inclusion in a breeding programme and a detailed review of the genetics of maternal and aggressive behaviours.

The wild pigs of Asia and Europe were selected for domestication in large part because of their behaviour. While there was considerable variation in behaviour within the species, the behavioural traits that favoured domestication included their omnivorous dietary needs, their medium body size, their typically (though not always) docile nature, their relatively weak maternal–neonatal bonds, their precocial nature and their general adaptability (Hale, 1969; Ratner and Boice, 1976). The domestic pig was selected from native strains of wild pigs in Europe and Asia. The wild ancestors, the oldest breeds of domesticated European pigs and common modern-day domesticated pigs share many behavioural characteristics (Mitichashvili et al., 1991). Extensive research has been carried out to study the natural behaviour of domesticated pigs and, in many cases, this has shown that fundamental aspects of behaviour differ only marginally from that of the wild ancestors with regard to their social, ingestive and exploratory behaviours (Jensen et al., 2008). Different though from its wild ancestors, the domestic pig has been selected to be more calm, quiet and less active than wild pigs (Robert et al., 1987). The very plasticity of the domestic pig allows it to grow and reproduce in less-than-ideal, and often variable, environments. To survive and prosper requires a plastic and forgiving genome and, therefore, combining this plasticity with the modern...
production environment will have ramifications on pig adaptability, which can then affect production, health and overall well-being.

Behaviour is an integral part of biological regulation, and an important factor in swine production (Mormède, 2005). In turn, the study of behaviour genetics explores the very nature of a species. We know that species are distinguished from each other both by their morphology and by their behaviour. Thus, while sheep are flocking animals, pigs tend to travel in small herds. Ruminants forage, while pigs scavenge as they forage. Intense selection may not make the scavenger a grazer, but it might increase its rate of feed intake and thus its rate of scavenging or feeding. Within a species, and within the behavioural types of that species, selected breeds or strains of animals often have unique behavioural traits that are heritable, or more commonly referred to as breeding true (like the sheepdog versus the lapdog). Pig behaviours vary among genetic lines within the species, some are lethargic, some hyperactive and many are in the middle of this range. The objectives of this review are to detail what is known today on pig behavioural genetics and to provide direction for further investigations.

**Behavioural Genetics and the Link to Swine Well-being**

The movement of pigs indoors has addressed some well-being concerns (Barnett et al., 2001), while creating others (Newman, 1994). Under these intensive conditions, behavioural characteristics may be: (i) manifestations of the pigs’ biological and psychological reactions to change or uniformity of environment; (ii) instinctual behaviours that are sources of difficulty under certain restrictive systems; or (iii) behaviours that are no longer necessary in the new environment. From Newman’s (1994) perspective, the genetics of behaviour may, in turn, provide clues to the well-being of the pig. To date, three main strategies have been used for the genetic improvement of swine: selection among breeds or strains, selection within breeds or strains, and crossbreeding. These breeding strategies have predominantly focused on production traits/outcomes or economically important traits (Falconer, 1990), but production (pigs per sow per year, body weight and pre-weaning mortality), physiology (blood profiles/immune status), anatomy (gait, bone density) and health are all interwoven, albeit directly or indirectly, with the behavioural repertoire of the pig. Moreover, there is more than just economic importance; there is concern about the undesired side effects of selection for high production efficiency and its association with swine well-being. Pig breeding organizations may need to balance these economically important traits with non-economic values such as emotional and societal issues (Kanis et al., 2005).

A new tool that could be used for many aspects of well-being is determining the molecular architecture of behaviour via quantitative trait locus (QTL) analyses. Knowledge of the molecular mechanism of behaviour might help to improve our understanding of behavioural problems such as aggression, environmental adaptation and overall swine well-being. Although there are limited QTL behavioural data for pigs, the Pig QTL database (PigQTLdb) and its peripheral tools make it possible to compare, confirm and locate on pig chromosomes the most feasible location for the genes responsible for quantitative traits important to pig behaviour (PigQTLdb, 2010). A recent study by Reiner et al. (2009) highlights the importance of this tool and its possible future applications. These authors mapped QTLs for a variety of behavioural indices in swine under healthy conditions and after infection with *Sarcocystis miescheriana*. The study was conducted using a Piétrain × Meishan F₂ family and six QTL-controlled behavioural indicators under healthy conditions, and another six, eight and nine QTL-influenced behavioural indicators during acute, subclinical and chronic disease, respectively. For example, time spent rooting was influenced by only two QTLs on pig (*Sus scrofa*, SCC) chromosomes 7 and 9 (SSC7 and 9), and solely during reconvalescence from acute disease (on day 28); Piétrain alleles were associated with higher activities. Two QTLs on SSC7 and 8 showed effects on walking activity, but exclusively before infection. Time spent performing social behaviour was influenced by one, two and two QTLs, before infection, during reconvalescence and...
during chronic disease, respectively. QTLs were located on SSC6, 7, 11, 12 and 13 and explained up to 20.4% of F2-phenotypic variance. Therefore, the authors concluded that behavioural indicators of pigs under healthy conditions and during subclinical, acute and chronic disease are influenced by multiple genes, and they suggested that future searches for QTL behavioural indices will be based on the founder swine breeds and on an environment that allows the expression of different behavioural patterns.

**Behavioural Traits that may be Useful for Inclusion in a Breeding Programme**

Behavioural genetics is a fascinating and important area of research, fascinating because of the complex and rich behavioural repertoire of the pig, and important because of its bearing on swine well-being and productivity (Rydhmer, 2005). Although the field of behavioural genetics has been around for several decades (Siegel, 1976), and the behaviour genetics of farm animals has been reviewed from time to time to synthesize the handful of literature (Siegel, 1976; Hohenboken, 1986, 1987) before the mid-1990s, relatively few investigations have been reported that seek to better understand the genetic basis of pig behaviours. This may be due to numerous behaviour studies requiring an enormous time commitment. However, with increased automation for collecting behavioural data (Hyun et al., 1997; Musial et al., 1999; Meiszberg et al., 2009), more peer-reviewed papers have now been published, mainly addressing feeding, aggressive and maternal behaviours of the pig as they relate to behaviour genetics. Ranking behaviours of the pig that may in turn be added to current breeding value programmes continue to be discussed, although several stumbling blocks prevail. Mormède (2005) elegantly summarized these challenges. First, in pigs only a few genetic lines are available that have been selected for behavioural traits. Secondly, within-breed variation is usually large and this important background noise reduces the influences of individual loci. Thirdly, and perhaps a major limit to behavioural changes in a breeding programme, is the limited basic knowledge about the psychobiological dimensions underlying behavioural trait variability, and the limited availability of reliable and meaningful measures of these that would be as free from environmental influences as possible. Fourthly, Jensen et al. (2008) noted that the age of the pig at which the study is performed may affect the trait to be selected, and that behavioural traits measured in young and adult pigs can have different underlying motivations, which means that they are basically different traits. Further challenges identified are how these behavioural measures can be collected objectively and accurately, and how repeatable they are over time. A final challenge is that the behavioural trait of interest could receive a different importance weight allocated by the breeding company concerned, based on the housing system in which the pig is kept. For example, rooting and pawing behaviours performed by the sow when housed in a farrowing stall may receive less weighting than if the sow is to be housed in a more loose housing system.

**Maintenance Behaviours (Feeding and Drinking)**

The pig, as a species, is known for its relatively copious feed intake. Pigs can reach feed intakes of 5% of their body weight, which exceeds the level of feed intake for most farm animal species (NRC, 1988). All other factors held constant, if we increase feed intake in young animals, we will not only improve the rate of weight gain but do so at a greater feed efficiency. Thus, we clearly have great economic reason to increase feed intake in lean lines of pigs. The genetic basis for control of average daily feed intake (ADFI) has been less well studied than the closely correlated trait, average daily gain (ADG), but, because ADG and ADFI are highly correlated, selection would be successful if one were to select for feed intake alone. A complete review of the genetics of performance traits can be located in Chapter 14.

Quite different problems of feed intake on commercial farms may show genetic variation
and thus are amenable to change. For example, adult sow feed intake on commercial high-quality diets can be too high, which wastes resources and, if left unchecked, causes excessive body fat and body size. The solution is to physically limit feed intake, which requires resources and alters sow behaviours. Another feed intake problem is the post-weaning feeding period in young pigs. Pigs do not usually consume much dry feed in the hours shortly after weaning. Gradually, over the first few days and weeks, feed intake increases to the species-typical level. A quicker start on dry feed would mean faster and probably more efficient weight gains. The problem of poor post-weaning gains was addressed by Akins (1989). She measured feeding behaviour in the post-weaning period for pigs weaned at 28 days of age for half and full siblings; she measured feeding durations, number of feeding bouts and time from weaning to first feeding experiences. In addition, she calculated a slope of onset of feed intake with time after weaning as X and feeding time as Y. In this model, pigs with a larger slope of feeding duration over time after weaning would have a quicker onset of feed intake. All measures of post-weaning feeding behaviours had very high heritability ($h^2 = 0.87$) estimates that were significantly greater than zero (Table 9.1). In the same work, the $h^2$ for post-weaning ADG was 0.84 ± 0.14. The comparison with post-weaning ADG suggested that the genetic variation for post-weaning feeding behaviour was similar to the genetic variation for ADG, although these $h^2$ values seem high compared with other research. Onset of feeding after weaning also showed significant $h^2$ for several measures. With standard errors of about 0.13, the estimates of $h^2$ ranged from 0.31 to 0.89 for the onset of dry feed intake. These data suggest that selection for early onset of intake of dry feed would have as great a chance of success as selection for body weight gain.

de Haer (1992) examined the genetics of feeding behaviour in group-housed growing pigs. Landrace males and females were compared with Yorkshires of both sexes for feed intake and feeding patterns. Landrace pigs spent more time at each meal; they ate fewer meals with larger intake per meal but had similar overall feed intakes to the Yorkshire pigs. Von Felde et al. (1996) examined voluntary feed intake, feed intake pattern and performance traits, which were recorded on 3188 group-housed boars of Landrace and Large White tested from day 100 to day 170. Measurements of feed intake and feed intake behaviour were obtained by electronic feed dispensers (ACEMO) under ad libitum conditions. Heritabilities of feed intake were estimated to range from 0.16 to 0.30 over the trial. Hall et al. (1999) compiled 1832 individual feeding records from 70 sire families using a food intake recording equipment system. The authors noted that daily feed intake had an $h^2$ of 0.21 (range 0.18–0.26), but that feeding pattern traits were low (0.06 to 0.11), with the exception of feed intake per visit (0.27) and number of visits per day (0.34).

When considering QTLs for feed intake analysis, feeding behaviour and growth traits,

### Table 9.1. Heritability estimates (on diagonal) and genetic (above diagonal) and phenotypic correlations (below diagonal) for selected behavioural traits in weanling pigs. Data were collected over 24 h after weaning and are from full-sibling analyses using Harvey's (1987) procedures (Akins, 1989).

<table>
<thead>
<tr>
<th>Type of behaviour</th>
<th>Feeding</th>
<th>Drinking</th>
<th>Attack</th>
<th>Submit</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeding</td>
<td>0.87 ± 0.13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>Drinking</td>
<td>0.34*</td>
<td>0.58 ± 0.13</td>
<td>0</td>
<td>0</td>
<td>0.14</td>
</tr>
<tr>
<td>Attack</td>
<td>−0.05</td>
<td>0.13</td>
<td>0 ± 0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Submit</td>
<td>−0.09</td>
<td>−0.09</td>
<td>0.03</td>
<td>0 ± 0</td>
<td>0</td>
</tr>
<tr>
<td>FB</td>
<td>0.25**</td>
<td>0.34*</td>
<td>−0.02</td>
<td>0</td>
<td>0.36 ± 0.13</td>
</tr>
</tbody>
</table>

*Differs from zero, $P < 0.01$.
**Differs from zero, $P < 0.05$.

FB, number of feeding bouts during the first 3 days after weaning at 28 days of age.
Houston et al. (2005) examined an F₂ population derived from a cross between Chinese Meishan and European Large White pigs. The only QTL that reached genome-wide significance was located on SSC2 at 28 cM, and this affected daily feed intake during the 55 to 80 kg growth period. The authors reported some suggestive evidence for QTLs that affected average daily gain (chromosomes 8, 11 and 17), daily feed intake (chromosomes 11, 13 and 17), feed conversion ratio (chromosomes 11, 12 and 14), daily feeding time (chromosome 6), average feed per visit (chromosomes 11 and 15), and average feeding rate (chromosomes 3 and 14), but these would need to be followed through in more detail.

Recently, Reiner et al. (2009) also reported that feeding behaviour in pigs was associated with eight QTLs. Baseline feeding behaviour was influenced by QTLs on SSC5, 18 and 7, and together they explained more than 40% of F₂ phenotypic variance. During a disease challenge, these effects vanished in favour of QTLs on SSC6, 8, 9, 16 and X. With one exception on SSC6, feeding behaviour was always positively influenced by Meishan alleles. The genetics of feeding behaviour should not be considered in isolation from the genetics for drinking behaviour, as pigs are prandial drinkers. Reiner et al. (2009) has begun addressing the QTLs associated with drinking in pigs, and reported that it is influenced by six QTLs on SSC5, 7, 11, 12, 15 and 16. None of these QTLs influenced the drinking behaviour of healthy pigs before infection. Two, three and one QTLs, respectively, showed effects during acute, subclinical and chronic sarcocystosis. With one exception (the QTL on SSC12), Piétrain alleles were associated with reduced drinking behaviour. During the subclinical stage of infection (day 28), three QTLs explained about 65% of F₂ phenotypic variance of this behavioural indicator. The QTL on SSC7 alone explained 27% of variance.

Zhang et al. (2009) has also identified QTLs for feed consumption and feeding behavioural traits in pigs. ADFI, feed conversion ratio (FCR), number of visits to the feeder per day (NVD) and average feeding rate (AFR) were recorded in 577 F₂ animals from a White Duroc × Chinese Erhualian resource population during the fattening period of 120 to 240 days. A whole-genome scan was performed with 183 microsatellites covering the pig genome across the entire resource population. A total of eight QTLs were identified on five pig chromosomes, including three genome-wide significant QTLs for FCR on SSC2, 7 and 9, one significant QTL for ADFI on SSC3, and one for NVD on SSC7. These QTLs were identified for the first time, except for the QTL for FCR on SSC2. Four of the five significant QTLs were adjacent to the known QTLs for growth, carcass and fat deposition traits, supporting the existence of gene(s) with pleiotropic effects on these traits. White Duroc alleles were generally associated with greater phenotypic values, except for those on SSC7 and 9. Comparison of QTLs for feed consumption and feeding behaviours indicated that distinct chromosomes had effects on the two types of traits. The authors noted that characterization of the causative gene(s) underlying the identified QTL would shed new light on the genetic basis of feed consumption and feeding behaviours in pigs.

**Sexual Behaviours**

Sexual behaviours include courtship behaviours by both sexes, male and female reproductive techniques and female expression of oestrus. Each of these categories of reproductive behaviour is possibly under genetic control in part or whole. Goy and Jakway (1969) espoused the merits of gaining a better understanding of the genetic basis of sexual behaviours, but little direct work has been reported in pigs. Even reviews (for example, Bichard and David, 1985) of the genetics of prolificacy have not appreciated the role of behaviour in its significant contribution to prolificacy. An excellent review by Flowers (2008) discusses the current understanding of phenotypic variation in reproductive traits of AI (artificial insemination) boars. The proportion of boars that cannot be trained for collection in commercial studs is low, and differences among genetic lines are small. In contrast, there is a considerable variation in sperm production and significant differences are present among genotypes. In general, sexual behaviours in boars have not been studied to the same extent as other reproductive traits.
The majority of the published work performed in the late 1970s concluded that crossbred boars exhibited greater libido than their purebred counterparts, and that boars with Duroc in their pedigrees were less interested in mounting and more reluctant to mate than males with Yorkshire breeding. We do know that certain breeds of pigs express oestrus, including oestrus behaviour, at a very young age. The Meishan breed of pigs from China shows much earlier oestrus onset than traditional European breeds of pigs. Reasons for this are not known for the most part. Onset of oestrus after weaning seems to have a genetic component. ten Napel et al. (1995) reported realized heritabilities of 0.17 to 0.36 ± 0.05 for weaning to oestrus interval. This trait might be amenable to selection or the identification of the major genes involved. The genetic mechanisms of sexual techniques such as mounting, mounting efficiency, coitus, lordosis and others, although generally thought to be innate in part, are not well understood. Male libido, at least, is an economically important trait, probably closely tied to gamete production. Male libido is another area that should be investigated for genetic variation. A complete review of the biology and genetics of reproduction can be found in Chapter 10.

Maternal Behaviours

A preferred measurable trait utilized in most breeding programmes is the number of piglets born or born alive per litter. Selection for an increased litter size at birth does not necessarily translate into a direct increase in the number of weaned pigs if the capacity of the sow to gestate and then take care of her litter is a limiting factor (Högberg and Rydhmer, 2000). Knol et al. (2002) noted that essentially nothing is known at present about the biological background of differences in genetic merit for maternal effects and mothering ability. Yet the sow–piglet reciprocal relationship through their performed behaviours is a critical aspect for successful swine production as it contributes to pre-weaning mortality and growth of the piglets (Simm et al., 1996).

The sow is an unusual ungulate, not only because she is polytocous and builds a nest, but also because licking of the fetal membranes of the newborn is rare. The genetic merit of estimated breeding value of mothering ability (EBV_{ma}) can be calculated as the maternal genetic effect of the foster sow on piglet survival until weaning and expressed as a deviation from the population herd, or even litter mean (Uitdehaag et al., 2008). Knol et al. (2002) concluded that the sow’s EBV_{ma} rather than the piglet’s genetic merit for vitality is related to the litter average time interval between piglet expulsion and first colostrum intake. Piglets born from a group of sows with a relatively high EBV_{ma} had a shorter mean interval from birth until first colostrum intake than piglets born from sows with a lower EBV_{ma} (40 min versus 100 min). Furthermore, these authors noted that the interval from birth until first colostrum intake between sows with different EBV_{ma} values was not related to differences in udder and teat morphology.

Several workers have studied the effects of sow rolling (lateral to sternum lying) on piglet mortality (Weary et al., 1996a,b; Marchant et al., 2001). Uitdehaag et al. (2008) hypothesized that, during parturition, sows with a higher EBV_{ma} would spend more time lying laterally, spend less time standing and have fewer postural transitions than sows with a low EBV_{ma}. These authors found that during parturition sows with a high weaning survival spent more time sitting and less time standing than sows with a low weaning survival, indicating that maternal behaviour during parturition has an effect on piglet survival at weaning; however, no effect of EBV_{ma} on maternal behaviours data collected during parturition was observed in this study.

Optimal sow maternal behaviour after parturition is a critical component of piglet survival and sow longevity within the breeding herd. It is likely that natural selection in the wild boar favours a willingness to protect piglets from intruders and predators, as suggested by theories that emphasize the adaptive functions of aggression- and fear-related defensive behaviours (Grandinson et al., 2003). Although the original function for such behaviour is piglet defence, it can be a practical problem when the behaviour is directed either at the stockperson or at the piglets themselves.

The most dramatic form of maternal aggression is cannibalism. This involves a sow biting, killing and eating newborn piglets.
Three behaviour subtypes, varying in degrees of expression, can be observed for cannibalistic sows. One form of cannibalism occurs when the sow is hyper-reactive following piglet birth and responds with agitation to piglet activities, including their vocalizations. Piglets killed in this situation may be wholly or partially eaten. The second form of cannibalism resembles neonatal rejection. The sow persistently avoids her piglets and this avoidance behaviour can lead to aggression directed towards piglets when they approach her. The third form of cannibalism in the sow resembles the general puerperal aggression condition. This resembles the mental illness sometimes occurring after childbirth in humans. The affected sow is hyperactive following parturition and shows aggression towards people or piglets that come within her range, and snaps aggressively at any intruding piglets. The aggressive piglet biting usually leads to the death of the entire litter. As with the previous forms of cannibalism, piglets killed by biting may be partially or wholly eaten, or left. All of these cannibalism forms expressed by sows are associated with hyper-excitability and are limited primarily to sows (gilts) in their first parity, although the behaviour can be observed in older experienced sows as well. Once cannibalism has started, the behaviour is likely to continue until the litter is lost entirely.

Two recent QTL studies have begun addressing maternal infanticide in sows. An affected sib-pair whole-genome linkage analysis was carried out with 80 microsatellite markers covering the 18 porcine autosomes and the X chromosome, with the aim of identifying chromosomal regions responsible for this abnormal behaviour (Quilter et al., 2007). The results identified four QTLs mapping on SSC2, 10 and X (SSCX). Chen et al. (2009) studied maternal behaviours from 5h before to 24h after parturition in 288 White Duroc × Erhualian intercross F₂ sows over three continuous farrowings. In the F₂ population, 12.8% gilts showed maternal infanticide in their first litter, while the incidences of maternal infanticide at their second and third farrowing reduced to 7.5% and 4.5%, respectively. All F₂ sows were genotyped for 194 microsatellite markers spanning the whole pig genome. The authors identified that seven chromosome regions on SSC2, 6, 14, 15 and X were significantly linked with maternal infanticide (P < 0.05). The QTLs on SSC2 and SSCX achieved P < 0.01. Chen et al. (2009) noted that the most promising QTLs, however, were detected on the X chromosome, where there were three peaks of negative logarithm P-values located at markers SW980, SW2456 and SW1608.

Protectiveness by the sow towards her piglets has been measured by how the sow reacts to fear and reaction to piglet screaming. Vangen et al. (2005) estimated heritabilities in a Norwegian herd for several maternal traits that could affect piglet survival. Heritability estimates were 0.00 for carelessness against piglets, 0.12 for reaction to piglets screaming, 0.17 for fear during routine management and 0.22 for protests by the sow when moved to the farrowing pen. Several authors have reported the importance of h² for maternal behaviours over lactation that may influence sow productivity and, in turn, longevity within the breeding herd. Grandinson et al. (2002) estimated the phenotypic and genetic relationships between the sow’s reaction to a piglet screaming and early piglet survival/growth. Twenty per cent of the sows did nothing, 47% reacted by looking for the sound, 12% sat up and 21% stood up. In addition, the females exhibiting the greatest response to the scream test included gilts and parity five and older sows. Gade et al. (2008) collected data from 32 nucleus and multiplier herds in Germany to estimate variance components and breeding values for five maternal behaviour traits in sows. A single-trait analysis was implemented to estimate these values using a threshold animal model. A total of 31,000 farrowings recorded from December 2003 until July 2005 were included. The h² estimates were 0.07 (SE 0.06) for group behaviour, 0.06 (SE 0.03) for attitude to people and 0.05 (SE 0.01) for maternal ability. Additionally, genetic correlations between the behaviour traits themselves, and between the behaviour traits and litter size, respectively, were estimated using restricted maximum likelihood (REML) procedures in which a linear model was implemented in the multivariate analysis. Low heritability and weak genetic correlations with litter size at birth indicate that it may be difficult to genetically improve maternal behaviour, and that selection for better mothering ability is not necessarily
accompanied by changing litter size at birth. In another study, Grandinson et al. (2003) recorded four behaviour traits that may play a role in maternal ability, and calculated $h^2$ estimates for the sow that included: (i) a sow’s response to the sound of a screaming piglet; (ii) a sow’s response towards the stockperson handling her piglets; (iii) avoidance of; and (iv) aggression towards the stockperson during the handling of the piglet (Table 9.2). The results of this work indicate that selection for a strong response in the scream test or selection against sows that avoid humans would result in a correlated genetic improvement in piglet survival.

Avoidance may reflect underlying fear, and selection for lower levels of fear in sows would improve both sow and piglet well-being.

The complex phenomenon of pre-weaning mortality has been well researched, with 30% of the losses attributable to a single factor, for example a piglet being trapped by the sow, and 70% to multiple factors, for example a piglet missing a meal, becoming lethargic and then being trapped by the sow (Le Dividich et al., 1996), but the reasons why sows crush and kill their young are as yet unexplained. Ethological theories relate to the sow farrowing so many young that some are dispensable. Other theories reflect that the environment or nutrient requirements are not met, or that predators are close, which can increase sow excitability (van der Steen et al., 1988). Crushed piglets not only represent one of the most important factors limiting sow productivity, but also present a substantial obstacle in any attempt to improve piglet well-being (Arey, 1993) (Table 9.2). These undesired aspects in the domestic sow’s maternal behaviour may have arisen during the domestication, especially during the intense selection breeding regimes for economically important traits that have been prevalent in the past few decades.

Špínska et al. (2000) conducted a study comparing domestic (Yorkshire sire × Dutch Landrace mothers) and wild (sire) × domestic crossbred sows and asked: (i) ‘did maternal behaviour change during domestication?’; and (ii) ‘can the inter-individual variability of maternal behaviour be subsumed into a few dimensions of maternal temperament?’ In this study, the crossbred wild × domestic sows did not differ from the domestic sows in any aspect of maternal behaviour except for a higher tendency to terminate nursing and a higher frequency of changing posture from lying to standing and back during the night. Factor analysis (based on a correlation matrix of 11 behaviours and cortisol variables calculated for all sows after removing the effect of breed) indicated that 82% of the variability in the data could be explained by three factors: (i) ‘calmness’, of which sow nighttime frequency of major posture changes, carefulness of laying-sow behaviour and a high propensity to remain in the nursing position after nursing is completed loaded positively; (ii) ‘protectiveness’, which had high loadings for reaction scores to the playbacks of piglet distress calls and human presence near the piglets; and (iii) ‘nursing activity’, which was strongly positively associated with nursing frequency. The authors concluded that most pig maternal behaviour aspects have not been significantly changed by domestication, and that substantial

<table>
<thead>
<tr>
<th>Author</th>
<th>Trait</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knap and Merks (1987)</td>
<td>Aggressive behaviour of sow to piglets</td>
<td>0.7–0.9</td>
</tr>
<tr>
<td>van der Steen et al. (1988)</td>
<td>Aggressive behaviour of sow to piglets</td>
<td>0.1–0.3</td>
</tr>
<tr>
<td>Hen (1996)</td>
<td>Aggressive behaviour of sow to piglets</td>
<td>0.17</td>
</tr>
<tr>
<td>Grandinson et al. (2002)</td>
<td>Scream test</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Crushing</td>
<td>0.07</td>
</tr>
<tr>
<td>Grandinson et al. (2003)</td>
<td>Avoidance</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Handling test</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Scream test</td>
<td>0.06</td>
</tr>
<tr>
<td>Gade et al. (2008)</td>
<td>Aggressive behaviour of sow to piglets</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Crushing</td>
<td>0.03</td>
</tr>
<tr>
<td>Hellbrügge et al. (2008)</td>
<td>Crushing</td>
<td>0.03</td>
</tr>
</tbody>
</table>
maternal behaviour variability exists between sows, perhaps in the form of several behaviour characteristics that encompass both behaviour and endocrine profiles of the sows.

Although the sow can be the main cause of pre-weaning mortality, either directly through crushing (Marchant et al., 2000) or indirectly via starvation (Algers, 1994; Weary et al., 1996a), the piglets’ behaviours and interactions with their dam can also contribute to their survival to weaning success. In populations with reliable registration of farrowing survival and pre-weaning survival, genetic variation can be quantified by the breeding value estimates for pig survival (EBVps). Estimated breeding value for pig survival for an individual pig represents its genetic merit to survive from parturition until weaning; values for individual pigs may, among other factors, be related to birth weight (Rohe et al., 2009), farrowing progression (Leenhouders et al., 2001) and the newborn piglet early postnatal behaviour (Fraser et al., 1997). Leenhouders et al. (2001) reported that farrowing survival ($P = 0.007$) and early postnatal survival ($P = 0.027$) both increased with increasing EBVps. Birth intervals tended ($P = 0.10$) to increase with increasing EBVps, and duration of farrowing was not related to EBVps. Time until first teat in mouth increased with increasing EBVps ($P = 0.05$), but the other piglet vitality behavioural indicators (first upright standing, time to first contact with udder and time to first colostrum intake) were not related to EBVps. The authors noted that information on these survival-related factors in piglets with known EBVps values will give insight into the biological background of genetic differences in piglet survival, and will contribute to an improved selection and management strategy.

### Stress-related Alarm, Fear and Other Emotional Behaviours

#### From stress to the psychobiology of adaptation

Responses to environmental challenges involve different types of mechanisms: psychological (emotional and cognitive), biological (nervous and neuroendocrine) and behavioural (fight and flight). Each of these includes a set of specific changes directly related to the stimulus as well as non-specific adaptations, usually referred to as stress responses. A great degree of variability exists among individuals of the same species (Mormède et al., 2002). In general, the concept of stress derives from the study of the physiological adjustments necessary to maintain homeostasis in a fluctuating environment. Walter B. Cannon (1935) noticed that the specificity of the sympathetic nervous system response usually seen with most stimulation was lost when the stimulus intensity overflowed the normal regulatory mechanisms, or in the case of an intense emotional activation, such as the case of a cat exposed to a barking dog. The cardiovascular and metabolic actions of the sympathetic nervous system were considered as necessary adjustments allowing an efficient behavioural adaptation, the ‘fight or flight’ response. Selye (1936) described the activation of the adrenocortical gland, in which glucocorticoid hormones were released under the control of the anterior pituitary gland and of the paraventricular nucleus of the hypothalamus (the hypothalamic–pituitary–adrenal (HPA) axis), as a non-specific response to a number of stimulations. Mason (1971) demonstrated that this non-specificity response was primarily the result of emotional activation by the environmental stimulus. Thus, stress research moved from pure physiology to psychobiology, with physiological responses being intimately interconnected to behavioural adjustments. Both response outputs are the coordinated expression of a central emotional state induced by environmental stimulations. However, these responses are not stereotyped, but depend upon the specific individual reactivity pattern and the behavioural control efficiency over the stimulus (Danzer and Mormède, 1983). For an excellent review of molecular genetic approaches to investigating individual variations in behavioural and neuroendocrine stress see Mormède et al. (2002).

### Genetics of emotional behaviours

Behavioural scientists have devised different tests to determine individual behavioural reactivity features in experimental settings. These include
the open field test, the human approach test, the tonic immobility test, the elevated plus maze test and the emergence test (Anderson et al., 2000). Numerous behavioural measures can be collected on each animal while undergoing these tests; for example, ambulation score, exploration time, vocalizations, faecal boluses and urinations are the classical measures of the emotional output, and involve both behavioural and autonomic responses. The tests can be further complicated; for example, if animals are placed into the open field test in the presence of food in the arena, to create a motivational conflict (neophobia test, Mormède et al., 1984), or by adding new objects during the test (Lawrence et al., 1991; Jensen et al., 1995a,b). Behavioural measures can be complemented by collecting physiological measures; for example, the adrenocortical axis response to such challenges can be measured by comparing plasma levels of adrenocorticotropic hormone (ACTH) and cortisol before and after the test.

Several divergent selection experiments involving laboratory rodents to measure emotionality have taken place, and rodents that display a low motor activity and a high elimination are considered more emotional (Hall, 1934; Archer, 1973). Furthermore, these emotions have a genetic component (Broadhurst, 1962; DeFries et al., 1966). However, little work has been done to determine emotionality and the link to behavioural genetics in pigs. Von Borell and Ladewig (1992) reported that the open-field score variability was larger between litters than within litters. Comparison among different breeds demonstrated considerable variation (Mormède et al., 1984). Mormède et al. (1994) studied F2 intercresses between Large White sires (with a high score of activity in the open-field test) and Meishan dams (with a low score in the test) in a novel environment and found that the character was normally distributed, suggesting that the trait is under multigenic control. Moreover, Large White pigs displayed higher defecation and locomotion scores than Meishan pigs, and F1 crossbred animals were intermediate for locomotion scores, suggesting that there were no dominance effects for this behavioural trait. For defecation scores, Meishan pigs were behaviourally dominant over Large Whites. The $h^2$ estimate for the locomotion score, calculated as four times the paternal half-sib correlations, was approximately 0.16. In another study, Barnett et al. (1988) showed that genetic selection for growth performance resulted in an increase in activity, but no change in exploratory activity, in a novel environment, together with a moderately increased feeding time in this novel environment, and a reduced motivation to interact socially.

Shea-Moore (1998) utilized the open-field test for segregated early-weaned pigs and two levels of body leanness. One group was selected for high levels of lean gain (Hi-L) and the other was selected for low levels of lean gain (Lo-L). Pigs were tested for 5 min, and vocalization, defecation and activity level were recorded. Salivary cortisol samples were collected immediately after the open-field test and also at 15, 30 and 45 min after the behavioural test. The results indicated higher levels of activity in the Lo-L treatment group than in the Hi-L group ($P < 0.05$). However, there was large variation in the number of vocalizations regardless of treatment and, because of this variability, no treatment differences were detected. Higher levels of activity usually suggest less anxiety and more interest in exploration of the environment, yet the level of defecation indicated an increase in anxiety in a novel situation. Although baseline cortisol levels were higher in the Lo-L group than in the Hi-L group ($P < 0.05$), over time there were no treatment differences. The author noted that pigs selected for a lower lean-gain showed a higher activity level in the open-field test, and that it may be possible that, by selecting for high lean gain, an animal’s ability to cope with a novel situation is changed, thus affecting the well-being of the animal. A more recent study by Fàbrega et al. (2004) investigated the effect of the \textit{RYR1} (known as the halothane gene, and associated with stress susceptibility, which can trigger malignant hyperthermia) genotype on the open-field behaviour in growing pigs. The study subjected 15 heterozygous ($Nn$) and 15 \textit{RYR1}-free ($NN$) gilts of 19 weeks old to three replicates of an open-field test 2 days apart from each other. The study measured the number of grid lines crossed and the defecation score in the test arena. There was a significant individual correlation among the three replicates.
of the test, both for the number of grid lines crossed and the defecation score \((P < 0.05)\). The RYR1 genotype had a significant effect on the number of grid lines crossed, with NN gilts showing more overall activity than Nn gilts \((P < 0.05)\). In this work, no significant differences in defecation score occurred between the genotypes, and the authors concluded that the RYR1 genotype may have an effect on the appraisal of novelty.

The social behaviours of pigs towards humans show large individual differences (Lawrence et al., 1991), and might be under genetic control. Hemsworth et al. (1990) measured pig fear in a human approach test. Pigs that approached more quickly were said to have less fear of humans, and the \(h^2\) estimate for ‘fear of humans’ was 0.38 ± 0.19. Tonic immobility (the back test) data can be collected by placing each pig on its back, restraining it in a supine position for a defined period of time, and counting the number of escape attempts and/or vocalizations. Hessing et al. (1993, 1994) described the back-test score distribution as bimodal, suggesting that two distinct phenotypes (active versus passive) exist. However, Forkman et al. (1995) could not confirm a bimodal distribution, and several other groups have reported a unimodal distribution of reactivity traits in the population (Lawrence et al., 1991; Mormède et al., 1994). Forkman et al. (1995) compared piglet behaviour using social (rank order, social dependence, aggression) and non-social (novelty, extinction, back test) tests, and used principal component analysis to study the relationships between these different measures. This analysis suggested three independent personality traits: aggression, sociability and exploration. This approach is quite recent in pigs, but confirms the multidimensional features of behavioural response to social and non-social challenges that have been demonstrated in laboratory rodents and other species, including pigs.

Velie et al. (2009) recently estimated repeatabilities andheritabilities for measures of pig behaviour and their relationship with performance. Measures of behaviour and performance included the back test, resident–intruder test, human approach test (HAT), novel object test (NOT), day 1 body weight (BW), backfat depth (BF), loin muscle area (LMA), ADG in the farrowing house, ADG, 21-day BW and 140-day BW (or W). Each behavioural trait was measured twice. The study consisted of 95 litters from 31 sires with an average of three litters per sire \((n \geq 457)\). Between 7 and 14 days old, the back test was conducted by placing each pig in a supine position for 60 s. Total time spent struggling (TTS) and total number of attempts to struggle (TAS) were recorded. The resident–intruder test involved two nursery pigs, a resident pig and an unfamiliar intruder pig. The resident pen was divided in half by a solid partition. A resident pig was placed in the test area, and an intruder pig was then introduced. Latency until an attack occurred (LAT) and total number of attacks over two tests (RIS) were recorded. The amount of time taken for each finishing pig to make snout contact with an unfamiliar human or object was recorded. Dam and sire effects influenced all traits \((P < 0.01)\). Sex and pen affected LAT, RIS, HAT and NOT \((P < 0.10)\). Repeatabilities of TTS, TAS, RIS, LAT, HAT and NOT were 0.38, 0.21, 0.07, 0.08, 0.17 and 0.11, respectively. The phenotypic correlations of TTS with TAS and HAT with NOT were 0.61 and 0.34, respectively. Phenotypic correlation between RIS and LAT was −0.85. TTS and TAS tended to be phenotypically correlated with 21-day BW and ADG in the farrowing house. TAS was phenotypically correlated with BF (0.15). LAT was phenotypically correlated with LMA (0.23). Resident intruder score was phenotypically correlated with ADG (−0.13), W (−0.13) and LMA (−0.21) and estimated to be slightly heritable \((h^2 = 0.12)\). The heritabilities of TTS and TAS were 0.31 and 0.53, respectively. The genetic correlation of TAS with ADG and W was 0.38. Genetic correlations of TTS with BF, W and TAS were 0.14, 0.18 and 0.81, respectively.

**Genetics and neuroendocrine emotional responses**

In pigs, most available data deal with the HPA axis, although large breed/strain variation has been reported for circulating catecholamines in pigs (Mormède et al., 1984), as well as in laboratory rodents (McCarty and Kopin, 1978).
Individual differences in circulating cortisol levels were shown by Hennessy et al. (1988) to be related to changes in the adrenal response to ACTH (Hennessy et al., 1988; Zhang et al., 1990, 1992), a result reported in humans as well (Bertagna et al., 1994). Indeed, the cortisol response to ACTH shows a very large range of inter-individual variation, but is a stable trait in individuals (in pigs, Von Borell and Ladewig, 1992; in humans, Bertagna et al., 1994). The cortisol response to ACTH can be influenced by chronic environmental stress, such as tight restraint (Janssens et al., 1994, 1995), and by genetic factors, as clearly demonstrated by several divergent selection experiments in poultry (Brown and Nestor, 1973; Edens and Siegel, 1975; Satterlee and Johnson, 1988). Large differences in circulating cortisol levels between European and Chinese pigs have been described (Mormède et al., 1984), and evidence exists for genetic control of adrenocortical activity from a study involving an F2 population from a cross between Large White and Meishan pigs (Mormède et al., 1994). Désautés et al. (1999) reported that the difference originated from the adrenal gland, and that it was independent of the ACTH drive. Taken together, these experiments show that multiple mechanisms may explain the genetic control of adrenocortical function.

Several experiments have demonstrated a link between HPA axis activity and production traits in domestic livestock species. Hennessy and Jackson (1987) and Barnett et al. (1988) showed that selection for growth performance was followed by reduced plasma cortisol levels and maximum corticosteroid binding capacity in pigs. Behavioural and neuroendocrine response modulation of stress in pigs as a result of genetic factors is now well established. Tools to investigate the molecular basis of genetic influences on the different stress responses are available, as demonstrated in mice by Flint et al. (1995). This approach of locating behavioural loci has been possible in the pig since the advent of pig genome maps (Archibald, 1994, 1995; Rohrer et al., 1994) and the more recent announcement of the first draft sequence of the pig genome.

Recent investigations have begun concentrating on the genetic relationships between different emotional reactivity traits and economically important production traits. Désautés et al. (2002) conducted a QTL analysis of behavioural and neuroendocrine responses to a novel environmental stress in a three-generation experimental cross between Meishan and Large White pig breeds. Locomotion, vocalization and defecation rate, as well as exploration time, were measured for 10 min. Blood samples were taken immediately before and after the test to measure plasma levels of ACTH, cortisol and glucose. Animals were typed for a total of 137 markers covering the entire porcine genome. The authors reported a highly significant gene effect for post-stress cortisol level \( (P < 0.001) \) and a significant effect for basal cortisol level \( (P < 0.05) \) at the end of the q arm of chromosome 7, which they noted as explaining 20% and 7% of the phenotypic variance. Meishan alleles were associated with higher cortisol levels and were partially dominant (for post-stress levels) over Large White alleles. Other significant gene effects on biological measures were detected on chromosomes 1 and 17 (ACTH response to stress) and 3, 5 and 8 (glucose levels). An interesting observation by Hutson et al. (2000), when comparing the responses of individual growing pigs to 60 stimuli from five sensory categories, was that significant variation occurred between individual pigs and that future studies of emotion, stress and reactivity should use stimuli that elicit high responsiveness but have little variation.

**Aggressive Behaviours**

Aggressiveness in pigs shows wide phenotypic variation and if it could be reduced could enhance health, well-being and growth performance (Bergsma et al., 2008). Excellent work has been conducted using the laying hen as a model to reduce aggressiveness (feather pecking) through selection programmes based on their behavioural repertoire (Muir, 1996). A recent study by Labouriau et al. (2009) used a high feather peck (FP) selection line, which has been selected for eight generations, and the authors presented data showing that the gene transcription profile of the birds performing high FP differs from the
Turner et al. (2006) determined the genetic contribution to aggressiveness in pigs by determining the relationship with skin lesions. A sample of 1132 pigs was mixed at an average weight of 27.9 kg into 96 pens on a commercial sire line nucleus unit. Post-mixing aggressiveness of pigs was assessed and an $h^2$ of 0.22 was estimated for the lesion score (LS) trait. The response to selection, when all selection pressure was placed on the LS trait, was a 25% reduction in LS per generation. Further work by Turner et al. (2008) used a Bayesian approach to estimate the heritability of three traits associated with aggressiveness in pigs during the 24 h post mixing: duration of reciprocal aggression, and whether in receipt of or delivery of non-reciprocal aggression (NRA). In this study, the genetic correlations were quantified between the behavioural traits and skin lesions (Table 9.3). The authors concluded that, based on the estimated genetic parameters, the selection of breeding values for reduced LS (especially LS for the central region of the body) is expected to reduce reciprocal aggression and the delivery of NRA, but will not change the receipt of NRA directly. In a follow-up study, Turner et al. (2009) estimated the genetic correlations between skin lesions and aggressive behaviour post mixing and under more stable social conditions as a potential means of selecting against pig aggressiveness. Aggressive behaviour was recorded continuously for 24 h after mixing, and a count of skin lesions (lesion count, LC) was recorded at 24 h and 3 weeks post mixing on 1663 pigs. Two behavioural traits were found to have a moderate to high $h^2$ similar to that of growth traits, whereas receipt of NRA had a lower $h^2$ (Table 9.3). Genetic correlations ($r_g$) suggested that lesions to the anterior region of the body 24 h after mixing were associated with reciprocal fighting ($r_g = 0.67 \pm 0.04$), receipt of NRA ($r_g = 0.70 \pm 0.11$) and, to a lesser extent, delivery of NRA ($r_g = 0.31 \pm 0.06$); lesions to the central and rear regions were primarily genetically associated with receipt of NRA ($r_g = 0.80 \pm 0.05, 0.79 \pm 0.05$, respectively). Genetic correlations indicated that pigs that engaged in reciprocal fighting delivered NRA to other animals but were less likely to receive NRA themselves (Table 9.3). Turner et al. (2009) noted that a genetic merit index using lesions to the anterior region as one trait and those to the centre or rear, or both, as a second trait should allow selection against animals involved in reciprocal fighting and the delivery of NRA. Positive correlations between LC 24 h and 3 weeks after mixing were found, especially for lesions to the central and rear regions of the body, indicating

<table>
<thead>
<tr>
<th>Behavioural traits</th>
<th>Reciprocal aggression</th>
<th>Delivery of NRA</th>
<th>Receipt of NRA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reciprocal aggression</strong></td>
<td>0.46</td>
<td>0.79</td>
<td>−0.16</td>
</tr>
<tr>
<td>2008 study</td>
<td>0.43</td>
<td>0.84</td>
<td>−0.04</td>
</tr>
<tr>
<td>2009 study</td>
<td>0.15</td>
<td>0.37</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Delivery of NRA</strong></td>
<td>0.39</td>
<td>0.31</td>
<td>−0.41</td>
</tr>
<tr>
<td>2008 study</td>
<td>0.39</td>
<td>0.31</td>
<td>−0.41</td>
</tr>
<tr>
<td>2009 study</td>
<td>0.62</td>
<td>−0.23</td>
<td>0.17</td>
</tr>
<tr>
<td>2009 study</td>
<td>0.31</td>
<td>0.23</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Reciprocal aggression defined as reciprocal damaging fighting lasting ≥1 s.

*Non-reciprocal aggression (NRA) defined as rest during a fight, withdrawal at the end of a fight and attack not associated with a fight.
that post-mixing lesions are predictive of those received under more stable group conditions. Therefore, these selection indexes could not only reduce immediate aggression at mixing, but also translate into a long-term impact on injuries from aggression, even after dominance relationships are established.

**Summary**

Swine domestication has altered aspects of the pig’s behaviour through genetic selection and the control of its environment. Over the past 5 years, work identifying QTLs for pig behaviour has become available, in particular for maintenance-related behaviours. We encourage more information on the identification of QTLs that might influence maternal, social, emotional and aggressive behaviours. In particular, the importance of QTL identification pertaining to social and aggressive behaviours will become paramount if legislation in the USA continues to ban gestation stalls and sows are placed into a group-housed gestation system.

**References**


Introduction

Reproduction is an essential process for the maintenance of a species. It has to be genetically controlled to ensure that the reproductive process is repeated with a strong degree of precision. Yet a considerable genetic variability exists between breeds for both male and female reproductive traits. For instance, average litter size of mature sows varies from four to 16 piglets per litter among breeds. These differences, combined with the appreciable genetic variation that also exists within breeds, have given the opportunity for substantial genetic improvement of sow prolificacy over the last 15 years (up to 30% of the mean in some breeds – see, e.g. Tribout et al., 2003). The main consequence of increased litter size has undoubtedly been large gains in the efficiency of pig production systems, but it has also resulted in some adverse effects, such as weaker oestrus symptoms (Rydhmer, 2000) or lower piglet survival (Tribout et al., 2003; Canario et al., 2007a). Such unfavourable correlative trends not only reduce the positive short-term effects of selection on efficiency, but may also have long-term consequences, such as decreased fertility owing to an increased proportion of undetectable oestrus or, for piglet mortality, may raise ethical problems. This emphasizes the importance of a broad perspective of the consequences of genetic improvement, which generates new questions, such as the societal or ethical consequences of selection, but it also strengthens the need for a good knowledge of the genetic (co)variation of a large number of potentially important traits. After a quick overview of pig reproductive biology, current knowledge on the genetic control of reproductive traits in pigs and their relationships with other
potentially important traits is reviewed. Both classical parameters characterizing among-breed (breed differences and heterosis effects) and within-breed (heritabilities and genetic correlations) genetic variability and new information on individual genes or genomic regions (quantitative trait loci, or QTLs) influencing reproductive traits are considered.

**Some Aspects of Pig Reproductive Biology**

**Female**

Puberty in gilts, usually defined as the moment of first ovulation, occurs at 3–4 months of age in the earliest maturing breeds (Chinese) and at an average of 6–7 months of age in the most widely used Western breeds. It generally coincides with the first oestrus, though ovulation without external manifestation of oestrus (silent heat) occurs occasionally in pigs, and generates a steroid-secreting activity of the corpora lutea. Age at puberty is quite tedious to measure and is often replaced in field studies by age at first mating or at first farrowing. Although strongly correlated, the two traits differ owing to the large variability in management strategies between breeders (Le Cozler et al., 1998; Rydhmer, 2000). Ovulations then occur every 3 weeks during the second half of a 2–3 day oestrous period in the absence of gestation, and have a mean duration of 2–5 h (Soede et al., 1992). The oestrous cycle is controlled by gonadotrophic hormones. Follicle stimulating hormone (FSH) stimulates recruitment and development of ovarian follicles. Ovulation and corpora lutea formation are stimulated by luteinizing hormone (LH). Ovulation rate increases with oestrus and parity number until the fourth or fifth parity. Conception rate in the pig is high (80–90%) and has increased with the generalization of multiple matings (two or sometimes three services 12 or 24 h apart during oestrus). Fertilization of the ova begins a few hours after mating after a necessary period of spermatozoa capacitation (Hunter, 1982). Fertilization rate is generally close to 100%. The rate of prenatal mortality in pigs is 30–40% on average. The largest part of losses (20–30%) occurs during the first month of gestation, with an additional 5–10% of fetal loss during late gestation (Wrathall, 1971; Christenson et al., 1987).

Farrowing lasts 3–5 h on average, with large variations between sows. It is initiated by the piglets, which produce corticosteroids, resulting in a production of prostaglandin by the placenta, which causes the regression of the corpora lutea and the initiation of the farrowing process. Parturition and the first few days after birth are critical periods for piglet survival. Up to 10% of piglets are stillborn in some populations (Holm et al., 2004a; Canario et al., 2007a), predominantly as a result of hypoxia during delivery. After birth, the time to get the udder, suckle and ingest an appropriate amount of colostrum and milk is a major determinant of piglet survival and growth. The sow has a strong control of milk ejection, with duration of milk flow of 10–20 s, and an average nursing interval of less than an hour. Colostrum and milk production are difficult to measure directly, but can be indirectly estimated from piglet weight gain (Etienne et al., 2000; Devillers et al., 2004). Colostrum is essentially produced during the first 24 h after parturition, and amounts to 4 kg on average. Milk production peaks at around 21 days of lactation, and may reach 1 kg daily for each piglet in sows nursing 10–12 piglets. With very few exceptions, the lactating sow has a very limited follicular development, and does not ovulate or show any oestrus symptoms. The total removal of the sow from her litter at weaning normally results in an acceleration of follicular growth and in ovulation within 4–10 days.

**Male**

Spermatogenesis in the boar starts at 4–6 months of age in most pig breeds, but may begin before 100 days of age in some early maturing breeds, such as the Chinese Meishan. Sperm quality and quantity then steadily increase with testicular development, testosterone production and libido until sexual maturity at 6–8 months of age, and then at a much lower rate until boars reach their adult body size. A parallel rise in male accessory glands (the seminal vesicle, prostate and bulbo-urethral
glands), which produce 95% of the seminal plasma, results in a correlated increase in the volume of the ejaculate. Sexual activity is controlled by gonadotrophic hormones. FSH stimulates spermatogenesis, whereas LH stimulates steroid hormone (testosterone, but also other steroids such as androstenone) synthesis and secretion by the interstitial Leydig cells. The action of LH is dependent on the FSH induction of LH receptors on the Leydig cells. Boar ejaculate is characterized by its large volume (around 300 ml on average) and spermatozoa number (80 to 120 billion when semen is collected once a week), which corresponds to total sperm reserves and widely exceeds daily sperm production (10 to 20 billion spermatozoa a day). As a consequence, spermatozoa number per ejaculate steadily decreases when the boar is used or ejaculate is collected more than once a week, in spite of a slight increase in sperm production with ejaculation frequency. Large amounts of spermatozoa and semen are necessary to ensure normal conception rate and prolificacy (50 ml of semen and 3 billion sperm are usually considered as minimum requirements for artificial insemination). Frozen boar semen can successfully be employed, with comparable or slightly lower conception rate and litter size than fresh semen, but it remains rather expensive and requires good technical expertise, so its commercial use remains quite limited.

### Between-breed Variation and Crossbreeding

#### Breed differences

Unlike production traits, breed comparisons for reproductive traits are not numerous owing to the size and costs of facilities that would be necessary to get accurate estimates of breed or line differences in a single environment. However, breed reproductive performance has until the mid-1990s remained almost unchanged, so that reasonably accurate estimates of breed differences could be obtained by compiling experimental results over time. The situation has radically changed over the last 15 years with the strong emphasis put on selection for litter size in pig dam lines and the large direct and correlative responses to selection that have been obtained. Old breed comparison results are no longer valid for traits related to litter size. For instance, the superiority of the Chinese Meishan breed over the Large White for litter size, which amounted to three to five piglets in the early 1990s (Bidanel, 1993; Haley et al., 1995), was reduced to 1.1 piglets in 2005 (Canario et al., 2006b). In such cases, large-scale on-farm data recording systems are often the only way to get up-to-date information on breed performance. Though differences are purely phenotypic and may also reflect differences in average management or environmental conditions, the data from Table 10.1 show the large differences that now exist between European breeds, as well as the differences in their ranking, with an advantage of Large White over Landrace in France and the reverse situation in Denmark.

If the advantage of Meishan sows regarding prolificacy is probably lower than it used to be, these sows still give birth to litters that have a lower risk of stillbirth (Canario et al., 2006b), in agreement with earlier studies (Blasco et al., 1995; Haley et al., 1995). Similarly, Meishan females still perform better than Large White females for prenatal survival at a given ovulation rate (Bidanel et al., 1990; Haley et al., 1995) and for pre-weaning survival adjusted for litter size born (Lee and Haley, 1995), as selection has not improved prenatal or pre-weaning survival in the Large White breed (Tribout et al., 2003; Canario et al., 2007a). Meishan pigs have also been shown to reach puberty about 100 days earlier than Western breeds both in females (Després et al., 1992; White et al., 1993) and in males (Prunier et al., 1987; Hochereau-de Reviers et al., 1999), and to have a higher conception rate and a lower return to oestrus interval (Després et al., 1992; White et al., 1993) than their Large White contemporaries.

Many other local breeds have reproductive performances that widely differ, but are generally lower than those of Large White or Landrace (Dobao et al., 1988; Gourdine et al., 2006); they will not be detailed here. Significant, although less important than with local breeds, differences in reproductive efficiency also exist between major commercial breeds. For instance, Landrace gilts have been shown to reach puberty
earlier (Hutchens et al., 1982; Allrich et al., 1985; Bidanel et al., 1996), while having a slightly lower ovulation rate and a higher prenatal survival rate than Large White gilts (Bidanel et al., 1996). Breed differences in age at sexual maturity and in semen production traits have also been suggested in boars (Babol et al., 2004; Smital et al., 2004); these variations could be related to differences in androstenone and skatole levels, the two major components of boar taint (Babol et al., 2004). In addition to a lower prolificacy, Hampshire, Piétrain and Duroc breeds tend to have a lower conception rate (Sonderman and Luebbe, 2008) and maternal abilities, as shown by higher pre-weaning mortality rates compared with Large White or Landrace breeds (Blasco et al., 1995).

Crossbreeding

Crossbreeding has for decades been known as an effective means of enhancing reproductive efficiency, in particular because of ‘hybrid vigour’, or heterosis. For traits such as litter traits, which are controlled by both piglets and sow genotypes, enhanced performance may come from both crossed piglets (i.e. litter, direct or individual heterosis effects) and crossbred dams (i.e. sow or maternal heterosis effects). A large number of crossbreeding experiments carried out in the 1970s and 1980s have provided a rather good picture of heterosis effects in pigs (Table 10.2). Most reproductive traits exhibit significant heterosis, the largest effects being associated with sow genes. Indeed, as compared with purebreds, crossbred sows reach puberty earlier, have a higher conception rate, a slightly larger ovulation rate and better embryo/fetal survival rates, and farrow larger litters and have better nursing abilities, i.e. they raise piglets that have a higher probability of survival and grow faster than piglets nursed by purebred sows. Piglet heterosis leads to slightly larger litter size at birth and, above all, to higher piglet growth and survival during the nursing period.

Heterosis effects vary according to breed combination. For instance, Large White × Landrace crosses generally exhibit somewhat lower heterosis values than other crosses between Western breeds. Yet crosses between Western and Chinese local breeds are the most striking illustration of breed combination differences in heterosis, with effects on sow heterosis being up to five times larger than the values usually encountered between European and/or American breeds (Table 10.2).

Several experiments have compared crossbred with purebred boars (see review of Buchanan, 1987). Crossbred boars reach sexual maturity earlier than purebreds. As a consequence, young crossbred males have larger testes size and a higher sperm production than their purebred counterparts. They also have higher conception rates and tend to have more libido and to be more aggressive than purebred boars. This superiority then tends to diminish as boars mature, but it is not clear whether it is completely eliminated in adult animals. Finally, differences in litter size sired by crossbred versus purebred boars are generally small and not significant (Buchanan, 1987).

It has to be emphasized that the improved reproductive performance of crossbred sows is a major reason for the generalized use of

Table 10.1. Examples of comparative reproductive performance of some pig breeds.a

<table>
<thead>
<tr>
<th>Breed</th>
<th>Dam line</th>
<th>Sire line</th>
<th>Landrace</th>
<th>Duroc</th>
<th>Piétrain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number born</td>
<td>14.2</td>
<td>–</td>
<td>14.6</td>
<td>9.9</td>
<td>–</td>
<td>DSP, 2008</td>
</tr>
<tr>
<td>Total number born</td>
<td>14.8</td>
<td>12.0</td>
<td>13.7</td>
<td>10.0</td>
<td>9.3</td>
<td>IFIP, 2009</td>
</tr>
<tr>
<td>Number born alive</td>
<td>13.6</td>
<td>11.0</td>
<td>12.5</td>
<td>9.3</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>Number weaned</td>
<td>11.4</td>
<td>9.4</td>
<td>10.8</td>
<td>7.8</td>
<td>7.8</td>
<td></td>
</tr>
</tbody>
</table>

aStandard errors of breed means range from 0.03 to 0.1.
two-generation crossbreeding schemes in pig production. Indeed, most individual pig producers practise crossbreeding through the use of specialized paternal and maternal genotypes (Moav, 1966; McLaren et al., 1987). In most cases, the maternal genotype is an F₁ sow, but both more complex schemes using great-grandparental populations to produce three-way parental sows and simple partial rotational crossbreeding schemes are also used at a significant level.

### Within-breed Genetic Variability

#### Components of genetic variation

Estimates of heritability for the major female and male reproductive traits are summarized in Table 10.3. Female reproductive traits have low-to-moderate heritabilities. The most heritable traits are those depending solely on the genotype of the female, i.e. age at puberty, at first mating or farrowing, ovulation rate, weaning to oestrus interval and measures of the intensity of oestrus symptoms. Fertility and prolificacy traits, i.e. conception or farrowing rate, litter size, piglet survival rates and, to a lesser extent, litter weight, which result from complex interactions between sow, boar and embryo or piglet genotypes, have low heritabilities and are more difficult to improve through selection.

Heritability value estimates for male traits are much less numerous than for female traits. As most body composition traits, testes and accessory gland measurements, which are indicators of the sexual development of young boars, have moderate-to-high heritability values and are expected to easily respond to selection. Sperm characteristics have low-to-moderate heritabilities when measured on individual ejaculates.
The much higher heritability values reported in some studies (Smital et al., 2005) are based on mean values per boar (averaged over ejaculates). Levels of steroid hormones have also been found to be moderately-to-highly heritable. Realized heritability values of 0.15 and 0.18 were obtained in lines divergently selected for basal testosterone levels (Robison et al., 1994). More numerous estimates are available for the levels of androstenone, a pheromone that accumulates in fat tissues of males and is a major component of boar taint. Heritability values are high, around 0.55 on average (Robic et al., 2008).

Some authors have quantified the relative importance of maternal, paternal and embryo/piglet effects on genetic variation in litter traits. They all have confirmed that the largest part of genetic variation is due to sow genes, except for fertility, where similarly low paternal and maternal heritabilities were described by Varona and Noguera (2001). The service sire has been shown to have a rather limited effect on litter size, only 1–5% of phenotypic variance (Beauvois et al., 1997; Van der Lende et al., 1999; Serenius et al., 2003; Hamann et al., 2004; Su et al., 2007). However, several studies conclude that taking this limited male influence into account improves genetic evaluation models for litter size (Serenius et al., 2003), and even gives the opportunity to detect boars as potential carriers of chromosomal abnormalities (Tribout et al., 2000).

Embryo/piglet genes also potentially influence their own survival and growth. Most estimates of direct genetic effects on piglet survival at birth or during the nursing period are low (below 0.05) compared with maternal effects (Knol et al., 2002a; Lund et al., 2002; Ibáñez-Esriche et al., 2009), although somewhat higher estimates (around 0.10) have been reported in some studies or populations (Van Arendonk et al., 1996). Similarly, growth during gestation is prominently due to the sow’s genes, with two to five times higher heritability values for maternal (0.17–0.26) than for direct (0.03–0.10) effects on piglet birth weight (Roehe, 1999; Knol et al., 2002a; Solanes

### Table 10.3. Heritability ($h^2$) estimates for female and male reproductive traits in the pig. Adapted from Rothschild and Bidanel (1998).

<table>
<thead>
<tr>
<th>Trait</th>
<th>No. of estimates</th>
<th>Mean $h^2$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female traits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at puberty</td>
<td>16</td>
<td>0.37</td>
<td>0–0.73</td>
</tr>
<tr>
<td>Oestrus symptoms</td>
<td>3</td>
<td>0.21</td>
<td>0.09–0.29</td>
</tr>
<tr>
<td>Ovulation rate</td>
<td>18</td>
<td>0.32</td>
<td>0.10–0.59</td>
</tr>
<tr>
<td>Conception rate</td>
<td>3</td>
<td>0.10</td>
<td>0–0.29</td>
</tr>
<tr>
<td>Prenatal survival rate</td>
<td>12</td>
<td>0.15</td>
<td>0–0.23</td>
</tr>
<tr>
<td>Total number born</td>
<td>103</td>
<td>0.11</td>
<td>0–0.76</td>
</tr>
<tr>
<td>Number born alive</td>
<td>118</td>
<td>0.10</td>
<td>0–0.66</td>
</tr>
<tr>
<td>Number weaned</td>
<td>54</td>
<td>0.08</td>
<td>0–1.0</td>
</tr>
<tr>
<td>Farrowing survival rate</td>
<td>12</td>
<td>0.07</td>
<td>0.03–0.14</td>
</tr>
<tr>
<td>Birth to weaning survival rate</td>
<td>12</td>
<td>0.05</td>
<td>0–0.13</td>
</tr>
<tr>
<td>Farrowing length</td>
<td>2</td>
<td>0.07</td>
<td>0.05–0.10</td>
</tr>
<tr>
<td>Litter homogeneity at birth</td>
<td>6</td>
<td>0.08</td>
<td>0.03–0.1</td>
</tr>
<tr>
<td>Litter weight at birth</td>
<td>18</td>
<td>0.24</td>
<td>0–0.54</td>
</tr>
<tr>
<td>Birth assistance</td>
<td>2</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Litter weight at 21 days</td>
<td>22</td>
<td>0.14</td>
<td>0.07–0.38</td>
</tr>
<tr>
<td>Weaning to oestrus interval</td>
<td>5</td>
<td>0.22</td>
<td>0.11–0.36</td>
</tr>
<tr>
<td>Rebreeding interval</td>
<td>3</td>
<td>0.23</td>
<td>0.03–0.36</td>
</tr>
<tr>
<td><strong>Male traits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testis width</td>
<td>8</td>
<td>0.37</td>
<td>0.02–0.61</td>
</tr>
<tr>
<td>Testis weight</td>
<td>5</td>
<td>0.44</td>
<td>0.24–0.73</td>
</tr>
<tr>
<td>Semen volume</td>
<td>6</td>
<td>0.19</td>
<td>0.14–0.25</td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>6</td>
<td>0.19</td>
<td>0.13–0.26</td>
</tr>
<tr>
<td>Sperm motility</td>
<td>6</td>
<td>0.11</td>
<td>0.06–0.18</td>
</tr>
<tr>
<td>% Abnormal sperm</td>
<td>4</td>
<td>0.10</td>
<td>0.06–0.17</td>
</tr>
<tr>
<td>Libido</td>
<td>13</td>
<td>0.15</td>
<td>0.03–0.47</td>
</tr>
</tbody>
</table>
et al., 2004; Rosendo et al., 2007b; Roehe et al., 2009). The influence of piglet genes then increases, but remains lower than maternal genetic effects as far as piglet growth mainly depends on sow milk, i.e. until 3 to 4 weeks of age (Kaufmann et al., 2000; Solanes et al., 2004; Rosendo et al., 2007b).

Maternal effects strongly reduce or even vanish after weaning. Yet several authors have suggested an effect of the pre-weaning environment provided by the female’s dam on litter traits, which would reduce the efficiency of selection for litter size. Indeed, significant maternal genetic effects have been reported by some authors (Southwood and Kennedy, 1990; Ferraz and Johnson, 1993; Irgang et al., 1994), although low estimates were reported in other studies (Perez-Enciso and Gianola, 1992; Chen et al., 2003a). Such variations may reflect differences between populations or management conditions (e.g. cross-fostering or age at weaning). In any case, maternal effects, if present, should be considered in genetic evaluation for prolificacy (Rohe and Kennedy, 1993). Similarly, the hypothesis that successive litters might have a somewhat different genetic determination has been put forward to explain lower than expected performances remain very high in all studies (³0.8), whereas much lower values have been obtained between first and, to a lesser extent, second parities, compared with later parity records (Irgang et al., 1994; Roehe and Kennedy, 1995; Hanenberg et al., 2001; Noguera et al., 2002). Values in commercial herds with poorer environments are generally lower.

Genetic correlations between reproduction traits

Phenotypic and genetic correlations between male genital tract measurements are generally large (Toelle et al., 1984; Sellier et al., 2000). Testes measurements are favourably related to sperm production or percentage spermatogenesis (Toelle et al., 1984; Young et al., 1986; Huang and Johnson, 1996). An increased daily sperm production, a higher sperm concentration, a greater number of spermatozoa per ejaculate and a lower proportion of abnormal cells were found in a line selected for size of the testes at 150 days of age (Rathje et al., 1995; Huang and Johnson, 1996). Divergent selection of boars for testosterone production was also effective, with realized heritabilities of 0.15 and 0.24 in the low and high lines, respectively, and significant line differences in testicular size, epididymal weight and volume density of Leydig cells, but without any change in daily sperm production per gram of testes (Robison et al., 1994; Walker et al., 2004). Estimates of the genetic parameters of sperm characteristics are not very numerous (Falkenberg et al., 1988; Grandjot et al., 1997b; Smital et al., 2005; Wolf, 2009), but clearly show a strong negative correlation between semen volume and sperm concentration, as well as between sperm motility and the proportion of abnormal cells. Conversely, motility appears as nearly independent from both sperm volume and concentration, except in the study of Falkenberg et al. (1988). Finally, a five-generation divergent selection experiment on fat androstenone level (Willeke et al., 1987) resulted in a significant direct response, mainly in the high line, as well as in large significant correlated responses on plasma testosterone and conjugated oestrogen levels.

Genetic relationships between male and female reproductive traits have been estimated in some studies to examine the interest of using male traits as indirect selection criteria to improve female reproductive performance. Estimates of genetic correlations between testes measurements and age at first oestrus, ovulation rate or litter size are generally low and do not show any consistent trend (Schinckel et al., 1983; Toelle and Robison, 1985; Young et al., 1986; Johnson et al., 1994). Conversely, Smital et al. (2005) reported significantly negative and positive genetic correlations (0.3–0.4) between, respectively, sperm volume and sperm motility of boars and the litter size of their daughters. Selection for low-fat androstenone levels was also shown to result in a delayed puberty of females in the selection experiment of Sellier and Bonneau (1988).
Phenotypic and genetic correlations between several female traits are shown in Table 10.4. Age at puberty has, on average, close to zero relationships with ovulation rate and litter size at birth and at weaning, and weak negative, i.e. favourable correlations, with litter weights. Mean values for genetic correlations between age at first mating and sow fertility or longevity are also low (≤0.10 in absolute value). However, these average values hide noticeable differences between studies (Ruiz-Flores and Johnson, 2001), which are partly due to the limited accuracy of genetic parameter estimates, but also probably result from both differences in the genetic make-up of pig populations and differences in management practices, e.g. mean age at first service or feeding level, which may affect the magnitude of competition effects between growth and reproduction for resource allocation. Results are more homogeneous regarding the association between age at first service and weaning to oestrus interval, which appears as clearly positive in the different studies investigating this relationship (Hanenberg et al., 2001; Holm et al., 2005; Imboonta et al., 2007). Finally, nearly zero genetic correlations between age at puberty and traits characterizing the intensity of oestrus symptoms were obtained by Rydhmer et al. (1994).

Studies on relationships between rebreeding performance and other reproductive traits are not numerous in the literature. Most studies reported close to zero genetic correlations between weaning to service interval and litter size at birth (Hanenberg et al., 2001; Holm et al., 2005; Imboonta et al., 2007; Lundgren et al., 2010). No measurable correlative responses to selection for a reduced weaning to oestrus interval were reported (ten Napel et al., 1998). Yet some authors (Sterning et al., 1990; Lundgren et al., 2010) have suggested a potential antagonism between the investment of sows to raise their litter and their subsequent rebreeding performance. Though evidence for an antagonism between performance and fertility is still limited in pigs — only Lundgren et al. (2010) reported a negative, but not significant, genetic correlation between the interval from weaning to oestrus and litter weight gain — it is well documented in other species with high performance levels such as dairy cattle and should receive some attention in the future.

Unlike fertility traits, the genetic determination of prolificacy traits is now rather well documented. On average, ovulation rate has a moderate genetic antagonism (negative correlation) with prenatal survival ($r_g = −0.36$, Table 10.4), so that selection for ovulation rate results in limited correlative improvement in litter size at birth (Cunningham et al., 1979; Rosendo et al., 2007a). Several selection experiments have been implemented to increase litter size using different indirect criteria: an index of ovulation rate and embryo/prenatal survival (Johnson et al., 1984; Ruiz-Flores and Johnson, 2001; Table 10.4.

Table 10.4. Means of literature estimates of genetic and phenotypic correlations* among reproductive traits.

<table>
<thead>
<tr>
<th></th>
<th>AP</th>
<th>OR</th>
<th>PSR</th>
<th>TNB</th>
<th>FSR</th>
<th>NBA</th>
<th>BWSR</th>
<th>NW</th>
<th>LBW</th>
<th>L21W</th>
<th>ABW</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>−0.06</td>
<td>−0.08</td>
<td>−0.04</td>
<td>−0.07</td>
<td>−0.09</td>
<td>−0.10</td>
<td>−0.15</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>0.05</td>
<td>−0.36</td>
<td>0.32</td>
<td>−0.27</td>
<td>0.24</td>
<td>−0.38</td>
<td>0.01</td>
<td>0.24</td>
<td>0.03</td>
<td>−0.23</td>
<td></td>
</tr>
<tr>
<td>PSR</td>
<td>−0.01</td>
<td>0.14</td>
<td>0.50</td>
<td>0.3</td>
<td>0.55</td>
<td>−0.25</td>
<td>0.42</td>
<td>0.30</td>
<td>0.10</td>
<td>−0.41</td>
<td></td>
</tr>
<tr>
<td>TNB</td>
<td>−0.03</td>
<td>0.13</td>
<td>0.60</td>
<td>−0.25</td>
<td>0.92</td>
<td>−0.15</td>
<td>0.73</td>
<td>0.62</td>
<td>0.40</td>
<td>−0.41</td>
<td></td>
</tr>
<tr>
<td>FSR</td>
<td>−0.06</td>
<td>−0.15</td>
<td>−0.08</td>
<td>0.01</td>
<td>0.17</td>
<td>−0.01</td>
<td>−0.10</td>
<td>0.05</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBA</td>
<td>−0.03</td>
<td>0.12</td>
<td>0.40</td>
<td>0.91</td>
<td>0.15</td>
<td>−0.14</td>
<td>0.81</td>
<td>0.64</td>
<td>0.55</td>
<td>−0.34</td>
<td></td>
</tr>
<tr>
<td>BWSR</td>
<td>−0.11</td>
<td>−0.14</td>
<td>−0.12</td>
<td>0.08</td>
<td>−0.22</td>
<td></td>
<td>0.15</td>
<td>−0.07</td>
<td>0.65</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>NW</td>
<td>−0.01</td>
<td>0.03</td>
<td>0.36</td>
<td>0.71</td>
<td>0.47</td>
<td>0.79</td>
<td>0.55</td>
<td>0.67</td>
<td>0.81</td>
<td>−0.23</td>
<td></td>
</tr>
<tr>
<td>LBW</td>
<td>−0.03</td>
<td>0.07</td>
<td>0.55</td>
<td>0.79</td>
<td>0.43</td>
<td>0.82</td>
<td>0.09</td>
<td>0.71</td>
<td>0.65</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>L21W</td>
<td>−0.04</td>
<td>0.02</td>
<td>0.08</td>
<td>0.42</td>
<td>0.36</td>
<td>0.46</td>
<td>0.65</td>
<td>0.80</td>
<td>0.61</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>ABW</td>
<td>−0.17</td>
<td>−0.32</td>
<td>−0.40</td>
<td>−0.44</td>
<td>−0.17</td>
<td>0.10</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Genetic correlations above the diagonal, phenotypic correlations below.

AP, age at puberty; OR, ovulation rate; PSR, prenatal survival rate; TNB, total number born; FSR, farrowing survival rate; NBA, number born alive; BWSR, birth to weaning survival rate; NW, number weaned; LBW, litter weight at birth; L21W, 21-day litter weight; ABW, average birth weight.
Rosendo et al., 2007a), uterine capacity (Christenson et al., 1987) or placental efficiency (Wilson et al., 1999). Significant improvements in litter size have been obtained in several experiments (Ruiz-Flores and Johnson, 2001; Rosendo et al., 2007a), but none of them outperformed direct selection for litter size. Interestingly, results from successful selection for litter size show no improvement in prenatal survival and a proportional increase in ovulation rate (Driancourt et al., 1992; Tribout et al., 2003).

Genetic correlations between litter size at birth and at weaning are strongly positive ($r = 0.71$, Table 10.4). Conversely, litter size traits have some antagonistic relationships with piglet survival. In particular, farrowing survival rate is unfavourably correlated with ovulation rate ($r = -0.27$) and total number born ($r = -0.25$), but independent of number born alive and number weaned. Birth to weaning survival rate is negatively linked with ovulation rate and litter size at birth, and has a slightly positive correlation with the number of piglets weaned. Responses to selection are consistent with genetic parameters. Selection for ovulation rate or total number born has been shown to result in an increase in the number of stillborn piglets (Blasco et al., 1995; Ruiz-Flores and Johnson, 2001; Canario et al., 2007a; Guéry et al., 2009), whereas no deterioration of farrowing survival is observed when selecting for number born alive (Guéry et al., 2009).

Average birth weight appears to have negative genetic correlations with litter size and positive ones with survival rates and litter weights. Yet the relevance of increasing piglet birth weight to improve survival is strongly debated, as it would be associated with a higher nutritional demand from the litter at the end of gestation and during lactation, which may be detrimental to the sow (Bergsma et al., 2008). Moreover, bigger piglets might be associated with increased farrowing difficulties (Canario et al., 2006b). Some authors have suggested that improving the homogeneity of piglet weight at birth would result in higher piglet survival. Indeed, several studies have reported positive relationships between within-litter variation in birth weight and pre-weaning mortality (Roehe and Kalm, 2000; Knol et al., 2002b; Milligan et al., 2002; Huby et al., 2003). Additionally, results from a divergent selection experiment for the homogeneity of birth weight in rabbits showed a significant direct response to selection and lower farrowing (16.6% versus 18.6%) and birth to weaning (17.7% versus 32.7%) mortality rates of kits in the 'homogeneous' than in the 'heterogeneous' line (Garreau et al., 2008). However, the efficiency of such a criterion has been questioned by some authors owing to the important amount of work represented by the individual weighing of piglets and to the low heritability obtained in some pig populations (Wolf et al., 2008). Piglet maturity at birth has been shown to be another important issue in piglet survival. Indeed, embryos/piglets from Meishan dams have been shown to have a higher probability of survival and to be more mature at birth than embryos/piglets from European and/or American dams (Le Dividich et al., 1991; Canario et al., 2007b). Moreover, selection for lean tissue growth rate has been shown to reduce piglet maturity at birth (Herpin et al., 1993; Canario et al., 2007b). Yet finding simple criteria to characterize piglet maturity currently remains a challenge.

Farrowing kinetics is another important determinant of farrowing survival. The probability of stillbirth has been shown to be increased (Holm et al., 2004a; Canario et al., 2006a) in prolonged farrowings or with longer birth intervals because piglets are more prone to be asphyxiated or suffer some degree of hypoxia. Longer farrowings may also affect birth to weaning survival, as piglets that have suffered from hypoxia take a longer time to reach the udder and have a higher probability of being crushed or suffering from hypothermia. After birth, piglet survival and growth largely depend on the ability of females to take care of their progeny and to provide a sufficient amount of colostrum and milk. Sow behavioural traits are lowly-to-moderately heritable depending on test standardization conditions (Grandinson et al., 2003; Hellbrugge et al., 2008) and may affect piglet survival. As reviewed by Grandinson (2005), sows with a high responsiveness towards their offspring, a low fear response towards humans and a high frequency of successful nursings tend to have fewer piglet losses. Colostrum consumption, which provides the piglet with both energy and maternal antibodies, and can be indirectly estimated
from piglet growth rate over the first 24 h of life (Devillers et al., 2004), has also been shown to be an important determinant of piglet survival (Le Dividich et al., 2005). However, contrary to milk production, colostrum production was found to remain unchanged after 21 years of selection in French Large White breed pigs (Canario, 2006).

Milk production is difficult to record, but can be indirectly estimated from litter growth rate (Etienne et al., 2000); production is influenced by the stimuli from the suckling piglet and has increased as a consequence of higher sow prolificacy and more active piglets (Mackenzie and Revell, 1998; Canario, 2006). As selection has led to a strong reduction of sow body fat reserves and has had limited effects on lactational sow feed intake (Tribout et al., 2003), it has probably resulted in a more pronounced negative energy balance of sows during lactation, with potentially unfavourable effects on health, reproduction and longevity (Bergsma et al., 2008). This negative energy balance could be limited by increasing lactational sow feed intake, by decreasing piglet early growth potential or by increasing lactation efficiency – defined as the ratio of energy output for lactation to sow energy input (Bergsma et al., 2008). A last important issue is the number of functional teats, which can be a limit to the number of piglets a sow can nurse. Teat number has been shown to have a moderate heritability, to be genetically independent of production traits and litter size at birth, and to have a positive genetic correlation with litter size at weaning (Ligonesche et al., 1995; Zhang et al., 2000).

Genetic correlations with other economically important traits

Estimates of genetic correlations of reproduction with growth and carcass traits are rather numerous in the literature. In males, testes measurements have been shown to have positive genetic correlations with growth traits when measured at a constant age (Toelle et al., 1984; Young et al., 1986; Lubritz et al., 1991; Johnson et al., 1994). Genetic correlations are more variable when measurements occur at a constant weight, as both positive (Schinckel et al., 1983) and negative (Sellier et al., 2000) estimates have been obtained. Growth rate has also been found to be positively correlated with plasma testosterone level (Lubritz et al., 1991) and to have positive or null associations with fat androstenone levels (Willeke and Pirchner, 1989). Genetic correlations of reproduction traits with backfat thickness are generally low (Toelle et al., 1984; Sellier et al., 2000).

Age at puberty of gilts exhibits negative, i.e. favourable, genetic correlations with growth rate (Young et al., 1978; Hutchens et al., 1981; Rydhmer et al., 1992; Bidanel et al., 1996), and has weak relationships with backfat thickness or carcass lean content (Rydhmer et al., 1992, 1995; Bidanel et al., 1996; Serenius et al., 2004). Results regarding relationships between fertility and production traits are much more limited. Rydhmer et al. (1994) obtained negative correlations between traits characterizing the intensity of oestrus symptoms and growth rate, whereas relationships with carcass lean-ness were close to zero. Significant differences in conception rate were obtained between a line selected for high lean growth on scale feeding and a line selected for low lean growth (64% versus 83%) by Kerr and Cameron (1996). Adamec and Johnson (1997) and ten Napel et al. (1998) reported rather low genetic correlations (≤0.20 in absolute value) between re-breeding performance and growth or carcass traits, but pointed out a risk of downward bias due to culling and mentioned that selection for litter weight gain and against backfat might result in prolonged intervals.

Early literature reviews (Brien, 1986; Haley et al., 1988) concluded that litter size and weights are, on average, weakly correlated with performance traits. Yet several more recent results suggest the existence of slightly unfavourable genetic relations between the two sets of traits in populations with high performance levels. Negative genetic correlations between growth and litter size were estimated by Ducos and Bidanel (1996), Hermesch et al. (2000a), Zhang et al. (2000), Holm et al. (2004b) and Tribout and Bidanel (2008). The majority of more recent estimates of genetic relationships between litter traits and carcass lean content are slightly positive, i.e. unfavourable (e.g. Zhang et al.,
2000; Chen et al., 2003b; Serenius et al., 2004; Imboonta et al., 2007). The situation is less clear for feed intake and feed efficiency because of a much lower number of estimates. Close to zero relationships were reported by Hermesch et al. (2000b), but significant correlated responses of litter size to selection for high and low feed intake and for low residual feed intake were obtained by, respectively, Kerr and Cameron (1995) and Sellier et al. (2010).

With the exception of the above-mentioned relationships between fat androstenone level and sexual development, estimates of genetic correlations between reproduction and meat quality traits are not numerous in the literature. Most available estimates with meat pH, meat colour and meat water-holding capacity do not significantly differ from zero (Larzul et al., 1999; Hermesch et al., 2000a; Serenius et al., 2004; Rosendo et al., 2010) and tend to indicate that the two groups of traits can reasonably be considered as almost genetically independent.

**Individual Genes and Quantitative Trait Loci (QTLs) Affecting Reproduction Traits**

The dramatic advances in molecular biology and genomics over the past 20 years have profoundly changed our knowledge of the genetic determination of economically important traits in major livestock species. In particular, the development of panels of genetic markers covering the whole genome has allowed the individual loci underlying the genetic variance of quantitative traits of economic importance to be systematically detected and mapped. A large number of experiments aiming at detecting these QTLs have led to the identification of a large number of such loci (more than 5600 QTLs are currently referenced in the PigQTLdb accessible at http://www.animalgenome.org/cgi-bin/QTLdb/SS/index (Hu et al., 2007). It has to be noted that these QTLs have essentially been detected using low-density maps based on panels of microsatellite markers and not the high-density linkage maps based on single nucleotide polymorphism (SNP) markers that have recently become available (Ramos et al., 2009). QTLs for male and female pig reproductive traits detected so far are summarized in Tables 10.5 and 10.6, respectively. The main functional candidate genes associated with female reproductive traits that have been investigated are given in Table 10.7.

QTLs for male reproductive traits have been mapped using crosses between early maturing Chinese and European/American breeds. Most of these have a rather limited effect (less than 10% of trait variation), but a limited number of loci, located on chromosome X, explain up to 20% of the variance of several male reproductive traits, i.e. testicular and seminal vesicle weight, as well as seminiferous tubular diameter. Alleles from Chinese breeds increase testes weight at young ages, but tend to reduce weight at older ages. The SERPINA7 gene, which regulates the availability of thyroid hormones within tissues, has been reported as an interesting positional candidate by Ren et al. (2009), but its potential implication remains to be investigated. No fine mapping of autosomal QTLs has been carried out as yet, and the localization interval of QTLs remains large (generally above 20 cM).

QTLs for female reproductive traits are given in Table 10.6. QTLs for teat number, which are much more numerous than QTLs for true reproductive traits because its measurement is very easy, are not reported here. The 71 QTLs for teat number reported in PigQTLdb are located on all chromosomes except SSC (Sus scrofa chromosome) 14 and 18. They generally explain a limited proportion of trait variability, so that no fine-mapping programme has been developed. The most numerous QTLs then concern the most heritable traits, i.e. age at puberty and ovulation rate (23 and 22 QTLs, respectively, on PigQTLdb). QTLs for age at puberty have been mapped to ten different chromosomes, with overlapping localizations from independent studies on SSC1, on SSC7 in the SLA (swine leucocyte antigen) complex region and at the extremity of the short arm of SSC8. QTLs for ovulation rate have been detected on nine different chromosomes, with QTLs from several independent studies located on two different regions of SSC8, the first one in the centromeric region (Wilkie et al., 1999) and
<table>
<thead>
<tr>
<th>Trait</th>
<th>Pig chromosome</th>
<th>Population</th>
<th>Size</th>
<th>% variance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ejaculation duration</td>
<td>6, 17</td>
<td>DU × ER</td>
<td>177</td>
<td>7.7–7.9</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Ejaculation times</td>
<td>6, 16, 17</td>
<td>DU × ER</td>
<td>177</td>
<td>5.9–11.8</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Epididymal weight at:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 days of age</td>
<td>2</td>
<td>DU × ER</td>
<td>347</td>
<td>4.5</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>180 days of age</td>
<td>3, 4, 10, 13, 15</td>
<td>LW × MS</td>
<td>487</td>
<td>1.9–4.3</td>
<td>Bidanel et al., 2001</td>
</tr>
<tr>
<td>300 days of age</td>
<td>3, 7</td>
<td>DU × ER</td>
<td>347</td>
<td>4.5–27.3</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>Length of bulbo-urethral glands</td>
<td>1, 3, 7, 13</td>
<td>LW × MS</td>
<td>485</td>
<td>3.3–5.1</td>
<td>Bidanel et al., 2001</td>
</tr>
<tr>
<td>Plasma FSH level</td>
<td>3, 10, X</td>
<td>WC × MS</td>
<td>315</td>
<td></td>
<td>Rohrer et al., 2001</td>
</tr>
<tr>
<td>Plasma testosterone level</td>
<td>7, 13</td>
<td>DU × ER</td>
<td>347</td>
<td>7.3–14.3</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>Semen volume</td>
<td>3, 15, 18</td>
<td>DU × ER</td>
<td>177</td>
<td>7.9–8.6</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Seminiferous tubular diameter at:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 days of age</td>
<td>5, 13, 14, X</td>
<td>DU × ER</td>
<td>347</td>
<td>8.4–14.8</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>300 days of age</td>
<td>16</td>
<td>DU × ER</td>
<td>347</td>
<td>14.8</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>Seminal vesicles weight</td>
<td>1, 3, 4, 7, 11,15, 16, X</td>
<td>LW × MS</td>
<td>481</td>
<td>2.5–21.8</td>
<td>Bidanel et al., 2001</td>
</tr>
<tr>
<td>Sperm abnormality rate</td>
<td>4, 9</td>
<td>DU × ER</td>
<td>177</td>
<td>8.8–11.8</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Sperm concentration</td>
<td>17</td>
<td>DU × ER</td>
<td>177</td>
<td>9.5</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Sperm motility</td>
<td>4</td>
<td>DU × ER</td>
<td>177</td>
<td>6.3</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Sperm pH value</td>
<td>2, 6, 9</td>
<td>DU × ER</td>
<td>177</td>
<td>5.7–9.8</td>
<td>Xing et al., 2009</td>
</tr>
<tr>
<td>Testicular weight at:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 days of age</td>
<td>3, X</td>
<td>DU × MS</td>
<td>449</td>
<td>5.0–9.0</td>
<td>Sato et al., 2003</td>
</tr>
<tr>
<td>90 days of age</td>
<td>1, X</td>
<td>DU × ER</td>
<td>347</td>
<td>9.1–20.6</td>
<td>Ren et al., 2009</td>
</tr>
<tr>
<td>180 days of age</td>
<td>4, 7, 10, 13, 17, X</td>
<td>LW × MS</td>
<td>487</td>
<td>3.5–19.6</td>
<td>Bidanel et al., 2001</td>
</tr>
<tr>
<td>220 days of age</td>
<td>X</td>
<td>WC × MS</td>
<td>315</td>
<td></td>
<td>Rohrer et al., 2001</td>
</tr>
<tr>
<td>300 days of age</td>
<td>1, 5, 7, X</td>
<td>DU × ER</td>
<td>347</td>
<td>4.8–14.7</td>
<td>Ren et al., 2009</td>
</tr>
</tbody>
</table>

*DU, Duroc; ER, Erhualian; LW, Large White; MS, Meishan; WC, White European breed cross.

QTLs affecting litter size traits have been detected on 13 different chromosomes, but most of them are putative results and there is limited overlap between studies. Indeed, the only overlaps concern the results of Wilkie et al. (1999), Bidanel et al. (2001) and Li et al. (2009) on SSC6, of De Koning et al. (2001), Tribout et al. (2008) and Li et al. (2009) on SSC7, of Bidanel et al. (2001) and Noguera et al. (2009) on SSC13 and, finally, of De Koning et al. (2001) and Noguera et al. (2009) on SSC17. Some overlaps with candidate gene studies have also recently begun to appear. The leptin receptor gene (LEPR), which has been shown to be associated with variations in litter size by Chen et al. (2004b), is located in the confidence interval of the above-mentioned QTLs on SSC6 (Wilkie et al., 1999; Bidanel et al., 2008; Li et al., 2009). The SSC7 QTLs (De Koning et al., 2001; Tribout et al., 2008; Li et al., 2009) are located in the region of the properdin locus (CFP), which was found to be associated with litter size by Buske et al. (2005). The prolactin receptor locus (PRLP; Vincent et al.,
J.-P. Bidanel (1998) is located on SSC16, close to the confidence interval boundary of the QTLs affecting number born alive detected by Tribout et al. (2008). Conversely, no QTL has so far been detected in the **ESR1** gene region, a candidate gene known to be associated with litter size in several populations (Table 10.7). It is currently not known whether these candidate genes are only additional markers associated with the observed variability or are responsible for these variations. The existence of non-significant results in some populations would tend to indicate that the polymorphism investigated is not the causative mutation. Although several non-significant results from very small populations have been reported in the literature (e.g. Drogemuller et al., 2001; Gibson et al., 2002) for the **ESR1** gene, a bias from unpublished non-significant results is likely to exist.

Finally, it should be emphasized that the simple additive models that have in most cases been used so far are likely to detect only a (small) fraction of the genetic variability of traits such as litter size that are known to be affected by non-additive gene effects. Indeed, no fewer than 18 epistatic QTLs affecting number born alive were detected by Noguera et al. (2009) using a bidimensional genome scan versus two QTLs using conventional one-dimensional scans.

The effects on reproductive traits of some of the major genes identified so far have been investigated. The halothane sensitivity gene

### Table 10.6. Quantitative trait loci (QTLs) for female reproductive traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Pig chromosome number</th>
<th>Populationa</th>
<th>Size</th>
<th>Variance (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at puberty</td>
<td>1, 4, 6, 7, 13</td>
<td>LW × MS</td>
<td>476</td>
<td>3.0–10.0</td>
<td>Bidanel et al., 2008</td>
</tr>
<tr>
<td></td>
<td>7, 8, 12</td>
<td>LW × LR</td>
<td>295</td>
<td>2.7–9.7</td>
<td>Cassady et al., 2001</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>LW × LR</td>
<td>295</td>
<td></td>
<td>Holl et al., 2004</td>
</tr>
<tr>
<td></td>
<td>1, 10</td>
<td>WC × MS</td>
<td>344</td>
<td></td>
<td>Rohrer et al., 1999</td>
</tr>
<tr>
<td></td>
<td>1, 7, 8, 17</td>
<td>DU × ER</td>
<td>454</td>
<td>2.0–8.0</td>
<td>Yang et al., 2008</td>
</tr>
<tr>
<td>Ovulation rate</td>
<td>4, 5, 7, 9, 13</td>
<td>LW × MS</td>
<td>502</td>
<td>3.9–5.9</td>
<td>Bidanel et al., 2008</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>LW × LR</td>
<td>295</td>
<td>3.4</td>
<td>Cassady et al., 2001</td>
</tr>
<tr>
<td></td>
<td>4, 8, 13, 15</td>
<td>LW × LR</td>
<td>114</td>
<td>5.1–10.9</td>
<td>Rathje et al., 1997</td>
</tr>
<tr>
<td></td>
<td>3, 8, 9, 10, 15</td>
<td>WC × MS</td>
<td>344</td>
<td></td>
<td>Rohrer et al., 1999</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>DU × MS</td>
<td>234</td>
<td></td>
<td>Sato et al., 2006</td>
</tr>
<tr>
<td></td>
<td>7, 8, 15</td>
<td>YO × MS</td>
<td>104</td>
<td></td>
<td>Wilkie et al., 1999</td>
</tr>
<tr>
<td>Number of embryos</td>
<td>9, 12, 18</td>
<td>LW × MS</td>
<td>468</td>
<td>2.8–7.2</td>
<td>Bidanel et al., 2008</td>
</tr>
<tr>
<td>Uterine capacity</td>
<td>8</td>
<td>WC × MS</td>
<td>187</td>
<td></td>
<td>Rohrer et al., 1999</td>
</tr>
<tr>
<td>Gestation length</td>
<td>1, 9, 15</td>
<td>YO × MS</td>
<td>104</td>
<td>9.4–23.6</td>
<td>Wilkie et al., 1999</td>
</tr>
<tr>
<td>Number mummified</td>
<td>2, 6, 12</td>
<td>LW × LR</td>
<td>279</td>
<td></td>
<td>Holl et al., 2004</td>
</tr>
<tr>
<td>Total number born</td>
<td>11</td>
<td>LW × LR</td>
<td>279</td>
<td>5.1</td>
<td>Cassady et al., 2001</td>
</tr>
<tr>
<td></td>
<td>7, 12, 14, 17</td>
<td>LW/LR × MS</td>
<td>269</td>
<td>2.7–8.8</td>
<td>De Koning et al., 2001</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>LW × MS</td>
<td>152</td>
<td></td>
<td>King et al., 2003</td>
</tr>
<tr>
<td></td>
<td>7, 15</td>
<td>DU × ER</td>
<td>299</td>
<td>2.8–4.3</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>13, 17</td>
<td>IB × MS</td>
<td>881</td>
<td></td>
<td>Noguera et al., 2009</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>YO × MS</td>
<td>104</td>
<td></td>
<td>Wilkie et al., 1999</td>
</tr>
<tr>
<td>Number of stillborn</td>
<td>5, 13</td>
<td>LW × LR</td>
<td>279</td>
<td>7.9</td>
<td>Cassady et al., 2001</td>
</tr>
<tr>
<td></td>
<td>12, 14</td>
<td>LW × LR</td>
<td>279</td>
<td></td>
<td>Holl et al., 2004</td>
</tr>
<tr>
<td></td>
<td>7, 8</td>
<td>DU × ER</td>
<td>299</td>
<td>3.7–5</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>6, 11, 14</td>
<td>LW, LR</td>
<td>152</td>
<td></td>
<td>Tribout et al., 2008</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>YO × MS</td>
<td>104</td>
<td></td>
<td>Wilkie et al., 1999</td>
</tr>
<tr>
<td>Number born alive</td>
<td>11</td>
<td>LW × LR</td>
<td>279</td>
<td></td>
<td>Cassady et al., 2001</td>
</tr>
<tr>
<td></td>
<td>6, 15</td>
<td>DU × ER</td>
<td>299</td>
<td>3.7–5</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>13, 17</td>
<td>IB × MS</td>
<td>881</td>
<td></td>
<td>Noguera et al., 2009</td>
</tr>
<tr>
<td></td>
<td>7, 16, 18</td>
<td>LW, LR</td>
<td>152</td>
<td></td>
<td>Tribout et al., 2008</td>
</tr>
</tbody>
</table>

*aDU, Duroc; ER, Erhualian; IB, Iberian pig; LR, Landrace; LW, Large White; MS, Meishan; WC, White European breed cross; YO, Yorkshire*
Table 10.7. Candidate genes associated with female reproductive traits in pigs. Adapted from Buske et al., 2006a; Onteru et al., 2009.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Trait</th>
<th>Population</th>
<th>Genotyped pigs (number)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESR1</td>
<td>Pvull site (intron)</td>
<td>TNB, NBA</td>
<td>MS × SL; LW</td>
<td>161; 1079; 4262</td>
<td>Rothschild et al., 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNB, NBA, TN</td>
<td>LW</td>
<td></td>
<td>Short et al., 1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNB, NBA</td>
<td>LW × MS</td>
<td>275</td>
<td>Van Rens et al., 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNB, NBA</td>
<td>LW</td>
<td>1030</td>
<td>Goliashova and Wolf, 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNB, NBA</td>
<td>LW</td>
<td>226</td>
<td>Horogh et al., 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNB, NBA</td>
<td>SL</td>
<td>408</td>
<td>Munoz et al., 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP</td>
<td>DU/LR × CW</td>
<td>376</td>
<td>Kuehn et al., 2009</td>
</tr>
<tr>
<td>PAX5</td>
<td>C/T (exon 5)</td>
<td>TNB, NBA</td>
<td>YO × ER</td>
<td>289</td>
<td>Li et al., 1998</td>
</tr>
<tr>
<td>FSHB</td>
<td>FSHBMS microsatellite (5' flanking region)</td>
<td>TNNB, NBA</td>
<td>LW × MS</td>
<td>402</td>
<td>Li et al., 2008</td>
</tr>
<tr>
<td>EPOR</td>
<td>Intron 4</td>
<td>NW, LWW, GL</td>
<td>LW × MS</td>
<td>4-way cross</td>
<td>Vallet et al., 2005;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uterine capacity</td>
<td>402</td>
<td></td>
<td>Nonneman et al., 2006</td>
</tr>
<tr>
<td>LEPR</td>
<td>Intron 2, exon 2, exon 18</td>
<td>Litter size</td>
<td>YO; DU</td>
<td>62; 246</td>
<td>Chen et al., 2004b</td>
</tr>
<tr>
<td>FUT1</td>
<td>Exon 2</td>
<td>TNNB, NBA</td>
<td>PBP</td>
<td>104</td>
<td>Horak et al., 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNNB, NBA</td>
<td>(LW × LR) × Le</td>
<td>123</td>
<td>Buske et al., 2006b</td>
</tr>
<tr>
<td>RNF4</td>
<td>C/T (intron 5)</td>
<td>TNNB, NBA</td>
<td>QP</td>
<td>159</td>
<td>Niu et al., 2009</td>
</tr>
<tr>
<td>BF</td>
<td>Intron 1</td>
<td>TNNB, NBA</td>
<td>QP</td>
<td>123</td>
<td>Buske et al., 2005</td>
</tr>
<tr>
<td>GNRHR</td>
<td>3′ UTR</td>
<td>OR</td>
<td>MS × LW</td>
<td>200</td>
<td>Jiang et al., 2001</td>
</tr>
<tr>
<td>OPN</td>
<td>Intron</td>
<td>TNB, NBA</td>
<td>SL</td>
<td>519</td>
<td>Korwin-Kossakowska et al., 2002</td>
</tr>
<tr>
<td>LIF</td>
<td>Exon 3</td>
<td>NBA</td>
<td>LR; LW</td>
<td>850; 604</td>
<td>Spotter et al., 2009</td>
</tr>
<tr>
<td>AKR1C2</td>
<td>Ile16Phe (Nt179 in coding region)</td>
<td>AP, OR, TN</td>
<td>1/4 MS</td>
<td>191</td>
<td>Nonneman et al., 2006</td>
</tr>
<tr>
<td>RBP4</td>
<td>(Intron)</td>
<td>TNNB, NBA</td>
<td>SL</td>
<td>1300</td>
<td>Rothschild et al., 2000</td>
</tr>
<tr>
<td>PRLR</td>
<td>Alu site</td>
<td>TNNB, NBA</td>
<td>LR; LW</td>
<td>850; 604</td>
<td>Spotter et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TNNB, NBA</td>
<td>LW; MS; LR</td>
<td>400; 261; 416</td>
<td>Vincent et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AP, OR, TN, NBA</td>
<td>SL</td>
<td>273</td>
<td>Drogemuller et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MS × LW/ LR</td>
<td>55–77</td>
<td>Van Rens and Van der Lende, 2002; Van Rens et al., 2003</td>
</tr>
<tr>
<td>LEP</td>
<td>(Exon 3)</td>
<td>TNNB, NBA</td>
<td>SL</td>
<td>519</td>
<td>Korwin-Kossakowska et al., 2002</td>
</tr>
<tr>
<td></td>
<td>Exon 3</td>
<td>Litter size</td>
<td>YO; LR</td>
<td>62; 170</td>
<td>Chen et al., 2004a</td>
</tr>
<tr>
<td></td>
<td>Intron 1</td>
<td>Litter size</td>
<td>DU</td>
<td>246</td>
<td>Chen et al., 2004a</td>
</tr>
</tbody>
</table>

\(^a\) ESR1, estrogen receptor 1; PAX5, paired box 5; FSHB, follicle stimulating hormone beta; EPOR, erythropoietin receptor; LEPR, leptin receptor; FUT1, fucosyl transferase 1; RNF4, ring finger protein 4; BF, properdin; GNRHR, gonadotrophin releasing hormone receptor; OPN, osteopontin; LIF, leukaemia inhibitory factor; AKR1C2, aldo keto reductase 1C2; RBP4, retinol binding protein 4; PRLR, prolactin receptor; LEP, leptin.

\(^b\) SSC, Sus scrofa chromosome.

\(^c\) AP, age at puberty; GL, gestation length; LWW, litter weight at weaning; NBA, number born alive; NW, number weaned; OR, ovulation rate; TN, teat number; TNNB, total number born.

\(^d\) DU, Duroc; ER, Erhualian; Le, Leicoma; LR, Landrace; LW, Large White; MS, Meishan; PBP, Prestice Black Pied; QP, Qingping; SL, Synthetic line; YO, Yorkshire.

\(^e\) 3′ UTR, three prime untranslated mRNA region.
(HAL; Fujii et al., 1991) and the RN gene (Milan et al., 2000) are generally considered to have limited effects on reproductive traits. Significant effects of the halothane gene region were reported in some early studies but, as discussed by Sellier et al. (1987), these are likely to be due to the presence of a QTL affecting reproduction in linkage disequilibrium with the halothane locus. Indeed, QTLs affecting reproduction traits have been detected in the vicinity of both the HAL and the RN genes (Tables 10.5 and 10.6), so that apparent effects of these two genes on reproduction would occur in the case of linkage disequilibrium with the QTLs concerned. Such disequilibrium could be an explanation of the effect of the IGF2-Intron3-G3072A mutation (Van Laere et al., 2003) on litter size reported by Buys et al. (2006).

**Gene Expression Results**

The possibility of large-scale gene expression analysis is a useful complementary approach to understand the biological basis of reproductive function. For instance, Bonnet et al. (2008) tried to identify the genes differentially expressed in pig granulosa cells along the terminal ovarian follicle growth. They showed in particular the down-expression of ribosomal protein, cell morphology and ion-binding genes, and the differential expression of genes involved in lipid metabolism. Ross et al. (2009) investigated gene expression during porcine conceptus rapid trophoblastic elongation and attachment to the uterine luminal epithelium. When comparing filamentous day 12 conceptuses with large spherical conceptuses, as many as 482 genes were statistically different, with a greater than twofold change in expression. The genes represented a large number of biological processes associated with cell motility, ATP utilization, cell growth, metabolism and intracellular transport. Expression analysis of genetic variation has so far been limited. A transcriptome analysis of pig folliculogenesis was performed on pigs from lines selected for reproductive traits by Caetano et al. (2004) and Gladney et al. (2004). These studies showed that genes involved in steroid biosynthesis (cytochrome P450 side chain cleavage enzyme, steroidogenic acute regulatory protein), tissue remodelling (plasminogen activator inhibitor III) and apoptosis (calpain light chain I) were differentially expressed between lines. Approaches such as eQTL (expression quantitative trait locus) analyses have not yet been used to investigate pig reproduction owing to cost considerations, but should be of great interest in deciphering the genetic variability of this complex function.

**Conclusions**

Large gains in pig reproductive efficiency have been achieved over the last two decades. Sows from the highest performing maternal lines now farrow almost 15 piglets a litter, i.e. 30% more than 20 years ago. This large improvement has often been necessary to allow breeders to maintain some profit from their activities. It has been accompanied by some adverse effects such as increased piglet mortality, and also raises questions about the ability of sows nursing very large litters without impairing their health, their longevity and their welfare. Much remains to be known in order to understand the physiological and genetic basis of the complex interactions between the sow and its litter, and to find ways to improve lactation efficiency without impairing sow longevity and piglet post-weaning growth. The new genomic revolution associated with the sequencing of the pig genome and the availability of new tools such as high-density SNP chips and new sequencing technologies give scientists and breeders the opportunity to benefit from the use of much more powerful methods to understand the biological bases of genetic variation and to manage livestock populations more efficiently. In particular, genomic selection tools will undoubtedly result in more efficient genetic improvement programmes and potentially give scientists huge amounts of data to decipher trait genomic variability, provided that more accurate phenotypes are available to understand the complex relationships between genotype and phenotype.
References


The emergence and utilization of new reproductive management strategies and tools have had a substantial impact on reproductive capabilities for the swine industry and for research labs using pigs as research models. Artificial insemination (AI) in pigs has significantly extended the genetic resources available to commercial and seed-stock producers as well as researchers on a global scale. In addition to AI, further development and commercial utilization of other techniques such as sperm and embryo cryopreservation, deep-intrauterine insemination and surgical embryo

Transgenics and Modern Reproductive Technologies

Jason W. Ross¹ and Randall S. Prather²

¹Iowa State University, USA; ²University of Missouri, USA

Introduction 242
Modern Reproductive Technologies 243
  Oestrus synchronization 243
  Breeding strategies 243
  Semen and embryo cryopreservation 244
  Embryo transfer 244
  Cloning 245
The Utilization of Transgenic Pigs 245
  Current use of transgenic pigs 245
Transgenic Pig Development 246
  Pronuclear injection 246
  Sperm- and intracytoplasmic sperm injection (ICSI)-mediated transgenesis 247
  Oocyte transduction 247
  Genetic modification of a somatic cell followed by nuclear transfer 247
Genetic Modification of Somatic Cells for Nuclear Transfer 248
  Exogenous DNA delivery into the potential donor cell 248
  Random integration 249
  Targeted modifications 250
  Conditional/inducible genetic engineering 250
  Type of donor cell affects success of somatic cell nuclear transfer (SCNT) 251
  Improving somatic cell remodelling and reprogramming 251
Future of Transgenic Pigs 252
  Reproductive efficiency 252
  Growth and development 252
  Disease resistance 253
  Transgenic pigs for xenotransplantation 253
  Transgenic pigs for biomedical models 253
Conclusions 254
References 254
transfer (ET) will enable producers and researchers to address continuing and emerging problems affecting pig production and research. Another area that has dramatically changed in recent years is the use of cloning of and transgenic modification to pigs. Transgenic modifications to pigs are occurring at a rapid pace and are significantly enhancing the pig’s potential for use in both production agriculture and biomedical disease modelling, while cloning permits the reproduction of animals demonstrating rare phenotypes. The rapid progression of transgenic modifications to pigs is largely a result of significant advancements in somatic cell nuclear transfer (SCNT). As SCNT efficiency continues to improve, novel strategies for genetic modifications to mammalian genomes will provide additional methods for studying conditional ‘knockouts’ and multiple ‘knock-ins’ to better understand the physiology of the pig and specific gene function. This knowledge will result in more efficient use of the pig for agriculture, biomedicine and disease modelling.

**Modern Reproductive Technologies**

In any livestock venture, reproductive performance is often the number one contributor to profitability. Improved reproductive performance requires an understanding and utilization of innovative approaches to the synchronization of breeding and farrowing, while maintaining adequate litter size and maximizing the contribution of elite genetics to future progeny. Below we have outlined a number of reproductive technologies that are currently being applied, as well as discussing where the science is going.

**Oestrus synchronization**

Synchronization of oestrus in swine herds is essential to maintaining large farrowing groups and has permitted efficiencies in production, management and marketing. The most common method of oestrus synchronization in sows is weaning litters at the same time. Generally, most of the sows will begin oestrus within 4–7 days following weaning. This method is widely used throughout the swine industry and is consistently reliable in sows. Gilts, in contrast, represent a significant proportion of the breeding inventories for most swine herds, and synchronizing gilt oestrus can be more challenging. One reliable method of synchronizing oestrous cycles in gilts is the synchronization of puberty onset. A well-established approach is the injection of PG 600 to hormonally induce follicular growth in prepubertal gilts (Guthrie, 1977); this can also be effective in acyclic post-pubertal females.

In cattle, a single injection of exogenous prostaglandin F2α effectively regresses the corpus luteum (CL), providing a premature return to oestrus. However, in pigs the CLs are largely unresponsive to a single dose of prostaglandin F2α and multiple injections are only effective during the later stages of dioestrus (Hallford et al., 1975; Estill et al., 1993). In the absence of competent embryos, the pig CL normally initiates regression around days 15–18. The requirements of multiple injections that have a short window of effectiveness make this a cost-prohibitive strategy.

While luteolytic agents are not widely used in the swine industry, the use of progesterone receptor agonists has become the method of choice in manipulating the oestrous cycles of females. The most common strategies involve the use of altrenogest (Matrix®). The majority of randomly cycling gilts fed 15 mg/head of altrenogest daily for approximately 14 days will come into oestrus 5–6 days after withdrawal of the altrenogest (Stevenson and Davis, 1982). The use of PG 600 (400 and 200 international units of pregnant mare serum gonadotrophin (PMSG) and human chorionic gonadotrophin (HCG), respectively) following oestrus synchronization with altrenogest can increase the number of ovulations by about eight, and shortens the altrenogest withdrawal to oestrus interval by about 24 h (Estienne et al., 2001). However, some evidence suggests that the additional embryos produced as a result of PMSG and HCG stimulation have compromised developmental ability (Ziecik et al., 2005).

**Breeding strategies**

AI has dramatically advanced the genetic improvement of pigs and is the standard
operating procedure on major swine farms. The downside of AI in pigs is that the long uterine horns of females require a volume and sperm count that limits the boar to producing an average of 20–30 AI doses per ejaculate. To overcome this, investigators have been developing approaches to insemination that can reduce both the volume and sperm numbers required per dose. These approaches require the delivery of the semen past the cervix (i.e. post-cervical or intrauterine insemination) or further into the uterine horns (deep intrauterine insemination, also referred to as low-dose insemination) (Martinez et al., 2001; Watson and Behan, 2002; Mezalira et al., 2005). Strategies capable of maintaining average litter size while significantly reducing sperm numbers are required to advance the utilization of sex-sorted sperm and frozen–thawed semen to widespread commercial application.

Finding and using techniques that reduce the volume and the number of sperm cells required for insemination and pregnancy establishment while maintaining litter size requirements will be beneficial for the exploitation of elite boars. Currently, swine producers typically utilize two doses of semen, each about 24 h apart, and beginning at the onset of oestrus. The reason that two doses are used is because the length of behavioural oestrus is highly variable between females (24–72 h), and, given that the timing of ovulation relative to the duration of oestrus is quite variable (Soede and Kemp, 1997; Almeida et al., 2000), it is difficult to establish appropriate insemination timing for a single dose. Investigators are now developing methods to couple ovulation synchronization to currently used oestrus synchronization strategies in order to limit AI to a single dose. These methods include the use of gonadotrophin-releasing hormone agonists (Baer and Bilkei, 2004; Brussow et al., 2007; Martinat-Botté et al., 2010) and luteinizing hormone (Degenstein et al., 2008).

Semen and embryo cryopreservation

Cryopreservation of semen and embryos from pigs has presented more of a challenge than in many other species. Both semen and embryos are sensitive to freeze–thawing. Although boar semen has been frozen and used for AI for 35 years (Pursel and Johnson, 1975; Westendorf et al., 1975), the success rates are still low. Boar semen is sensitive to changes in osmotic balance, oxidative stress, low temperatures and the toxic effects of exposure to cryoprotectants (Rath et al., 2009). There are also significant differences between individual boars in the ability of their sperm to survive cryopreservation (Rath et al., 2009; Roca et al., 2009). As sperm survivability after cryopreservation can be low, deep uterine insemination is the preferred method of AI (Roca et al., 2009). With improvements in procedures for cryopreservation and AI, the application of other technologies, such as using gender-sorted semen, might become commonplace.

Similar to sperm, pig embryos are especially sensitive to a decrease in temperature. Interestingly, removal of the numerous lipid droplets appears to alleviate this sensitivity (Wilmut, 1972; Polge et al., 1974; Nagashima et al., 1994; Johnson et al., 2005). However, the lipid removal techniques generally compromise the zona pellucida, thus creating the possibility of pathogen entry. A non-invasive method, involving no compromise of the zona pellucida, has been developed for high-throughput cryopreservation of in vitro-produced embryos (Li et al., 2009), and we are currently perfecting the technique with in vivo-produced embryos. Other strategies, such as destabilizing the cytoskeleton (Dobrinsky et al., 2000), altering the vitrification conditions (Berthelot et al., 2001; Misumi et al., 2003; Beebe et al., 2005; Somfai et al., 2008) or using a solid surface for cryopreservation (Somfai et al., 2008), have been successfully employed. Additional research is needed to develop these technologies for widespread commercial application.

Embryo transfer

Utilization of ET in the swine industry is much less advanced than the significant progress in the beef industry. The greatest advantage of any ET programme is that it permits an increase of the genetic contributions of select females to the herd. While this is a tremendous advantage to the beef and dairy industries, where females
produce only a single progeny per year, the pay-off is not as significant in pigs. Most sows are capable of farrowing an average of 2.4 litters a year, giving them the ability to transmit their genetics to more than 20 progeny a year.

Perhaps one of the most limiting aspects of ET in pigs is that non-surgical collection and transfer of embryos is not nearly as successful as surgically removing and transferring embryos. While investigators have been capable of producing piglets following non-surgical ET (Li et al., 1996; Cuello et al., 2005), the efficiency still lags significantly behind surgical methodologies. Because of the inefficiency of non-surgical methods, ET is being used for primarily specialized purposes, such as the transfer of transgenic and/or cloned embryos and, for example, determining the developmental competency of embryos following in vitro manipulations (Li et al., 2009).

More recently, pig ET has become somewhat commercialized in an effort to flush embryos encapsulated in their zona pellucida from diseased herds, appropriately wash/sanitize them (Bureau et al., 2005) and transfer them to recipient pigs that have a specific pathogen-free status. While this strategy is not economically viable for many swine herds, it is being utilized as a method for salvaging elite and/or valuable genetic stock from herds before depopulation.

The Utilization of Transgenic Pigs

A transgenic pig is one whose genetic material has been altered to influence the expression and abundance of specific genes, with the ultimate objective of producing a specific, desired phenotype. This is most commonly done by introducing exogenous DNA, produced through recombinant DNA technology, into the genome of the pig. The appropriate production method and expected results can vary significantly, depending on whether the desired modification is random, targeted and/or tissue specific. The ability to produce transgenic pigs has dramatically improved over the past 10–15 years. Here we will describe the uses of transgenic pigs, the methods used to produce them and the strategies being utilized to improve the efficiency of transgenic pig production.

Current use of transgenic pigs

While large animals have proven to be very functional for our understanding of physiology, cell biology, molecular biology and genetics, our current ability to make genetic modifications to pigs has dramatically expanded their usefulness. Utilization of transgenic pigs to create models of human diseases and pathologies is already of value to the scientific community.

One unique example of the value that transgenic livestock offer is the use of their mammary glands for the production of heterologous proteins. Surprisingly, in addition to dairy breeds of cattle and goats, the pig is also being used as a source of heterologous proteins, such as human coagulation factors VIII and IX (Paleyanda et al., 1997; Lindsay et al., 2004) and protein C (Van Cott, 1999). While, clearly, pigs are not the only species capable of being genetically modified to produce a desired protein product in their milk, they offer some unique advantages compared with other species. The reason these genetic modifications have been made in pigs in contrast to in a dairy breed of cattle is that pigs are
smaller and less expensive to create and maintain, have a shorter gestation length, reach puberty faster and, being litter bearing, allow for a rapid expansion of the transgenic herd. Also of potential importance is that post-translational modification of transgene protein products is thought to occur more accurately in pigs (Van Cott et al., 2004). This could be particularly important with regard to the production of proteins such as haemophilic factors that require specific post-translational events that may influence their biological activity and half-life, such as the appropriate glycosylation of glycoproteins.

Valuable disease models in pigs include retinitis pigmentosa (Petters et al., 1997; Ross et al., 2009), cystic fibrosis (Rogers et al., 2008a,b), diabetes (Renner et al., 2008), mammary tumours (Yamakawa et al., 1999) and, more recently, Alzheimer’s syndrome (Kragh et al., 2008), to name but a few. Genetic modifications to study the incidence of pathologies such as cardiovascular disease in pigs and in humans have also been made (Lai et al., 2006). In addition to these applications, animals have been made that have tetracycline-dependent expression (Kues et al., 2006), or express marker genes such as the enhanced green fluorescent protein (eGFP) (Park et al., 2001), thus enabling a novel method of tracking cells transplanted into wild-type animals.

### Transgenic Pig Development

Numerous milestones have been achieved in the development of methods capable of producing cloned and transgenic pigs (Table 11.1). The development of efficient methods for producing transgenic large animals has been and still is essential to understanding the biology and increasing the utility of pigs. The first transgenic animal created was a mouse, produced nearly 30 years ago (Gordon et al., 1980; Brinster et al., 1981; Costantini and Lacy, 1981; Wagner et al., 1981). A few years later, gene ‘knockout’ mice were created by utilization of homologous recombination in embryonic stem (ES) cell lines (Doetschman et al., 1987; Thomas and Capecchi, 1987). Transmission of these methodologies to pigs is rapidly progressing, although not without some difficulty. Several primary methods have been used to create genetically modified pigs (Table 11.2), each of which can have specific advantages and limitations depending on the desired outcome. Below we will briefly describe the primary methods that have produced the majority of transgenic pigs.

#### Pronuclear injection

The first transgenic animals created were mice produced during the early 1980s through pronuclear injection of DNA in the one-cell stage zygote (Gordon et al., 1980; Brinster et al., 1981; Costantini and Lacy, 1981; Wagner et al., 1981). The process involves the injection of linearized DNA into the pronuclei of a single-cell zygote, and is repeatable and capable of producing offspring with a randomly integrated transgene. The utilization of the technique developed in rodents followed suit in large animal species as well, and has resulted in the production of transgenic cattle (Bondioli et al., 2008).

### Table 11.1. A timeline of some of the significant achievements leading to the ability to produce transgenic pigs.

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transgenic pig via pronuclear injection</td>
<td>Hammer et al., 1985</td>
</tr>
<tr>
<td>Cloned pig from embryonic cells</td>
<td>Prather et al., 1989</td>
</tr>
<tr>
<td>Sperm mediated gene transfer in pigs</td>
<td>Lavitrano et al., 1997</td>
</tr>
<tr>
<td>Cloned pig from fetal and adult cells</td>
<td>Polejaeva et al., 2000</td>
</tr>
<tr>
<td>Transgenic pig via oocyte transduction</td>
<td>Cabot et al., 2001</td>
</tr>
<tr>
<td>Transgenic pig via somatic cell nuclear transfer (SCNT)</td>
<td>Park et al., 2001</td>
</tr>
<tr>
<td>Gene ‘knockout’ pig using SCNT</td>
<td>Lai et al., 2002b</td>
</tr>
<tr>
<td>Intracytoplasmic sperm injection (ICSI)-mediated transgenic pig</td>
<td>Kurome et al., 2007</td>
</tr>
<tr>
<td>Gene ‘knock-in’ pig via SCNT</td>
<td>Rogers et al., 2008a</td>
</tr>
</tbody>
</table>
et al., 1991), pigs (Hammer et al., 1985; Vize et al., 1988; Petters et al., 1997; Bleck et al., 1998) and goats (Wang et al., 2002).

Sperm- and intracytoplasmic sperm injection (ICSI)-mediated transgenesis

Another strategy for producing transgenic animals is by using sperm for transgene delivery (Lavitrano et al., 1989). Sperm-mediated transgenesis can occur by AI or by intracytoplasmic sperm injection (ICSI), with sperm carrying the transgene concerned. ICSI-mediated transgenesis occurs following the chemical or mechanical disruption of the sperm membrane, a short incubation with double-stranded DNA, injection into a metaphase II-arrested oocyte and subsequent activation (Perry et al., 1999; Kurome et al., 2006; Pereyra-Bonnet et al., 2008). The procedure overcomes the tedious task of introducing the DNA into the pronuclei by utilizing the fractured membrane of sperm to carry exogenous DNA into the oocyte. In pigs, ICSI-mediated delivery of a transgene produced seven fetuses from 219 oocytes fertilized via ICSI-mediated gene transfer, of which two were transgenic (Kurome et al., 2007). Sperm-mediated transgene delivery has been utilized for the production of human decay accelerating factor in transgenic pigs (Lavitrano et al., 1997) and appears to be somewhat efficient in incorporating multiple transgenes simultaneously (Webster et al., 2005).

Oocyte transduction

Genetic modification of germ cells has to date only been achieved in the oocyte. For example, a replication-defective retrovirus can be used to infect an unfertilized oocyte. These vectors integrate into metaphase chromosomes. Because the oocyte is arrested in metaphase II of meiosis, this is an excellent strategy for obtaining transgene integration. The process was first demonstrated in cattle by injecting the vector under the zona pellucida, permitting infection and integration, fertilizing the oocyte, culturing to the blastocyst stage, and then ET (Chan et al., 1998). A similar strategy was used in pigs to create the first eGFP-transgenic pigs (Cabot et al., 2001).

Genetic modification of a somatic cell followed by nuclear transfer

SCNT involves the enucleation of a metaphase II-arrested oocyte and replacement of the original nucleus with the nucleus from a somatic cell, followed by activation of the reconstructed embryo. Somatic cell cloning has been essential for use in domestic livestock species because of the inability to isolate true embryonic stem cells in these species. The first domestic animal cloned from adult somatic cells was Dolly the sheep, born in 1996; the cloning used mammary gland cells for donor nuclei (Wilmut et al., 1997). Since this original demonstration that the mammalian oocyte is capable of remodelling and reprogramming a differentiated nucleus, somatic cell cloning of other large animal species (Cibelli et al., 1998; Baguisi et al., 1999; Galli et al., 2003; Woods et al., 2003; Jang et al., 2007; Oh et al., 2008; Yin et al., 2008), including pigs (Polejaeva et al., 2000), has occurred rapidly. While large animals had been previously cloned (Prather et al., 1987,
1989; Sims and First, 1994) using embryo-derived cells, cloning with differentiated somatic cells presented the opportunity to identify somatic cell types that can be produced in greater quantities and are easier to genetically modify. Making genetic modifications to the somatic cell before nuclear transfer has enabled the production of numerous transgenic pigs.

While pronuclear injection, ICSI-mediated transgenesis, sperm-mediated gene transfer and oocyte transduction are each capable of producing transgenic pigs, SCNT following genetic modification to the donor nuclei has become the preferred method for a variety of reasons. SCNT is advantageous over oocyte transduction, pronuclear injection and ICSI-mediated transgenesis as all offspring produced through SCNT are expected to be transgenic, and the modifications are always expected to demonstrate germline transmission. In addition, SCNT using clonal transgenic cell lines allows the investigator to preliminarily characterize the modification by addressing the number of integration sites, the transgene copy number and the chromosomal location before creating piglets. Importantly, because homologous recombination occurs at such a low frequency, the process of knocking in or out a gene has only been accomplished in pigs through SCNT using appropriately modified somatic cells.

**Exogenous DNA delivery into the potential donor cell**

Efficient generation of genetic modifications in donor cell genomes relies not only on the targeting/integration strategy but also on the efficient introduction of the exogenous DNA into the potential donor cell. To this end, a variety of methods have been quite effective, including lipid-based delivery (Hyun et al., 2003; Lee et al., 2005), viral delivery (Lai et al., 2002a; Rogers et al., 2008a) and electroporation (Arat et al., 2001; Ramsoondar et al., 2003; Watanabe et al., 2005; Ross et al., 2010). It has more recently been demonstrated that utilization of the Amaxa Nucleofection System was capable of introducing exogenous DNA into 79% of surviving cells as determined by eGFP expression, compared with 53% of surviving cells using electroporation (Nakayama et al., 2007). We were able to achieve similarly high transient transfection efficiency through electroporation by using repeated square-wave pulses. Using optimized electroporation conditions, more than 80% of stably transgenic colonies were PCR positive for the transgene when the copy number ratio of selectable marker to transgene was 1:1 (Ross et al., 2010). Because of this frequency, transgenesis via randomly integrated gene addition is relatively feasible.

**Genetic Modification of Somatic Cells for Nuclear Transfer**

Genetically modifying a somatic cell followed by its use for nuclear transfer has been effectively used to generate cloned transgenic pigs (Lai et al., 2006; Rogers et al., 2008a). The process from designing the construct to characterizing the transgenic offspring requires consideration of numerous factors, some of which are presented in Fig. 11.1, to produce the desired outcome. The primary methods of gene targeting (i.e. gene knockout) and gene addition (i.e. gene knock-in) are improving in terms of both efficiency and specificity. Gene-knockout animals were first demonstrated by modification of the genome using homologous recombination (Doetschman et al., 1987; Thomas and Capecchi, 1987). Because homologous recombination is such a rare event, it is not an effective strategy for genome manipulation during the initial stages of embryonic development – the point at which transgenic modification occurs following pronuclear injection or ICSI-mediated transgenesis. A major advantage of using mice for transgenesis by homologous recombination was the availability of ES cells that can grow indefinitely and are capable of being introduced into early-stage embryos and contributing to chimeric offspring. Significant efforts have been invested into the creation of ES cell lines in large domesticated animals, although these have been largely unsuccessful (Talbot and Blomberg, 2008).

Efficient generation of genetic modifications in donor cell genomes relies not only on the targeting/integration strategy but also on the efficient introduction of the exogenous DNA into the potential donor cell. To this end, a variety of methods have been quite effective, including lipid-based delivery (Hyun et al., 2003; Lee et al., 2005), viral delivery (Lai et al., 2002a; Rogers et al., 2008a) and electroporation (Arat et al., 2001; Ramsoondar et al., 2003; Watanabe et al., 2005; Ross et al., 2010). It has more recently been demonstrated that utilization of the Amaxa Nucleofection System was capable of introducing exogenous DNA into 79% of surviving cells as determined by eGFP expression, compared with 53% of surviving cells using electroporation (Nakayama et al., 2007). We were able to achieve similarly high transient transfection efficiency through electroporation by using repeated square-wave pulses. Using optimized electroporation conditions, more than 80% of stably transgenic colonies were PCR positive for the transgene when the copy number ratio of selectable marker to transgene was 1:1 (Ross et al., 2010). Because of this frequency, transgenesis via randomly integrated gene addition is relatively feasible.
Random integration

For successful production of transgenic pigs carrying a gene addition, exogenous DNA must be delivered to the nucleus and, following a double-stranded DNA break, be integrated into the host genome. While gene knockout in pigs has only occurred through somatic cell modification followed by SCNT, random integration of an exogenous transgene can occur through any of the above-discussed strategies for producing a transgenic pig. Transgene constructs used for gene additions are commonly represented by coupling the coding sequence of a gene to a specific promoter to produce a protein product expressed in a tissue-specific (Paleyanda et al., 1997) or a ubiquitous manner (Whitworth et al., 2009).

One of the greatest difficulties with respect to integrating a transgene into the genome of a mammal is producing sufficient offspring to
screen and identify those with the most appropriate phenotype. The reason for this is the multiplicity of events that can vary between each founder; the precise site into which the transgene integrates, the number of integration sites and the number of transgene copies in each integration site can all vary between founders. Most integration events in mammalian cells following exogenous DNA delivery via electroporation result in only a single integration site (Nakanishi et al., 2002). However, the number of transgene copies within the integration site and the ability of the transgene to be expressed from different loci vary considerably. An example of this is the recent production of a new miniature pig model of retinitis pigmentosa (Ross et al., 2009). Six male founder piglets were created by adding the P23H human rhodopsin gene into the pig genome. While each founder is PCR positive for the transgene, fluorescent in situ hybridization indicated that transgene integration occurred on a variety of chromosomes and that the progression of retinal degeneration as determined through electroretinography varied significantly between founders.

Targeted modifications

The established method of gene targeting by homologous recombination, established in mouse ES cells (Thomas and Capecchi, 1987; Doetschman et al., 1987), remains the primary method by which targeted genetic modifications in other mammals have been accomplished. A targeting vector possessing a desired trait mutation is introduced into somatic cells (e.g. fetal fibroblasts) followed by colony selection and screening to identify appropriately modified cell lines for SCNT. Most genomic modifications will be the result of random incorporation of the targeting vector into the genomic DNA (Iizumi et al., 2008). However, in rare events, the arms of the targeting vector match up with homologous chromosomal DNA, and the exogenous DNA (situated between the arms of homology in the targeting vector) will be inserted into the genome by homologous recombination (Smithies et al., 1985). Positive and negative selection facilitates the identification of transgenic somatic cells potentially having the desired targeting event.

Another potentially useful method of introducing targeted modifications are zinc finger nucleases (ZFNs). ZFNs introduce DNA double-strand breaks at specific DNA sequences and have the capacity to increase homologous recombination-mediated gene targeting frequency by nearly 1000-fold (Porteus and Carroll, 2005; Morton et al., 2006; Moehle et al., 2007). ZFNs comprise a non-specific FokI (restriction endonuclease) cleavage domain and three to four Cys$_2$His$_2$ zinc finger (ZF) DNA-binding domains (Tan et al., 2003). Gene disruption and repair (i.e. gene therapy) can be achieved with ZFNs in human and Caenorhabditis elegans somatic cells (Urnov et al., 2005; Morton et al., 2006). While ZFNs have significant potential in terms of target specificity, designing and delivering functional ZFNs can be quite challenging (Ramirez et al., 2008). Efforts are currently under way to design and test ZFNs to target genes in swine fibroblasts, with intentions to develop transgenic swine via SCNT (Whyte et al., 2008).

Other strategies, including the use of recombinant adeno-associated virus (AAV) vectors, were used to deliver genetic constructs that were effective for gene-specific targeting in pig fetal fibroblasts (Rogers et al., 2008a). Not only was the vector highly efficient in the production of cystic fibrosis transmembrane regulator (CFTR)-targeted fetal fibroblasts, but these authors also demonstrated that individual cell lines have very different targeting efficiencies. For example, out of six cell lines used for targeting the CFTR gene, the frequency of targeted clonal cell lines as a percentage of transgenic lines ranged from 0.07 to 10.93% (Rogers et al., 2008a).

Conditional/inducible genetic engineering

Incomplete understanding of all specific gene functions can result in embryonic lethality in developing embryos possessing a trait knockout, as has been demonstrated in the mouse (Tsumura et al., 2003). Overcoming the embryonic lethality of specific gene knockouts underscores the importance of developing targeting strategies that can be conditionally induced. Conditional mutation by site-specific recombination has been reviewed in great detail elsewhere (Gossen and Bujard, 2002; Glaser et al., 2005). Essentially, a
site-specific recombinase, commonly Cre, is used to conditionally control gene expression. Cre recombinase catalyses a DNA recombination event removing DNA between two *loxP* sites (34-bp recognition sites), each composed of two 13-bp inverted repeats flanking an 8-bp central element (Primrose and Twyman, 2006). Cre has utility in excising unwanted selectable markers, such as the neomycin resistance gene (*neo*) in transgenic constructs (Kaartinen and Nagy, 2001). The basic premise is to design targeting vectors with part of the preferred endogenous gene flanked on each side by *loxP* sites, also termed ‘floxed’ (Sauer, 1998). Insertion of the *loxP* sites into introns flanking an exon critical for protein function allows normal expression to continue until Cre recombinase is present, driven by temporal or tissue-specific promoters (Primrose and Twyman, 2006). Because Cre recombinase does not require accessory proteins or cofactors for its activity, it is suitable for use in mammalian cells (Glaser et al., 2005).

One example of this is the conditional knockout of Dicer, which codes for a critical enzyme in miRNA (micro RNA) biogenesis during oocyte development (Tang et al., 2007). This was accomplished by mating a transgenic mouse containing the floxed Dicer allele (Yi et al., 2006) with a mouse transgenic with the ZP3 promoter driving the expression of Cre recombinase (de Vries et al., 2000). Because ZP3 is an oocyte-specific promoter, the result was the ability to determine specific temporal and spatial impacts of miRNA biogenesis during oocyte growth. Modifications such as these will be essential in further expanding the pig as a research model for human health and physiology.

**Type of donor cell affects success of somatic cell nuclear transfer (SCNT)**

Despite significant progress in creating clones of large animals without the utilization of ES cells, the process is still inefficient. Because the efficiency largely depends on the extent of nuclear remodelling and reprogramming, numerous cell types have been tested for their ability to produce clones through SCNT. A variety of cells has been utilized for SCNT in pigs, such as pre-adipocytes (Tomii et al., 2005), salivary gland-derived progenitor cells (Kurome et al., 2008), fetal somatic stem cells (Hornen et al., 2007), skin-derived stem cells (Hao et al., 2009), adult fibroblasts (Beebe et al., 2007; Brunetti et al., 2008) and fetal fibroblasts (Park et al., 2001; Lai et al., 2002a,b; Hyun et al., 2003). Fetal fibroblasts have proven valuable for the production of transgenic pigs because they are easily collected and cultured, are capable of being genetically modified and possess the ability to produce live offspring. While fetal fibroblasts, like other somatic cells, are capable of producing live offspring, the efficiency at which nuclear remodelling and reprogramming are sufficient to produce live offspring remains low. It is hypothesized that many of the differences in developmental efficiency between donor cells following SCNT may be a result of genomic methylation differences (Bonk et al., 2007).

**Improving somatic cell remodelling and reprogramming**

During normal, *in vivo* reproduction, shortly after fertilization, both the male and female pronuclei undergo global demethylation, which contributes to the production of a totipotent zygote. The zygote then undergoes several rounds of holoblastic cleavage and the activation of the embryonic genome at around the four- to eight-cell stage of development in the pig. As the embryo continues to develop, the methylation patterns and chromatin structure (i.e. histone code) begin to differentiate between cells, enabling the regulation of their transcriptomes and proteomes to contribute to the production of multiple cell types by the blastocyst stage of development. During SCNT, the oocyte must remodel (i.e. change the structure of the chromatin) and reprogramme (i.e. alter the transcriptional profile) the somatic cell genome to enable the appropriate spatial and temporal gene expression required for successful embryonic and fetal development.

The mechanisms involved during SCNT are complex and poorly understood, and require cytoplast-mediated reprogramming of the donor nucleus from a differentiated state to a totipotent condition (Mitalipov et al., 2007). Remodelling and reprogramming of differentiated somatic nuclei into a totipotent embryonic state by SCNT are not efficient, and the
mechanism by which this remodelling occurs is not well understood. Many epigenetic modifications occurring during normal embryonic development are likely required by the somatic cell genome for SCNT to be successful.

Accruing evidence suggests that inefficiencies in epigenetic reprogramming in SCNT embryos are the result of errors in DNA methylation and chromatin remodelling, which may contribute to the poor survival of SCNT-derived embryos (Dean et al., 2001; Kang et al., 2001; Santos et al., 2003; Ohgane et al., 2004; Bonk et al., 2007). This is supported by the observation of molecular abnormalities in the placenta and tissues of live-cloned animals (Humpherys et al., 2002; Inoue et al., 2002; Jiang et al., 2008). One potential contributor to the epigenetic reprogramming of SCNT-derived embryos is histone modification following SCNT. The histone code is a cellular memory responsible for maintaining the identity of differentiated cells (Bird, 2002) and correlates with gene activation through its effect on chromatin accessibility (Th'ng, 2001). Treatment of pig SCNT embryos with trichostatin A (TSA), a histone deacetylase inhibitor (HDACi), resulted in higher in vitro embryonic development than in controls (Zhang et al., 2007; Li et al., 2008). A downside to TSA is that it can be teratogenic when the concentration is high or the exposure is long (Svensson et al., 1998). Scriptaid, a HDACi with low toxicity that enhances transcriptional activity and protein expression (Su et al., 2000), significantly enhanced the development of SCNT embryos to the blastocyst stage when outbred Landrace fetal fibroblast cells (FFCs) or NIH mini pig inbred FFCs were used as donors, compared with results from an untreated group (Zhao et al., 2009, 2010). Considering that the reprogramming of nuclei following nuclear transfer only happens during a limited time before zygote genome activation, the relaxation of chromatin structure by histone acetylation, which corresponds to a transcriptionally permissive state, might contribute to successful cloning.

In addition to promoting the efficiency and capacity of the oocyte to appropriately remodel and reprogramme the somatic cell genome, the identification of somatic cells that are capable of being genetically modified and have a pluripotent phenotype, such as ES cells in mice, may improve the efficiency of SCNT to make transgenic pigs. While true ES cells have been difficult to obtain in pigs, a renewed interest in the development of induced pluripotent stem cells has recently developed (Esteban et al., 2009; Ezashi et al., 2009; Roberts et al., 2009).

**Future of Transgenic Pigs**

The increased efficiency of SCNT, coupled with the rapidly increasing understanding of the biological functions of encoded gene products, will enable the production of transgenic pigs that can accurately address issues affecting pig production and the use of the pig as a biomedical model. While the technologies have advanced to the degree where the types of genetic manipulations are becoming unlimited, some potential areas that could be significantly affected by genetically modified pigs are discussed below.

**Reproductive efficiency**

While litter size in pigs has increased over the past few decades, pigs still offer significant potential to increase the number and size of piglets weaned. One strategy to improve milk production was the creation of transgenic sows over-expressing bovine alpha-lactalbumin in their milk (Bleck et al., 1998). Piglets raised by the transgenic sows over-expressing bovine alpha-lactalbumin demonstrated an increased milk production during the first 9 days of lactation and improved litter weight gain (Noble et al., 2002).

It was also thought that ovulation rate may be increased in transgenic pigs that express the human beta-cell leukaemia/lymphoma-2 (BCL2) gene in follicles by preventing or reducing apoptosis (Guthrie et al., 2005). Unfortunately, an improvement in reproductive function was not observed, but the study underscores the usefulness of expressing a transgene to better understand biological systems for affecting reproductive efficiency in pigs.

**Growth and development**

A variety of transgenic modifications have been made in pigs to enhance and/or study the mechanisms of growth and development. These include pigs transgenic for the mouse
metallothionein-1 promoter driving expression of human or bovine growth hormone, human growth hormone releasing factor or human insulin-like growth factor-I (Pursel et al., 1990, 1999).

As the world population and demand for animal protein continue to increase, the demand for animals capable of more efficiently converting cereal grains into lean tissue will increase. Discovery and understanding of the molecular controls of nutrient transport and utilization will enable the manipulation of genomes to enhance or diminish the activity of specific proteins that contribute to the efficient utilization of feed.

**Disease resistance**

Infectious diseases cause a significant negative economic impact on production agriculture. Either identification of animals that are naturally resistant to disease, as with brucellosis in cattle, and cloning those animals (Westhusin et al., 2007), or modifying the animal so that it is resistant to disease, as with prions (Golding et al., 2006), should result in a population that is more resistant to disease. Many infectious agents attach to cell-surface molecules and are then internalized and initiate infection of the cell. Modification of cell-surface molecules has already been achieved in pigs (Lai et al., 2002b; Kolber-Simonds et al., 2004), and this results in a change in the antigenicity of the cell surface (Kuwaki et al., 2005). Similar modification to other cell surface molecules may render the cells/pigs resistant to infection by a variety of infectious agents.

**Transgenic pigs for xenotransplantation**

Because pigs have anatomy similar to humans they represent potential models as a species amenable for the harvest of organs for xenotransplantation. In fact, this area of research has gained significant interest in the past decade after the ability to make genetic modifications to the pig cells used for SCNT. Pigs have long been considered a source of organs for xenotransplantation to humans. Their anatomy, physiology and genetics are similar to those of humans (Prather et al., 2008). An initial problem with xenotransplantation is the hyperacute rejection (HAR) that occurs following transplantation from pigs to primates, as a result of natural antibodies that are capable of initiating complement activation and eliciting a cascade of events that end in immediate rejection of the organ (Dalmasso et al., 1992).

A large number of genetic modifications have been made to make the pig organs less immunogenic to the primate. These include making modifications to eliminate HAR, such as adding complement modifiers (Langford et al., 1994; Diamond et al., 1996, 2001; Zhou et al., 2005), remodelling the cell surface carbohydrates (Bondioli et al., 2001; Miyagawa et al., 2001), preventing antithrombosis (Petersen et al., 2008) and knocking out alpha-1,3-galactosyltransferase (Dai et al., 2002; Lai et al., 2002b; Ramsoondar et al., 2003). Additional modifications are now focusing on post HAR (Tu et al., 2003; Klose et al., 2005).

As progress continues in the production of a transgenic pig for which the organs are capable of being utilized for transplantation into humans, other issues will also need to be addressed. One of these is the matter of porcine endogenous retroviruses (PERVs) (Dieckhoff et al., 2008). While their impact on xenotransplantation is unclear, certain types of PERVs have been documented to be transmissible across species in vitro (Patience et al., 1997; Wilson et al., 1998). Because of their multiplicity throughout the pig genome, their entire removal by “knocking out” each retrovirus through transgenesis or by developing a breeding strategy that is capable of producing pigs lacking all known retroviruses both represent daunting tasks. Instead, some efforts have focused on the development of short interfering RNA (siRNA) constructs whose expression can effectively knock down PERV expression in pigs (Ramsoondar et al., 2009).

**Transgenic pigs for biomedical models**

While a tremendous amount of basic biological information has been learned from transgenic mice, an incredible amount is also being gained from the pig (Lunney, 2007). In some cases,
modifications to the pig genome to model human diseases have been shown to produce a phenotype more similar to human conditions than other existing models (Rogers et al., 2008b). The larger size of the pig makes it more expensive to maintain as a research model relative to rodents; however, the size can be used advantageously for issues such as determining drug dosage in preclinical trials and developing therapeutic strategies requiring surgical intervention. Transgenic pigs will continue to be made to model human genetic diseases, such as muscular dystrophy, sickle cell anaemia and many others, including genetic anomalies that cause a predisposition to specific cancers such as BRCA1 and BRCA2 mutations in human breast cancer.

Conclusions

As basic scientific discovery exposes specific molecular mechanisms that can be controlled or influenced by the modification of specific genetic sequence(s), the opportunities to improve animal production agriculture and biomedicine using the pig become unlimited. While there is no doubt that scientists and producers will continue to improve pig production efficiency through better management practices that utilize novel tools in reproduction, the ability to make transgenic modifications to the pig has given us tremendous opportunities to better understand physiology and improve the health and well-being of both animals and humans.

References


Plate 1. (A) Segregation of coat colour among F$_2$ animals from an intercross between European wild boar and Large White domestic pigs. Two alleles are segregating at the Extension/MC1R locus: the wild-type allele (E$^+$) and the recessive allele for black spotting (E$^p$). Three alleles are segregating at the Dominant white/KIT locus: the recessive wild-type allele (i) and the alleles for Patch (P) and Dominant white colour (I). The genotypes of the depicted animals are, from left to right: 1: i/i, E$^p$/E$^p$; 2: i/i, E$^p$/E$^p$; 3: i/i, E$^+$/-; 4: P/i, E$^+$/-; 5: I/-, I/-.

Photo: Mats Gerentz, Swedish University of Agricultural Sciences. (B) Median-joining network tree of MC1R alleles in wild and domestic pigs from Europe and Asia. Alleles are represented by circles and a four-digit allele designation (see Table 3.3), while squares represent predicted intermediate forms that have not yet been found. Thin black lines perpendicular to lines connecting alleles represent synonymous changes, while thicker red lines represent non-synonymous changes. Colours inside circles and squares represent observed and predicted phenotypes, respectively; question marks inside two of the squares indicate that the associated phenotypes cannot be predicted because they are intermediate forms between alleles differing by two non-synonymous substitutions. The asterisks associated with the synonymous substitutions leading to alleles 0203 and 0104 indicate the only instance of an identical mutation at different locations in the tree. Based on Fang et al. (2009). Publication of this figure was generously supported by PIC (part of Genus plc), Hendersonville, Tennessee, USA.
Plate 2. One of the first comparative Zoo-FISH maps of the pig and human genomes. An index of colour patterns represents individual human chromosomes. The homologies between the human and porcine karyotypes are demarcated by these colour patterns on the pig chromosomes. Loci mapped to individual pig chromosomes at that time are arranged on the right side of each chromosome. Courtesy of Frönicke et al. (1996) (colour).
Plate 3. Porcine–human comparative map. The comparative map was based on human genome build 37 and porcine genome build 9, which represents 89% of the porcine genome. Colour codes refer to the 23 human chromosomes.
Plate 4. Chester White, male, USA. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 5. Duroc, female, USA. Photo: National Swine Registry, USA.
Plate 6. Hampshire, female, USA. Photo: National Swine Registry, USA.
Plate 7. Hereford, female, USA. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 8. Poland China, female, USA. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 9. Spotted, female, USA. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 10. Yorkshire, female, USA. Photo: National Swine Registry, USA.
Plate 12. Piau, female, Brazil. Photo: Simone Guimarães, Universidade Federal de Viçosa, Brazil.
Plate 13. Berkshire, male, England. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 15. Middle White, female, England. Photo: David Merrett, Creative Commons, Daventry, UK.
Plate 16. Tamworth, female, England. Photo: Mapes Livestock Photos, Milford Center, Ohio, USA.
Plate 17. Angeln Saddleback, female, Germany. Photo: Hedwig von Ebbel, Wikimedia, Germany.
Plate 19. Landrace, female, Denmark. Photo: National Swine Registry, USA.
Plate 20. Mangalitsa, female, Romania. Photo: Daniel Ciobanu, University of Nebraska, USA.
Plate 22. Da Min, male, China. Photo: Max Rothschild, Iowa State University, USA.
Plate 23. Fengjing, male, China. Photo: Max Rothschild, Iowa State University, USA.
Plate 24. Luchuan, female, China. Photo: Kenneth Stalder, Iowa State University, USA.
Plate 25. Meishan, female, China. Photo: Max Rothschild, Iowa State University, USA.
Plate 26. Mong Cai, male, Vietnam. Photo: Kenneth Stalder, Iowa State University, USA.
Plate 27. Tongcheng, female, China. Photo: San-Ping Xu and Bin Fan, Huazhong Agricultural University, China.
Developmental Genetics

Michael Dyck\textsuperscript{1} and Anatoly Ruvinsky\textsuperscript{2}

\textsuperscript{1}University of Alberta, Canada; \textsuperscript{2}University of New England, Australia

Introduction \hfill 264

Developmental Stages of the Pig Embryo \hfill 264
  Gametogenesis \hfill 264
  Fertilization and development within the zona pellucida \hfill 266
  Post-hatching development \hfill 269

Genetic Control of Pre-implantation Development \hfill 272
  Expression of maternal genes \hfill 272
  Activation of the embryonic genome \hfill 274
  Reprogramming, methylation pattern and genomic imprinting \hfill 275
  Gene expression during blastocyst formation, expansion and elongation \hfill 276
  Genetic control of gastrulation \hfill 278
  Establishment of axial identity \hfill 279
  Trophoblast gene expression \hfill 280

Genetic Regulation of Implantation, Maternal Recognition of Pregnancy and Placentation \hfill 281
  Implantation and placental development \hfill 281
  Maternal recognition of pregnancy \hfill 282
  Cytokines and implantation \hfill 282
  Insulin-like growth factors \hfill 283
  Angiogenesis \hfill 284
  Pregnancy-associated glycoproteins (PAGs) \hfill 284

Genetic Control of Post-implantation Development \hfill 285
  The three germ layers and their derivates \hfill 285
  Development of segment identity and \textit{HOX} genes \hfill 285
  Organogenesis \hfill 286
  Gene regulation of muscle development \hfill 288

Sex Differentiation \hfill 289
  The major steps in gonad differentiation \hfill 289
  The genes involved in sex differentiation \hfill 290
  Cycle of the X chromosome \hfill 292
  Anomalies in sex determination and differentiation \hfill 293

Summary \hfill 293
Note \hfill 294
References \hfill 294
Introduction

The study of mammalian development has historically relied on morphological observations and the evaluation of developing offspring in utero. This painstaking work has resulted in detailed and meticulous descriptions of the events and structures that arise during this elegantly orchestrated process (Patten, 1948). The developmental genetics of the pig, as with many other mammals, was hindered for a long time owing to the tremendous complexity of development and the lack of appropriate methods. However, recent advances in molecular techniques and gene expression profiling have resulted in the generation of a vast amount of data concerning the genetics of mammalian development. A great deal of this work has been carried out in the mouse, but there are increasing amounts of information on the embryology and genetics of development in economically important agricultural species such as the pig. Drawing on this information, this chapter is intended to provide an overview of the current understanding of mammalian development as it relates to pigs.

Developmental Stages of the Pig Embryo

Gametogenesis

As in any mammal, the beginning of development in the pig, as well as the origin of its genetic material, lies in the gametes that come together to ultimately create a porcine conceptus. As a result, the development and developmental competence of the conceptus can be affected by gametogenesis. The ovary of newly born females contains a lifetime supply of oocytes stored in quiescent primordial follicles (for a full review see Hunter, 2000; Picton et al., 2008; Edson et al., 2009). With oogenesis commencing in the female fetus, meiosis of oogonia begins as early as 40 days after conception and, about 35 days after birth, all oocytes are formed and arrested at the prophase of the first meiotic division (i.e. the germinal vesicle stage). At puberty, and continuing throughout the reproductive life of the female, under appropriate hormonal control, pools of primordial follicles are recruited to grow. In the pig, an estimated 500,000 primordial follicles are present at birth and this number decreases slightly to approximately 400,000 around puberty. Follicular development follows a well-coordinated series of events characterized by follicular and oocyte growth, as well as cell proliferation and differentiation. Morbeck et al. (1992) were able to determine the time needed for porcine follicular development from the initiation of primordial follicle growth to the pre-ovulatory stage; a primordial follicle is estimated to require approximately 84 days before reaching this stage (Fig. 12.1). Another 14 days is then necessary for an early antral follicle to develop to a 3-mm follicle. Finally, based on an estimated growth rate of 1 mm/day for follicles larger than 3 mm, a 3-mm follicle requires 5 days to reach pre-ovulatory status. Therefore, roughly 100 days are needed for a porcine primordial follicle to ovulate. Oocytes also grow throughout follicular development, starting at less than 30 mm in primordial follicles and reaching an average of 120 μm in the pre-ovulatory follicle. Oocyte growth is almost linear until the follicle reaches 300 μm in diameter, after which the size of the oocyte remains relatively constant until ovulation. Oocyte growth is also associated with meiotic and developmental competence, as well as with the ability of the oocyte to undergo meiotic maturation, fertilization and development. In vivo, a surge of luteinizing hormone (LH) initiates oocyte maturation and induces ovulation of the oocyte. A key component to oocyte maturation is maturation promoting factor (MPF) kinases. In the pig, MPF activity is very low in the germinal vesicle stage, and peaks at metaphase I and II stages (Dedieu et al., 1996). Other kinases, such as MOS and mitogen-activated protein (MAP), which are referred to as cytoplasmic factors (CFs), are also involved in the regulation of meiotic events. It has long been established that the cMos gene affects the block in the meiotic cell cycle at metaphase 2 in mammalian oocytes (Colledge et al., 1994). The Gs-coupled receptor, GPR3, has also been identified as an essential regulator of meiotic arrest in the mouse oocyte (Mehlmann et al., 2004). Gpr3 RNA has been localized in oocytes, and oocytes
from mice lacking the Gpr3 gene undergo spontaneous oocyte maturation within fully grown, intact follicles, independent of an increase in LH. Also, Andreu-Vieyra et al. (2006) provide a comprehensive overview of mouse models that demonstrate the importance of oocyte-expressed genes in fertility, including the genes for: growth differentiation factor 9 (Gdf9); bone morphogenetic protein 15 (Bmp15); zona pellucida glycoproteins (Zp1, Zp2, Zp3); helix-loop-helix transcription factor (Figla); connexins (Gja1, Gja2); and the homeobox gene (Nobox). Although maturing oocytes are both transcription and translation active, the presence of a specific mRNA in the oocyte does not ensure that the protein is actually expressed, particularly given that mammalian oocytes contain a large percentage of mRNA that is not polyadenylated and therefore not translated. As a result, proteomic analysis is favoured over gene expression in identifying molecules that are differentially expressed during oocyte maturation (Vitale et al., 2007).

Spermatogenesis is a cyclic and highly coordinated process that begins with spermatogonial stem cells, which, in contrast to the process of oogenesis, continually proliferate and replenish the testicular seminiferous tubules by mitotic division to produce a virtually unlimited number of gametes. The spermatogonia proceed through two meiotic divisions, which are followed by spermiogenesis, in which haploid spermatids develop into spermatozoa (Merchant-Larios and Moreno-Mendoza, 2001). All types of male germ cells are found in a single section of the seminiferous tubule epithelium. The cycle of the seminiferous epithelium refers to the complete progression through this series of cellular stages and is unique for each species, as is the duration of spermatogenesis. The entire progression of a seminiferous cycle in the pig is 8.3 days (Franca and Russell, 1998), while the complete process of spermatogenesis from spermatogonia A to a fully formed spermatozoa requires 39 days in the boar (Franca and Russell, 1998). An important event related to the meiotic
division of spermatogenic cells occurs during the pre-leptotene phase, when complete DNA replication occurs, forming tetrads without separation. The resulting chiasmata facilitate chromosomal recombination, so that the prophase of the first meiotic division during spermatogenesis is a major source of genetic heterogeneity. Owing to the complexity and duration of the spermatogenic process, the underlying genetic mechanisms are not fully understood, particularly in domestic animals such as the pig; however a review by Yan (2009) reveals 20 different genes that, when deleted, adversely affect male fertility.

During both oogenesis and spermatogenesis, epigenetic mechanisms that regulate gene transcription are also active in the form of DNA methylation. The primordial germ cells which are the starting point of gametogenesis in the early embryo have highly methylated DNA (Reik et al., 2001; Hajkova et al., 2002). However, once these cells have populated the developing gonads they are generally hypomethylated. In the pig, this demethylation process is complete by day 28 (D28) of development (Petkov et al., 2009). During gametogenesis there is a de novo methylation of the gametic genomes catalysed by DNA (cytosine-5)-methyltransferases (DNMTs) that regulate genes during development and play a role in genomic imprinting. Genomic methylation patterns are erased and reacquired differentially in the developing male and female gametes, are further modified in the early embryo and become relatively stable by late embryogenesis. Sex-specific DNA methylation of particular domain sequences occurs, forming the basis for paternally and maternally imprinted genes (discussed later). It is proposed that the erasure and resetting of DNA methylation that take place as a part of gametogenesis are likely to be important in preventing DNA methylation defects from being passed from one generation to the next.

**Fertilization and embryonic development within the zona pellucida**

Fertilization occurs within a few hours of ovulation in an inseminated sow, and the initial cleavage to a two-cell embryo within 12–16 h post ovulation. For this review, day 0 (D0) of development is considered to be the day that ovulation and fertilization occur, and all other temporal reference to developmental stages use this as the starting point. The porcine conceptus then develops through a series of critical stages prior to implantation. These steps and the timing of critical events in the prenatal development of the pig are summarized in Table 12.1. Following fertilization of an oocyte, the arrested meiotic process resumes, the second polar body is extruded into the perivitelline space, and then the male and female pronuclei form (Fig. 12.2a). The completion of meiosis is facilitated by various signal transduction pathways that converge to inactivate the MPF (Colledge et al., 1994). The MOS MAP kinase pathway is also considered to be responsible for the phosphorylation of spindlin (Oh et al., 1997), which is associated with the meiotic spindle and the changing of the metaphase spindle into an anaphase configuration that requires the presence of calcium/calmodulin-dependent protein kinase II (CaM) and has been characterized in the porcine oocyte. (Fan et al., 2003; Xu et al., 2009). Oocytes of cMos knockout mice undergo spontaneous activation, owing to a lack of spindlin phosphorylation, which destabilizes metaphase II arrest (Hashimoto et al., 1994). Fertilization also triggers waves of increased calcium concentration passing through the cytoplasm, which is thought to be mediated by the soluble sperm protein oscillin (Parrington et al., 1996). These calcium waves, also referred to as the calcium transient, lead to a remodelling of the cytoplasm and nuclear compartments, suggesting that they play a role in the initiation of transcription. Study of the calcium transient in pigs has generally been related to the refining of in vitro fertilization procedures (Funahashi et al., 1995; Ito et al., 2003), although it would appear that this is only one of the regulatory events that may proceed in waves through the cytoplasm at the time of fertilization. Additional factors are being discovered as promoting oocyte cytoplasmic and nuclear maturation, including the Sonic hedgehog signalling (Shh) pathway (Nguyen et al., 2009).

Gene expression studies in porcine gametes around the time of fertilization are limited,
but the presence of transcripts of the genes for clusterin (CLU), protamine 2 (PRM2), calfmegin (CLGN), cAMP-response element modulator protein (CREM), methyltransferase 1 (DNMT1), linker histone 1 (H1), protamine 1 (PRM1), TATA box-binding protein associated factor 1 (TAF1) and TATA box-binding protein (TBP) in porcine spermatozoa, mature oocytes, zygotes and two-cell stage embryos has been assessed (Kempisty et al., 2008). Transcripts for all of these genes were detected in spermatozoa, while oocytes contained only CREM, H1, TAF1 and TBP mRNAs. The zygote and two-cell stage embryos contained transcripts for the CLU, CREM, H1, PRM1, PRM2, TAF1 and TBP genes. This suggests that porcine spermatozoa may deliver CLU, PRM1 and PRM2 mRNAs to the oocyte, and that these probably contribute to zygotic and early embryonic development. By 48 h of development (D2) the porcine embryo is transformed from the two- to the four-cell stage (Fig. 12.2b and c) and, at the end of the third cell cycle, the transition from maternal to embryonic genome expression will

Table 12.1. Essential events and timing of pig prenatal development. Compiled from: Patten (1948); Evans and Sack (1973); Hyttel and Niemann (1990); Schultz (2002); Blomberg et al. (2005); Vejlsted et al. (2006b); Oestrup et al. (2009).

<table>
<thead>
<tr>
<th>Day of development*</th>
<th>Developmental stage/event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 0 (14–16 h post ovulation)*</td>
<td>Cleavage to two-cell stage</td>
</tr>
<tr>
<td>Day 1</td>
<td>Cleavage to four-cell stage</td>
</tr>
<tr>
<td>Day 2</td>
<td>Development to four-to-eight cell stage; genome activation</td>
</tr>
<tr>
<td>Days 2–3</td>
<td>Transition from oviduct to uterus</td>
</tr>
<tr>
<td>Day 3</td>
<td>Blastomere compaction; morula development</td>
</tr>
<tr>
<td>Day 5</td>
<td>Blastocoeal development</td>
</tr>
<tr>
<td>Day 6</td>
<td>Blastocyst expansion</td>
</tr>
<tr>
<td>Days 7–8</td>
<td>Blastocyst hatches from zona pellucida</td>
</tr>
<tr>
<td>Days 9–10</td>
<td>Conceptus is an expanded sphere; embryonic disc forms</td>
</tr>
<tr>
<td>Days 11–12</td>
<td>Pre-streak stage; maternal recognition of pregnancy; gastrulation begins; elongation of trophectoderm; primordial germ cell formation; beginning of implantation</td>
</tr>
<tr>
<td>Day 12</td>
<td>Primitive streak develops; gastrulation continues</td>
</tr>
<tr>
<td>Days 13–14</td>
<td>Open neural tube forms</td>
</tr>
<tr>
<td>Days 14–15</td>
<td>First somites apparent; neural tube developing and begins to close; optic vesicles apparent; attachment of conceptus to uterine endometrium; implantation is accomplished</td>
</tr>
<tr>
<td>Day 16</td>
<td>Neural tube closed; heart and auditory pits apparent; forelimb bud forming</td>
</tr>
<tr>
<td>Days 17–18</td>
<td>Amnion completely developed; optic vesicle present; hind limb bud forming</td>
</tr>
<tr>
<td>Day 19</td>
<td>Allantois fills and contacts the chorion; lens and optic cup well developed</td>
</tr>
<tr>
<td>Days 20–21</td>
<td>Olfactory pits present; eye pigments form; intestines herniated into umbilical cord</td>
</tr>
<tr>
<td>Day 22</td>
<td>End of somite formation</td>
</tr>
<tr>
<td>Days 24–26</td>
<td>Embryonic gonads visible</td>
</tr>
<tr>
<td>Day 28</td>
<td>Hair follicles appear; eyelids forming; external genitalia differentiate; digits forming</td>
</tr>
<tr>
<td>Day 30</td>
<td>Chorion fully vascularized by allantoic vessels</td>
</tr>
<tr>
<td>Days 34–35</td>
<td>Palate fused; facial clefts closed</td>
</tr>
<tr>
<td>Day 44</td>
<td>Prepuce, scrotum, labia and clitoris present</td>
</tr>
<tr>
<td>Days 46–49</td>
<td>Eyelids cover eyes</td>
</tr>
<tr>
<td>Days 112–116</td>
<td>Birth</td>
</tr>
</tbody>
</table>

*Day 0 (D0) of development is considered to be the day that ovulation and fertilization occur, and all other temporal references to developmental stages use this as the starting point.
Fig. 12.2. Early pre-implantation pig embryos: (a) fertilized zygote with two polar bodies and sperm on zona pellucida; (b) two-cell stage embryo; (c) four-cell stage embryo; (d) eight-cell stage embryo; (e) blastocyst in zona pellucida; (f) hatched blastocyst.
commence (Brevini et al., 2007). Subsequently, the embryo continues to cleave and initiates compaction as early as the 8- to 16-cell stage (Fig. 12.2d), and, by D2–3 of embryo development, this leads to formation of a compacted morula and transfer from the oviductal environment to that of the upper uterine horn. During the compaction process, blastomeres begin to flatten and cell-to-cell contacts become more pronounced; internal and external cells steadily differentiate and obtain some degree of polarity (Rossant and Tam, 2009). In the uterine environment, formation of the blastocyst, or blastulation, occurs. The blastocoeil or internal cavity develops next and, as a result, formation of the blastocyst occurs during D5 of development (Fig. 12.2e). Within the following 24 h, the blastocyst attains its typical structure and two distinct cell lineages: the inner cell mass (ICM) and trophoectoderm (TE). As a part of this process, the outer layers, closest to the zona pellucida, become connected by tight junctions and desmosomes to seal the expanding blastocyst cavity in which the ICM forms as a tight cluster of cells. The ICM cells communicate via gap junctions, facilitating the filling of the blastocoeil and, in turn, expansion of the embryo (D6), until it ultimately hatches from the zona pellucida by D7–8 (Fig. 12.2f).

**Post-hatching development**

At the time of hatching, the TE makes up the majority of the external cells of the blastocyst; it will develop an epithelial phenotype, form much of the extra-embryonic tissue, and play a critical role during implantation and formation of the trophoblast layers of the placenta. The ICM makes up the remaining cells, and differentiates into the epiblast and the hypoblast (primitive endoderm). At D9 of development, there is a disintegration of the TE polar region covering the epiblast to create the structure known as the embryonic disc. By D10 the conceptus is enlarged (8–10 mm), but still spherical, and the embryonic disc is fully formed (Fig. 12.3). The embryo may be referred to as being at the pre-streak 1 stage at this point (Vejlsted et al., 2006a). During D11–12, the embryonic disc transforms from a circle into an oval structure, and at one pole of the disc a prominent crescent-shaped thickening appears. This represents the pre-streak 2 stage and the first signs of anterior–posterior polarity of the embryo. At this point, a dramatic elongation of the conceptus occurs, in which it converts from a spherical to a long, filamentous structure, often over 1 m in length. The conceptus will also signal its presence during this period to allow for maternal recognition of pregnancy to occur in order that the corpora lutea (CL) are maintained and the uterine environment remains such that it will support and promote pregnancy. In the pig, blastocysts begin to produce oestrogens by D11 of development, which through a series of processes (discussed below) prevent secretion of the uterine luteolytic factor (PGF2 alpha) in an endocrine direction, while allowing secretion in an exocrine direction (i.e. into the uterine lumen), thereby protecting the CL from luteolysis or regression (Spencer et al., 2004). Then, by approximately D12–13 of development, the primitive streak appears at the posterior end of the embryonic disc (Fig. 12.4), corresponding to the onset of gastrulation.

Gastrulation tends to initiate in parallel with elongation of the TE (D11–12), and involves a complex sequence of cellular differentiation and movement that ultimately facilitates the generation of uniquely distinct structures and tissues within the conceptus. The major result is the formation of the three primary germ layers: endoderm, mesoderm and ectoderm. The resulting ectoderm will eventually give rise to the nervous system and epidermis, the mesoderm develops into the cardiovascular, urogenital and muscular systems, while the endoderm is the starting point for the digestive, pulmonary and endocrine systems. In addition to these somatic germ layers, the primordial germ cells are also formed (Flèchon et al., 2004a). The initiation of gastrulation as a developmental phase precedes neurulation, but its completion overlaps with this later process. The first sign of neurulation is a thickening of the anterior ectoderm as the primitive streak regresses and the formation of neural plate folds to become the neural groove (van Straaten et al., 2000). In the pig, the neural groove develops at approximately D14–15, which also coincides with the initiation of segmentation (Fig. 12.5).
Segmentation is the developmental process that subdivides the body into a series of subunits that result in the different segments of the body. In vertebrates, the earliest form of segmentation is the development of somites, which result from a thickening of the mesoderm in the midline of the embryo to form blocks of mesodermal cells. The first somites in the pig are apparent on D14 of development (Vejlsted et al., 2006b). In parallel to somite formation, the neural tube progressively forms and begins to close by the five-to-seven somite stage (D15) and is complete by the 28-somite stage (D16–17) (van Straaten et al., 2000). Soon after gastrulation in porcine embryos, the endoderm germ layer forms a primitive gut tube, which subsequently leads to organ specification (foregut, midgut and hindgut), then formation of organ buds, and finally to more specialized cell lineages (D20–22).

In the pig, fertilization of ovulated oocytes generally exceeds 95%, but 30–50% of these fertilized oocytes generally do not survive to parturition 112–116 days later (Geisert et al., 1982). It has been estimated

---

**Fig. 12.3.** (a) Porcine conceptus at D10 of development with embryonic disc clearly visible; (b) Schematic showing the loss polar TE (trophoectoderm) and emergence of the embryonic disc (ICM, inner cell mass); (c) Schematic showing stage of first mesenchyme cell migration. Figures (b) and (c) reproduced from Fléchon et al. (2004b), with the kind permission of the authors.
that 20–30% of porcine embryonic mortality occurs between D12 and D13, and another 10–15% occurs by mid-gestation (Pope et al., 1982; Stroband and Van der Lende, 1990). Another critical period is D20–30, as insufficient placental development during this stage influences subsequent fetal growth and survival in the uterine environment (Freking et al., 2007). Tayade et al. (2007) propose there to be three critical events for successful porcine development and survival in utero: (i) activation of the embryonic genome occurs at the four-to-eight cell stage (D2); (ii) blastocyst expansion and development of embryonic polarity and morphological differentiation; and (iii) attachment of the elongating conceptus to the endometrium by D15 and subsequent establishment of the placenta to sustain the developing fetus (Geisert et al., 1982). The remainder of the chapter will deal with the genetic control of these critical steps.
Genetic Control of Pre-implantation Development

Expression of maternal genes

As previously mentioned, unovulated porcine oocytes are arrested at the prophase of the first meiotic division, but both transcription and translation are very active and under ‘maternal command’. Mouse oocytes express about 5400 genes and transposable elements, many of which are conserved in chordates (Evsikov et al., 2006). Numerous newly synthesized mRNAs are stored and used later during oocyte maturation and up until embryonic genome activation (EGA), which occurs at the two-cell stage in the mouse (Hamatani et al., 2006) and the four-cell stage during pig development (Oestrup et al., 2009). Meanwhile, depletion of maternal mRNA intensifies prior to fertilization and continues until activation of the embryonic genome. By this time, nearly 90% of maternal mRNA is degraded and the majority of such transcripts are exclusively expressed from the oocyte genome (Bettegowda et al., 2008). In the mouse, and probably also other mammals, including the pig, ‘housekeeping’ genes are under-represented in the oocyte and early embryo transcriptomes. It has been suggested that this unique feature indicates that the core function of the oocyte is to act as a ‘reprogramming machine’
to create a totipotent embryo (Evsikov and Marin de Evsikova, 2009).

While current understanding of the transition towards mature oocyte and embryonic development is only emerging, and some species-specific deviations are possible, it is useful to develop a more comprehensive view. Several genes, some identified recently, guide this process (Fig. 12.6). Among such genes is Eif41b, which is involved in translational repression of maternal mRNAs. In the mouse, an oocyte-specific mammalian form of eukaryotic translation initiation factor 4E coded by the Eif41b gene may influence the speed of oocyte maturation (Evsikov et al., 2006). Another example is the inhibitory phosphorylation of the CDC2 protein with kinase activity, which is catalysed by pig Wee1B protein. This involves meiotic arrest of porcine oocytes. The inactivation of Wee1B gene, in combination with other factors, leads to the resumption of meiosis (Shimaoka et al., 2009). In mature oocytes, the degradation of maternal transcripts becomes more prominent, and seems to be nearly completed by the two-cell stage, when the so-called minor zygotic genome activation takes place. In fact, the ZAR1 gene (zygote arrest 1) is one of the few known oocyte-specific maternal-effect genes essential for the beginning of embryo development (Wu et al., 2003). Surprisingly, some Zar1(−/−) mice are viable and look normal. However, Zar1(−/−) females are infertile, probably as a result of the arrest of embryonic development in the majority of zygotes at the one-cell stage, and the fact that maternal and paternal genomes remain separate in such zygotes. These Zar1(−/−) embryos show a marked reduction in the synthesis of the transcription-requiring complex, with fewer than 20% of them progressing to the two-cell stage, and none develop to the four-cell stage (Wu et al., 2003). This gene is evolutionarily conserved, and the protein plays a role in transcription regulation during oocyte maturation and early post-fertilization development (Uzbekova et al., 2006). Several more maternal-effect genes, which were identified recently, are depicted in Fig. 12.6 and described in detail by Bettegowda et al. (2008).

Contrary to what is known for other mammals, in porcine zygotes there is an absence of the DNA replication checkpoint, meaning that the onset of DNA replication occurs very early in the maternal pronucleus, before activation of the paternal pronucleus (Vackova et al., 2006). The first divisions of the mammalian embryo are largely controlled by proteins and transcripts stored during oogenesis and oocyte maturation; porcine embryonic development is no different in that sense. Zygote genome activation (ZGA) in mice follows two stages: a minor one before cleavage and a major one at the two-cell stage and later (Hamatani et al., 2006). In the pig and other mammals, ZGA occurs slightly later, after the four-cell stage. Then the nucleoli, which are essential for

![Fig. 12.6. Current knowledge on genetic regulation of the oocyte-to-embryo transition in mammals relevant to porcine development. Redrawn from Bettegowda et al. (2008), with modifications compiled from several sources, including Magnani and Cabot (2008, 2009); Shimaoka et al. (2009); Evsikov and Marin de Evsikova (2009). Question marks indicate those genes for which activities have not, as yet, been confirmed in porcine embryos. EGA, embryonic genome activation.](image-url)
riboosomal RNA (rRNA) and ribosome production, develop in order to support protein synthesis. After fertilization, structures resembling the nucleolar remnant are established in the pronuclei; they are engaged in the re-establishment of fibrillo-granular nucleoli during the major activation of the embryonic genome (Maddox-Hyttel et al., 2007).

It is well known that in Drosophila melanogaster and Caenorhabditis elegans gradients of morphogens in the zygote and early embryo are crucial for establishing positional information (St Johnston and Nüsslein-Volhard, 1992; Nüsslein-Volhard, 1996). These gradients are essentially products of maternal gene expression. To what degree similar gradients and elements of the cytoskeleton are important during the earliest stage of mammalian development still is not clear. Increasing cell polarity was described at the eight-cell stage of mouse and rat development (Reeve, 1981; Gueth-Hallonet and Maro, 1992). Cell fate, controlled by positional information, seems reversible, and provides the developing embryo with a certain degree of flexibility. In cattle, cellular polarization occurred in some blastomeres at the 9- to 15-cell stage, but typical distinct polarity was not manifested until after the 16-cell stage, with approximately 40% polar cells per embryo (Koyama et al., 1994). Chimeric murine embryos, constructed from two-cell stage blastomeres from which the animal or the vegetal poles have been removed, can develop into normal fertile adult mice. Although polarity of the post-implantation embryo can be traced back to the eight-cell stage and, in turn, to the organization of the oocyte, its role is not entirely clear (Ciemerych et al., 2000). It seems that in mammals axis specification during oogenesis and through to the early stages of cleavage is under strong regulation. This is unlike what is observed in other metazoans, and may be related to viviparity (Evsikov and Marin de Evsikova, 2009). If so, then the gradients which are so important in insects and worms may not be crucial for the very early stages of mammalian development. The establishment of axial polarity during cleavage and blastocyst formation is considered later (Johnson, 2009).

### Activation of the embryonic genome

The first wave of EGA, the so-called minor activation, occurs primarily in a male pronucleus up until the two-cell stage and results in the synthesis of only a few specific polypeptides. The second wave, starting from the two-cell stage, leads to massive changes in the gene expression pattern. Data relevant to murine development show that the most significant activation of the embryonic genome takes place at the two-cell and particularly at the four-cell stages (Table 12.2). Many hundreds of other genes remain continually active during this period as well. In porcine embryos the transition to more intensive transcription is shifted towards the four-cell stage (Whitworth et al., 2005). The transcriptional dynamics of some porcine embryonic genes at very early stages provide useful information. For instance, the ZP3 and ZP4 genes coding for the major components of the mucoprotein layer of the zona pellucida have very high levels of expression in the germinal vesicle oocyte and progressively decline at the four-cell and the blastocyst stages. A similar pattern, slightly shifted towards the blastocyst stage, was observed for the DNMT genes responsible for the DNA methyltransferase involved in the regulation of transcription and genomic imprinting (Ko et al., 2005; Jeong et al., 2009). Another porcine gene coding for prothymosin alpha, which is involved in chromatin remodelling,

### Table 12.2. Dynamics of stage-specific gene activation during early mouse development. Data are extracted from Hamatani et al. (2006), Fig. 2.

<table>
<thead>
<tr>
<th>Stage</th>
<th>No. of genes specific to each stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfertilized Zygote</td>
<td>63</td>
</tr>
<tr>
<td>Two-cell stage</td>
<td>298</td>
</tr>
<tr>
<td>Four-cell stage</td>
<td>469</td>
</tr>
<tr>
<td>Eight-cell stage</td>
<td>13</td>
</tr>
<tr>
<td>Morula</td>
<td>4</td>
</tr>
<tr>
<td>Blastocyst</td>
<td>98</td>
</tr>
</tbody>
</table>
among other functions, peaks in activity at the four-cell stage and then steadily declines. Expression of the dihydrolipoamide dehydrogenase gene increases from the oocyte to four-cell and blastocyst stages, thereby reducing lipid and protein peroxidation. Published data also point out that the \textit{SMARCA2} and \textit{SMARCA4} genes, which are active at the early stages of development, play essential roles in controlling the expression of other genes during early mammalian embryogenesis (Magnani and Cabot, 2009). The expression of the actin coding gene (\textit{ACTB}) in porcine embryos increases over 20-fold starting from the two-cell stage (Whitworth et al., 2005).

Significant activation of transcription is an essential prerequisite for the intensification of translation that follows. \textit{RPL23}, one among 80 genes controlling ribosomal proteins, activates in the porcine embryo at the blastocyst stage before a major increase in translation (Whitworth et al., 2005). Bjerregaard et al. (2004) demonstrated in the pre-implantation porcine embryo the nucleolus-related gene expression of several proteins involved in rRNA transcription (upstream binding factor, \textit{UBF1}; topoisomerase I, \textit{TOP1}; RNA polymerase I, \textit{POLR1}; and the RNA Pol I-associated factor PAF53, \textit{POLR1E}) and processing (fibrillarin, \textit{FBL}; nucleophosmin, \textit{NPM1}; and nucleolin, \textit{NCL}). The first significant activation of the activity of these proteins was observed at the four-cell stage, and it then increased significantly at the blastocyst stage. Another marker that has been used to characterize genome activation is elongation initiation factor 1A mRNA (\textit{eIF1A}). Magnani et al. (2008) observed activation of \textit{eIF1A} at the two-cell stage in IVF (\textit{in vitro}-fertilized) embryos. As previously stated, activation of the embryonic genome occurs at the four-to-eight cell stage (\textit{D2}), while the dominant role of embryonic genes is established only after gastrulation (\textit{D11–12}) (de Vries et al., 2008).

### Reprogramming, methylation pattern and genomic imprinting

During the first 24 h or so after fertilization, the mammalian oocyte and sperm undergo natural reprogramming that gives rise to a totipotent zygote (de Vries et al., 2008). Genomic reprogramming is a complex process involving numerous mechanisms. Protein and mRNA molecules accumulated in the oocyte facilitate reprogramming through chromosome remodelling, as well as differential utilization and degradation of mRNA. As previously mentioned, DNA methylation is erased from chromatin very early during development, thus creating a critically important condition for the next cycle of life. In porcine embryos, demethylation continues up to the eight-cell stage, which is followed by \textit{de novo} methylation. Incomplete epigenetic reprogramming is common for nuclear-transfer generated embryos and contributes to the low efficiency of the cloning procedure (Dean et al., 2001).

The following developmental stages lead to the occurrence of pluripotent cell types with narrowed potential. Gene-expression programmes operating in these pluripotent cells steadily become more defined, the production of core transcription factors begins, and the expression of pluripotency-associated genes commences. At least three genes (\textit{OCT4}, \textit{NANOG} and \textit{SOX2}) coding for transcription factors have been identified. They are responsible for activation of other genes essential for the maintenance of pluripotency and the temporary repression of genes required for further differentiation (de Vries et al., 2008).

Genes that are required later in development are repressed by histone marks, which confer short-term, and therefore flexible, epigenetic silencing (Reik, 2007). As soon as demethylation is accomplished, a new wave of DNA methylation begins, and this leads to stable and long-term epigenetic silencing of certain genetic elements, such as transposons, imprinted genes and pluripotency-associated genes. Whether such DNA methylation epigenetic marks play a key role in determining cell and lineage commitment still remains an open question (Reik, 2007).

Porcine development shows all the typical features discovered in other mammalian species, including reprogramming and epigenetic modification. For instance, methylation of the lysine residue 27 of histone H3 (H3K27), which is an essential epigenetic marker linked to transcriptional repression, was confirmed during cleavage in porcine embryos. Three genes
were described (EED, EZH2 and SUZ12) that regulate the H3K27 methylation process starting from the four-cell stage. The data indicate that H3K27 trimethylation is an epigenetic marker of maternally derived chromatin which undergoes global remodelling (Park et al., 2009). The paternal genome in the pig is actively demethylated within several hours of fertilization, while the maternal genome, on the contrary, is demethylated passively by a replication-dependent mechanism after the two-cell embryo stage (Haaf, 2006). This type of demethylation is related to DNA repair processes (Reik, 2007).

Gametic or genomic imprinting is a developmental phenomenon typical for eutherian mammals and is based on the differential expression of maternal and paternal alleles in certain genes. These genes are essential for the regulation of embryonic and placental growth. In genes such as IGF2, only the paternal allele is expressed (maternal imprinting); in contrast, in genes such as H19, only the maternal allele is expressed (paternal imprinting). Imprint acquisition occurs before fertilization and imprint propagation extends up until the morula–blastocyst stage (Shemer et al., 1996). In H19, the 2-kb region is methylated on the paternal allele during spermatogenesis; the maternal allele has a different methylation pattern (Davis et al., 1999). In pigs, H19 is exclusively expressed from the maternal allele in all major organs, similarly to what has been observed in other species. On the contrary, the majority of IGF2 transcripts are expressed paternally from promoters 2–4 (Li et al., 2008).

The molecular mechanisms of gametic imprinting are still under investigation. It seems possible that primary gametic signals are not simply copied from the gametes, but rather that the methylation pattern typical of imprinted genes establishes gradually during early development (Shemer et al., 1996). The regulatory elements that control genomic imprinting have differential epigenetic marking in oogenesis and spermatogenesis, which results in the parental allele-specific expression of imprinted genes during development and after birth (Feil, 2009). Both DNA and histone methylation are essential for imprinting. The latest data also show that DNA methylation is involved in the acquisition and/or maintenance of histone methylation at imprinting control regions (Henckel et al., 2009).

The developmental function of gametic imprinting is also not absolutely clear, but an explanation proposed by Moore and Haig (1991) is widely accepted. This is based on the concept of genetic conflict arising during pregnancy between maternally and paternally inherited genes. Thus, it is likely that gametic imprinting evolved in mammals to regulate intrauterine growth and to increase the safety of embryonic development. Lack of maternally or paternally derived alleles, or the abnormal expression of such alleles in a zygote, may lead to embryonic mortality and impose strict requirements on the stability of imprinting signals. According to the database of imprinted genes (Imprinted Genes Catalogue, 2010), at the time of writing the number of such genes so far identified in the pig is 61. Imprinting was recently confirmed for DIRAS3, DLK1, H19, IGF2AS, NNAT, MEST, PEG10, PHLDA2, PLAGL1, SGCE and SNORD107 pig genes (Bischoff et al., 2009). The majority of imprinted genes in the pig (48) that have accumulated in the database were found in different quantitative trait loci (QTLs). While identification of these genes is still an objective for future studies, it is possible that information about QTL-related imprinting effects will sooner or later be adopted by selection programmes (de Vries et al., 1994).

### Gene expression during blastocyst formation, expansion and elongation

At the morula stage of development, activation of a few selected genes is critical for synthesis of morphogenetically important proteins such as actin and the actin-associated proteins like alpha-fodrin, vinculin, and E-cadherin (Reima et al., 1993). These molecules are distributed evenly in blastomeres during early cleavage, but then gradually accumulated towards the blastocyst stage in the regions of intercellular contact (Reima et al., 1993). This seems to be essential for development of tight gap junctions related to blastocoeel formation, and is particularly relevant to the outer cell layer.
As outlined, formation of the blastocyst, including the blastocoel, ICM and TE within the zona pellucida, occurs on D5–6 of development. Despite the simplicity of the blastocyst structure, the mechanisms of its formation are still elusive. Three models – mosaic, positional and polarization – have been suggested and extensively studied (Johnson, 2009). As previously discussed, the ICM differentiates into the epiblast and primitive endoderm. The epiblast gives rise to the embryo itself and also to some extra-embryonic tissues. The TE is responsible for the development of the remaining extra-embryonic tissues and plays a critical role during implantation and formation of the trophoblast layers of the placenta. It has been found that Cdx2, which encodes a caudal-related homeodomain protein, is a key regulator of the TE lineage (Rossant and Tam, 2009), even though expression of this gene begins earlier. Mouse gene Tead4, which produces transcription enhancer, is tentatively considered as an upstream factor relevant to Cdx2; Eomes, on the contrary, is a downstream-located factor (Rossant and Tam, 2009). The Cdx2 and Eomes proteins are restricted to the outer layer cells. These two genes, which are required for specifying the pluripotent cells of the ICM-like Oct4 (Nichols et al., 1998), Sox2 (Nichols et al., 1998; Avilion et al., 2003) and Nanog (Chambers et al., 2003; Mitsui et al., 2003), are initially expressed in all blastomeres but progressively become restricted to the ICM cells after blastocyst formation. Adjaye et al. (2005) identified in developing human blastocysts marker transcripts specific to the ICM (e.g. OCT4/POU5F1, NANOG, HMGB1 and DPPA5) and TE (e.g. CDX2, ATP1B3, SFN and IPL). The emergence of pluripotent ICM and TE cell lineages from the morula is controlled by the metabolic and signalling pathways, which include, inter alia, WNT, mitogen-activated protein kinase, transforming growth factor-beta, NOTCH, integrin-mediated cell adhesion, phosphatidylinositol 3-kinase and apoptosis.

Polarity of the cells in the blastocyst increases as a result of an accumulation of protein kinase3, polarity protein Par3 and ezrin in the apical domain of blastomeres and apical membrane; other proteins, such as Lg1 and Par1, are exclusively found in the basal portion of blastomeres (Rossant and Tam, 2009). Connexin proteins are differently expressed both temporally and spatially in the pig embryo, influencing the formation of gap junctions in the trophoblast and later controlling the exponential growth of the trophoblast in pre-implantation pig blastocysts (Fléchon et al., 2004a).

Although the rate of embryonic development differs in pigs and mice, there is a correlation between the developmental stage and cytoskeletal organization in both species. Likewise, in the expanded bovine blastocyst, the distribution of several cytoskeletal and cytoskeleton-related proteins appeared similar (Shehu et al., 1996). Extracellular fibronectin was first detected in the early blastocyst before differentiation of the primitive endoderm and, at this stage, was localized at the interface between the TE and the extra-embryonic endoderm (Shehu et al., 1996). Cingulin, the tight junction peripheral membrane protein, also contributes to morphological differentiation in early mouse development, and it is likely that other mammals have the same gene. The synthesis of cingulin is tissue-specific in blastocysts, and is up-regulated in the TE and down-regulated in the ICM (Javed et al., 1993).

It is commonly accepted that proto-oncogenes are involved in numerous processes of embryonic development and that they encode a series of nuclear transcription factors, intracellular signal transducers, growth factors and growth factor receptors. For example, activation of the c-fos and c-jun proto-oncogenes in sheep conceptuses occurs during the period of rapid growth and elongation (Wu, 1996), and a similar pattern possibly occurs in porcine embryos. These proto-oncogenes are involved in the regulation of gene expression, cell proliferation and differentiation.

Following hatching at D7–8, the blastocysts transform from a spherical to a tubular shape, the elongation process begins (Patten, 1948), ICM develops into the embryonic disc, and differentiation of the epiblast and the hypoblast starts. By D11–12 of porcine development, when elongation and transformation to a thin filamentous structure is accomplished, the pattern of gene expression becomes rather complex (Ross et al., 2003; Blomberg et al., 2008). Ross et al. (2003) identified 142 genes that were differentially...
expressed among spherical, tubular and filamentous conceptuses. It was also found that during the transition from tubular to early filamentous conceptuses, sadenosylhomocysteine hydrolase and heat shock cognate 70kDa expression were significantly enhanced. Conceptus expression of \textit{OCT4} on D11–12 was approximately 7.7-fold and 11.6-fold greater than expression on D15 and D17, respectively. This result suggests that down-regulation of \textit{OCT4} may be important during conceptus expansion, following implantation and gastrulation in the pig (Spencer et al., 2006). Comparisons between ovoid and filamentous conceptuses showed elevated expression of genes involved in steroidogenesis, such as cytochrome P-450sc (\textit{CYP11A1}), aromatase (\textit{CYP19A}) and steroidogenic acute regulatory protein (\textit{STAR}), and also oxidative stress response pathways (\textit{MGST1}) and copper-zinc superoxide dismutase (\textit{SOD1}) (Blomberg and Zuelke, 2005). The involvement of oestrogen and interleukin-1-beta was also confirmed (Blomberg et al., 2008).

Comparisons between the spherical and filamentous conceptuses (D12), as well as filamentous D12 versus filamentous D14, showed that expressions of 482 and 232 genes, respectively, were statistically different. The most significant differences were observed in genes coding for interferon-\(\gamma\) (\textit{IFNG}), heat shock protein 27kDa (\textit{HSPB1}), angiomotin, B-cell linker (\textit{BLNK}), chemokine ligand 14 (\textit{CXCL14}), parathyroid hormone-like hormone (\textit{PTHLH}) and maspin (Ross et al., 2009).

A comparison of bovine transcriptomes from the blastocyst (D7) and conceptus (D14) stages revealed that \(\sim 500\) genes were up-regulated between these developmental stages, and only 26 genes were down-regulated (Ushizawa et al., 2004).

**Genetic control of gastrulation**

On D11 of development, at the stage preceding gastrulation (pre-streak), there is a heightened rate of cell proliferation in the posterior section of the epiblast in the porcine conceptus. Migration of the cells that are proposed to be the precursors of the primitive streak and expression of the \textit{Brachyury} (\textit{T}) gene begin (Fléchon et al., 2004b). This gene belongs to the \textit{T}-box family of genes, which contribute greatly to tissue specification, morphogenesis and organogenesis (Müller and Herrmann, 1997). \textit{Brachyury} interacts with the \textit{Goosecoid} gene (\textit{GSC}), which codes for a homeboxoid protein, whose very low expression in porcine embryos can be detected as early as D7–8 of development. However, the intensity of expression becomes more pronounced by D12, particularly in differentiating mesodermal cells that ingress from the epiblast via Hensen’s node (van de Pavert et al., 2001). This process finally leads to formation of the mesoderm and embryonic endoderm. \textit{Goosecoid} overexpression may repress the \textit{Brahyury} gene and affect normal development (Boucher et al., 2000). In porcine embryos at the expanding hatched blastocyst stage, \textit{OCT4} is confined to the inner cell mass. Following separation of the hypoblast and formation of the embryonic disc, this marker of pluripotency was selectively observed in the epiblast. Progressive differentiation of germ layers and tissues leads to silencing of this gene, with exception of its expression in the primordial germ cells (Vejlsted et al., 2006a).

The migratory cells converge at the mid-line of the posterior part of the epiblast, which creates a thickened longitudinal band known as the primitive streak (Patten, 1948; van de Pavert et al., 2001). At the pre-streak stage, which precedes gastrulation and migration of the extra-embryonic mesoderm, the embryonic disc becomes polarized (Fléchon et al., 2004b). The early primitive streak is characterized by both high pseudostratified epithelium with an almost continuous but unusually thick basement membrane, and \textit{Brachyury} expression. \textit{Brachyury} is crucial for notochord development in all examined chordates, and at least 44 notochord-expressed genes are its transcriptional targets (Capellini et al., 2008). Expression of the \textit{NODAL} gene is essential for axial patterning during early mammalian gastrulation, as well as for induction of the dorso-anterior and ventral mesoderm (Jones et al., 1995). Interestingly, the round shape and gradual posterior displacement of the porcine anterior pre-gastrulation differentiation appears to be species specific, and correlates with development of the primitive streak and
extra-embryonic mesoderm (Hassoun et al., 2009). As gastrulation proceeds, the primitive streak extends anteriorly, and at its distal end the Hensen’s node is developed; this is composed of a mass of epithelium-like cells without cilia (Blum et al., 2007) (Fig. 12.4). Expression of the Goosecoid gene is typical for these cells, which are the origin of the notochord.

The notochord is a flexible fibro-cellular cord lying ventral to the developing central nervous system, and represents the major axial structure of the embryo, which plays an important role in the induction of the neural plate, chondrogenesis and somite formation. Glycoproteins compose a core of the notochord, with cells encased in a sheath of collagen fibres. Two genes controlling notochord formation encode laminin b1 and laminin g1, and are essential for building the scaffold on which individual cells organize the rod-like structure typical of the notochord (Parsons et al., 2002). Higher protein production of integrin subunits that regulate interactions with collagens and laminin is known for notochordal cells (Chen et al., 2006). In vertebrates, the notochord is replaced during development by the vertebrate column. The notochord grows anteriorly from the Hensen’s node below the embryonic disc, and is composed of cells derived from a certain kind of differentiating mesodermal cells that ingress from the epiblast. Three key porcine genes, SOX17, NODAL and Brachyury (T), are involved in early development of the axial structure during gastrulation (Hassoun et al., 2009). According to Zorn and Wells (2009) the Nodal signalling pathway is necessary and sufficient for initiation of endoderm and mesoderm development, and it is required for proper gastrulation and axial patterning. Nodal ligands are members of the TGFβ family of secreted growth factors. NOTO is another gene that is required for the formation of the-caudal part of the notochord, as well as for ciliogenesis in the posterior notochord. The data also show that Noto acts during murine development as a transcription factor upstream of Foxj1 and Rfx3. According to Beckers et al. (2007), this genetic cascade is important for the expression of the multiple proteins required for cilia formation and function. Later, these processes influence dorsal and ventral axis specification and neural tube and spinal cord patterning.

Clearly, the activation of the nuclear genes responsible for basic morphogenetic rearrangements is the prerequisite for notochord formation and development. The T gene, which was first described as the Brachyury mutation in mice 80 years ago, is an important participant of the events required for the differentiation of the notochord and the formation of the mesoderm during posterior development. The T protein is located in the cell nuclei and acts as a tissue-specific transcription factor (Kispert et al., 1995). Cloning and sequencing of the T gene led to the discovery of the T-box gene family, which is characterized by a conserved sequence called T-box (Bollag et al., 1994). This ancient family of transcription factors, which underwent duplication around 400 million years ago, is common to all vertebrates (Ruvinsky and Silver, 1997). There are indications that several murine T-box genes are essential for the formation of different mesodermal cell subpopulations, and one of the T-box genes is essential for the development of early endoderm occurring during gastrulation (Papaioannou, 1997). Involvement of the T-box genes Tbx2–Tbx5 in vertebrate limb specification and development has also been demonstrated (Gibson-Brown et al., 1998). Formation of the notochord leads to several key ontogenetic events, including induction of the neural tube and then the central nervous system. A putative morphogen, Shh, secreted by the floor plate and notochord, specifies the fate of multiple cell types in the ventral aspect of the vertebrate nervous system and motoneurons. Shh, in turn, induces expression of the oncoprotein Gli1, which affects later development of the dorsal midbrain and hindbrain (Hynes et al., 1997). Expression of the SHH gene is also important for establishment of the ventral pole of the embryonic dorsal–ventral axis (Echelard et al., 1993). Unlike the notochord cells, other emerging mesodermal cells spread out more or less uniformly and give rise to numerous organs and structures.

Establishment of axial identity

The most recent review on the development of axial polarity indicates difficulties in establishing the initial causes of polarity in the
very early stages of mammalian development (Johnson, 2009). It is obvious, however, that the early blastocyst, and possibly even the 16-cell stage, have some degree of polarization and emergent axial identity. Several genes that significantly contribute to the emerging polarity have been identified so far. Genes encoding ezrin, the PAR family proteins and CDX2 are probably the key regulators. Other proteins, such as CDC42, E-cadherin, β-catenin and Hippo, are strongly involved in the process, and laminin and integrins play some role (Johnson, 2009). Development of the primitive streak and the notochord is the convincing demonstration that both anterior–posterior (AP) and dorsal–ventral (DV) axes are strictly determined. The left–right (LR) axis may look like an automatic consequence from the earlier-defined AP and DV axes, as it is perpendicular to both (Levin, 2004). While the AP and DV axes can be influenced by gravity or sperm entry point, there are no known factors that differentiate left from right. If this is so, it may mean that the LR symmetry existing in mammals is effectively caused by the AP and DV axes. However, the cause of LR asymmetry in vertebrates, and in mammals in particular, is a difficult question. Levin (2004) compiled a long list of genes which may affect the symmetry. More recent findings show that in the developing mouse embryo, leftward fluid flow on the ventral side of Hensen’s node determines LR asymmetry. Morphological analyses of the node cilia demonstrated that the cilia stand not perpendicular to the node surface, but tilted posteriorly (Nonaka et al., 2005). This morphological asymmetry can produce leftward flow. A genetic cause of LR asymmetries of the internal organs in vertebrates is steadily becoming clearer. Gros et al. (2009) considered two possibilities. The first is that initial asymmetric cell rearrangements in chick embryos create a leftward movement of cells around the Hensen’s node; this is relevant to expression of Shh and Fgf8 (fibroblast growth factor 8). The alternative is a passive effect of cell movements. It has also been shown that a Nodal-BMP signalling cascade drives LR heart morphogenesis by regulating the speed and direction of cardiomyocyte movement (Medeiros de Campos-Baptista et al., 2008).

**Trophoblast gene expression**

In the pig, the TE makes up the majority of the external cells of the blastocyst, and this portion of the conceptus expands dramatically and elongates along the villous folds of the uterus. The TE, therefore, provides the functional point of fetal and maternal contact, which is known as the trophoblast. Differentiation of the trophoblast begins early in embryonic development and ultimately results in functionally diverse cells. Again, much of the work on trophoblast gene expression has been done in the murine model, whose placental tissues and placentation are quite different from those of the pig. Roberts et al. (2004) describe this process, and consider that a key first step in trophoblast differentiation is the down-regulation of OCT4, which normally acts as a negative regulator of genes required for further differentiation (de Vries et al., 2008). OCT4 acts in the pluripotent ICM to silence genes related to differentiation but, once this restraint is removed, these genes, discussed below, can come under the control of transcriptional activators. Knofler et al. (2001) provide a comprehensive overview of the key regulatory factors involved in trophoblast development and differentiation. Of these, the T-box gene Eomes, which is considered to be one of the earliest trophoblast-determining factors in the pre-implantation embryo, seems to be required for trophoblast differentiation (Russ et al., 2000). Both Eomes and the homeodomain protein CDX2 are absent in the ICM, but are present in the TE (Beck et al., 1995). Cdx2 and Eomes murine knockout embryos fail to implant, and only develop to the blastocyst stage (Chawengsaksophak et al., 1997; Russ et al., 2000).

A family of transcription factors of basic helix-loop-helix (bHLH) proteins have been shown to be important in trophoblast development. In the mouse, this includes *Mash-2* (*Ascl2*), whose expression is crucial in the specification of the trophoblast lineage, and particularly in spongiotrophoblast development. The transcription factor encoded by the *Hxt* gene is also expressed in the trophoblast, and is considered
to have a positive role in promoting the formation of trophoblast giant cells in the mouse (Cross et al., 1995). This family of factors also includes Hand1, which is important for trophoblast giant cell formation in the mouse. Mice lacking the Hand1 gene show defects in the development of these cells (Riley et al., 1998); Hand1 expression may also be related to the regulation of Mash-2 (Ascl2) (Scott et al., 2000). Although ruminants do develop giant binucleate cells in their placentas, structures equivalent to spongiotrophoblast or trophoblast giant cells do not appear to exist in the pig.

There are various transcription factors that are widely expressed in embryonic, fetal and adult tissues that, based on knockout studies, seem to be necessary for placental development, as their deletion is consistently associated with trophoblast abnormalities. These include ETS2 (Yamamoto et al., 1998), AP1 (Schorpp-Kistner et al., 1999; Schreiber et al., 2000) and AP2gamma (Auman et al., 2002; Werling and Schorle, 2002). Schultz et al. (1997) also described the genetic determination of integrin trafficking, which regulates adhesion to fibronectin during differentiation of the mouse peri-implantation blastocyst. In addition, the regulation of several metalloproteinases and their corresponding genes may also shed additional light on the process of implantation and further trophoblast development (Bass et al., 1997). As the trophoblast forms and matures, it eventually produces a series of products and hormones, including, but not limited to, growth factors, interferons and pregnancy-associated glycoproteins, which will be discussed later. Given that placental development in rodents is dramatically different from that of the pig, it is important to determine whether these murine trophoblast genes are relevant to porcine trophoblast development.

Genetic Regulation of Implantation, Maternal Recognition of Pregnancy and Placentation

Implantation and placental development

As with other ungulates, the porcine conceptus remains unattached in the uterine lumen for a relatively extended period, and is considered to associate with the uterine endometrium as of D11–12 of development, with attachment being complete by D15 (Geisert et al., 1982). During this initial period of free movement, the developing embryos are uniformly distributed in the two uterine horns. Also, the embryo produces significant amounts of oestrogens and interferons during this pre-implantation period. Unlike mice or humans, the porcine placenta is not highly invasive and relies on the elongated conceptus to establish maximal surface contact with the uterine endometrium. Formation of the extra-embryonic membranes is an obligatory step in establishing the ability of the conceptus to attach to and interact with the uterus. The membranes that originate from the primary germ layers include the yolk sac, chorion (serosa), amnion and allantois. The yolk sac is formed from the ICM, which evaginates to create a cavity in the developing embryonic gut, which becomes the sac. This transient structure is visible shortly after gastrulation and regresses in size as the conceptus develops. The amnion and chorion both form the primitive endoderm and mesoderm. The amnion is a fluid-filled membrane that surrounds the developing embryo, while the chorion is the outermost extra-embryonic membrane, which will ultimately interact with the uterine endometrium. The first indication of the amnion is in the form of the amnionic folds, which are evident shortly after the primitive-streak stage; it then quickly develops into a fluid-filled membrane that encases the developing embryo, while the chorion is the outermost extra-embryonic membrane, which will ultimately interact with the uterine endometrium. As the amnion is developing, the allantois emerges as a sac-like evagination from the primitive gut. As the embryo grows, the allantois fills and eventually contacts the chorion, usually by D19 in the pig. Continued and increasing contact between the chorion and allantois, as well as a period of rapid angiogenesis, results in the fusion of these two membranes and infiltration of the chorion by allantoic vessels. By D30, the chorion is extensively vascularized by allantoic vessels, resulting in the formation of the allantochorion membrane. At this point the allantochorion does not truly implant, but interacts with the endometrial glands. Therefore, the porcine placenta is classified as a diffuse epithelial-chorial type with numerous folds that
interlock into the endometrial folds, and six layers of tissue between the maternal and fetal tissues. As previously mentioned, this D20–30 period is critical in pigs, as insufficient placental development during this period influences subsequent fetal growth and survival in the uterine environment. The placenta continues to grow until D60–70, after which its growth rate decreases (Freking et al., 2007). By D85, the placenta folds deepen and become more complex and, once the placenta is fully established, both fetal and maternal mitogenic and morphogenic factors continue to support development until parturition.

Maternal recognition of pregnancy

The mechanisms involved in the maternal recognition of pregnancy in pigs were first described by Bazer and Thatcher (1977). The underlying process in this event is the production of PGF2 alpha by the uterine endometrium; this is luteolytic and will, therefore, cause regression of the progesterone-producing CL. In order to maintain the CL and the progesterone levels necessary to support pregnancy, the conceptus will produce oestrogens that are anti-luteolytic. Therefore, it is believed that the oestrogens produced by the conceptus between D11 and D12 provide the signal for maternal recognition of pregnancy in swine. A second period of oestrogen production also occurs between D15 and D30, and is considered important for CL maintenance. As mentioned above, developmental expression of the key genes involved in steroidogenesis have been studied in the pig, and the filamentous conceptus shows elevated expression of CYP11A1, CYP19A, STAR, MGST1 and SOD1 (Blomberg and Zuelke, 2005). These oestrogens influence the expression of receptors for maternal hormones (e.g. prolactin) in the uterine tissues which, in turn, alter the secretion of PGF2 (Young et al., 1989, 1990). Although it is well established that oestrogen secretion is essential for the maternal recognition of pregnancy in pigs, other secretory proteins from the conceptus and endometrium are important for the appropriate development and survival. For a comprehensive review of conceptus–uterus interactions in the pig see Johnson et al. (2009).

Cytokines and implantation

As described in a recent review by Bazer et al. (2009), there are numerous strategies for implantation in mammals. Uterine receptivity to implantation involves the expression of interferon (IFN)-stimulated genes (ISGs), and these genes have many roles. A comprehensive list of ISGs is available from the Interferon Stimulated Gene Database (http://www.lerner.ccf.org/labs/williams/xchip-html.cgi). IFNs are pro-inflammatory cytokines secreted in the uterus during early pregnancy. They are cell-signalling proteins and, in most cases, the action of IFNs on ISGs is preceded by induction of the genes by progesterone, which is elevated during pregnancy. This process has been well described in several species (Bazer et al., 2008). In the pig, the conceptus is able to secrete two major types of IFN: Type I IFN delta (SPI-IFN) and Type II IFN gamma (IFNG). Between D13 and D20 of development, SPI-IFN and IFNG transcripts are detectable (Cencic and La Bonnardiere, 2002; Joyce et al., 2007a). In ruminants, IFN tau (IFNT) is well established as the factor expressed by the conceptus that acts as the signal for recognition of pregnancy, and therefore impedes regression of the CL until the end of pregnancy. In the pig conceptus, IFNs do not show anti-luteolytic effects that alter the length of the oestrus cycle or the concentration of circulating progesterone (Cencic and La Bonnardiere, 2002; Cencic et al., 2003). However, beyond the initial priming effects of progesterone, interactions of IFNs and oestrogen regulate cell-specific gene expression of various genes in the endometrium, and orchestrate the interactions between the conceptus and the uterine environment. Several genes are induced by oestrogen in the uterine luminal epithelium by D12, including IFN regulatory factor 2 (IRF2). IRF2 is a regulator of transcription for various IFN-influenced genes that are likely to play roles in the establishment of pregnancy, including cellular proliferation, attachment and development of the conceptus (Ka et al., 2001;
Hicks et al., 2003; White et al., 2005; Joyce et al., 2007a). In pigs and sheep, IRF2 expression may restrict expression of most IFN-stimulated genes to the endometrial stroma and glandular epithelium. By D15, IFNs up-regulate a large array of genes in the uterine stroma and glandular epithelium, including: IFN-stimulated gene 15 (ISG15); the interferon-related developmental regulator 1 gene (IRF1); the signal transducer and activator of transcription gene (STAT1); the swine leucocyte antigen genes (B1PJV4 and Q9TSW4); and the beta 2 microglobulin gene (B2MG) (Hicks et al., 2003; Joyce et al., 2007a,b, 2008). The role of these pregnancy-specific IFN-stimulated uterine genes is unclear, but it could involve: (i) protecting the conceptus from immune rejection; (ii) limiting the ability of the conceptus to invade the endometrium; and (iii) stimulating uterine/placental angiogenesis.

There are various other cytokines whose mRNA expression has been detected in conceptus and placental tissues, including: interleukin-6 (IL-6) (Mathialagan et al., 1992), interleukin-1beta (IL-1B) (Tuo et al., 1996), colony stimulating factor-I (CSF-I) (Tuo et al., 1995) and leukaemia inhibitory factor (LIF) (Anegon et al., 1994). More recently, the expression of a family of small cytokines, referred to as chemokines, has been detected in the reproductive tissues of pregnant mice, humans and pigs (Girard et al., 1999; Townsend and Nibbs, 2002; Martinez de la Torre et al., 2007; Wessels et al., 2007). Although various functions for these cytokines have been proposed, their role in mice and humans may not be the same in the pig, in which the type of placentaion and degree of invasiveness of implantation are quite different. Therefore, the relevance of these pro-inflammatory cytokines to the success or failure of pregnancy in the mouse and human may be very different from that in the pig (Croy et al., 2009).

**Insulin-like growth factors**

Insulin-like growth factors (IGFs) are pleiotropic growth factors that are required for uterine and conceptus growth and development (Simmen et al., 1990). Insulin-like growth factors 1 and 2 (IGF1 and IGF2) are small polypeptides that promote cellular differentiation, proliferation and migration, and inhibit apoptosis. These processes are involved in the extensive remodelling that occurs during the development of the placenta and its endometrial attachment site. The IGFs bind with high affinity to IGF receptors, namely IGF1R and IGF2R. IGF1R is a member of the tyrosine kinase family and is structurally related to the insulin receptor (Jones and Clemmons, 1995). IGF1R binds with equal affinity to both IGF1 and IGF2, whereas IGF2R binds only IGF2 with high affinity (Pollak, 2008). The bioavailability and biological actions of IGFs are regulated by at least six insulin-like growth factor binding proteins (IGFBPs) (Clemmons, 1997), which may augment or inhibit IGF action, and several proteases cleave IGFBPs, reducing or eliminating their ability to bind IGFs.

The IGF system has been extensively studied during the porcine oestrus cycle and in early pregnancy (Simmen et al., 1992; Ashworth et al., 2005). IGF1 mRNA decreases while IGF2 mRNA increases as pregnancy advances, with the highest level found in the placenta, followed by the endometrium and myometrium (Simmen et al., 1992). There is a spatio-temporal relationship between increased uterine IGF1 and IGF2 and oestrogen synthesis in the elongating conceptuses at D10–13 (Letcher et al., 1989; Geisert et al., 2001). Increased uterine IGF is associated with the expression of CYP11A1 and CYP19A by the conceptus, and therefore by oestrogen synthesis (Green et al., 1995). Ashworth et al. (2005) documented premature loss of IGFs during the period of conceptus elongation following early exposure of pregnant gilts to oestrogen. The loss of IGFs during this period was attributed to proteolysis of IGFBPs due to endocrine disruption (Ashworth et al., 2005). Recent research has been conducted to address whether IGF1 and IGF2 are directly linked with porcine conceptus development at D20 and D50 (Croy et al., 2009) in which expression of the IGF1 and IGF2 genes was assessed. At D50, trophoblasts had more IGF transcripts than D50 endometrial samples, while, in D20 endometrial samples, IGF1 transcripts was more abundant than in D50 endometrium. These data are consistent with other reports
that IGF1 expression declines as pregnancy advances (Simmen et al., 1992; Ashworth et al., 2005). IGF2 transcripts were tenfold higher in both endometrial and trophoblast samples than IGF1 transcripts. These results are in agreement with previous reports that IGF2 is predominantly expressed over IGF1 during porcine pregnancy (Ashworth et al., 2005). This is also consistent with a series of gene knockout studies in mice that clearly show that IGF2 rather than IGF1 plays important roles in mouse placental and fetal development (Baker et al., 1993; Liu et al., 1993). Given that IGF2 and its receptor gene (IGF2R) are both imprinted genes, it has been shown that only disruption of the paternal IGF2 and/or the maternal IGF2R genes will affect embryonic growth and development (DeChiara et al., 1990; Filson et al., 1993).

Angiogenesis

Angiogenesis, the process by which new blood vessels are generated from an existing vascular system, occurs extensively during pregnancy to support the conceptus. A variety of factors support angiogenesis, but vascular endothelial growth factor (VEGF) and its receptors (VEGFR1, VEGFR2, VEGFR3) appear to be of primary importance (Loges et al., 2009). Of these receptors, VEGFR2 is dominant, and placenta growth factor (PGF) is a high-affinity ligand for VEGFR1 (Carmeliet et al., 2001). Angiogenic gene profiles have been studied for VEGF, PCF, VEGFRI and VEGFRII in porcine endometrial tissues, and in the trophoblast associated with the healthy conceptus and with conceptuses experiencing arrested development (Tayade et al., 2006; Linton et al., 2008). At D20 and D50 of development, fewer VEGF transcripts were detected in the endometrial tissues associated with the arrested conceptuses compared with the healthy ones, but the amount of VEGF transcripts in the trophoblast were not different. Also, from D15 to D28 of porcine development, there is dramatic onset of angiogenic activity that coincides with elevated numbers of a unique lymphocyte type referred to as uterine natural killer cells (uNK) (Engelhardt et al., 2002). When endometrial lymphocytes from these same D20 and D50 conceptuses were screened for angiogenic gene expression, the endometrial lymphocytes were found to have a greater abundance of the VEGF transcripts than the endometrial endothelium or the trophoblasts. However, for the conceptuses demonstrating arrested development, their attachment sites show severely reduced VEGF expression and an increase in PGF expression by the lymphocytes. The uterine lymphocytes preferentially expressed VEGFRI, while the trophoblasts were abundant in VEGFII transcripts, indicating that mechanisms that regulate angiogenesis differ between the maternal and embryonic/fetal compartments (Tayade et al., 2007). This molecular evaluation of the porcine conceptus attachment sites shows a clear role for immune cells in the acceleration of angiogenesis in these tissues.

Pregnancy-associated glycoproteins (PAGs)

The pregnancy-associated glycoproteins (PAGs) belong to a group of aspartic proteinases that include pepsins (gastric enzymes) and cathepsins D and E (lysosomal enzymes) (Davies, 1990). The production of PAGs and PAG-like factors has been identified in various species during pregnancy, and their expression consistently initiates at the time of implantation and continues in the TE as pregnancy proceeds (Szafranska et al., 1995). In ruminants, the expression of several bovine PAG (BPAG) transcripts has been identified in binuclear cells, while other distinct BPAG transcripts are expressed throughout the TE (Green et al., 2000). Porcine PAGs have been localized in the chorionic tissues (Majewska et al., 2006). To date, at least 66 distinct PAG complementary DNAs (cDNAs) have been cloned from various species (Majewska et al., 2006), but only eight pig PAG genes (PPAG) have been identified (PPAG1–6, PPAG8 and PPAG10) (Szafranska and Panasiewicz, 2002; Panasiewicz et al., 2004; Szafranska et al., 2006). The PPAGs are classified into two subfamilies: PPAG1-like (PPAG1, 3 and 5) and PPAG2-like (PPAG2, 4, 6, 8 and 10), based on
their nucleotide and amino acid sequence homology/identity. Majewska et al. (2009) have recently reported the chromosomal assignment of the PPAG gene family to the long arm of porcine chromosome 1. Although the definitive role of PAGs has yet to be clarified, Roberts et al. (1996) suggested that they may interfere with maternal immune function and thereby protect the conceptus. In vitro studies have revealed a potential role of the PAG family as chorionic signalling ligands that interact with gonadotrophin receptors in cyclic pigs and cows (Szafranska et al., 2007), as well as during the early pregnancy of the pig (Panasiewicz et al., 2007). However, the overall role of PAGs is still under investigation in several species. These secretory proteins are easily detectable in maternal blood circulation and are used for pregnancy diagnosis (see Szafranska et al., 2006).

**Genetic Control of Post-implantation Development**

**The three germ layers and their derivates**

As mentioned earlier, by the end of gastrulation, three germ layers are established: endoderm, mesoderm and ectoderm. Embryonic and postnatal derivates of each of these layers were described and generally understood many decades ago (Patten, 1948). Molecular genetic mechanisms driving this highly complex combination of processes began to emerge relatively recently. Zorn and Wells (2009) published one of the first reviews covering the entire process of endoderm development and organ formation. Here we can highlight only the major regulatory systems influencing the variety of genes and processes involved in endoderm morphogenesis and the formation of certain organs. The Nodal signalling pathway is necessary and sufficient to initiate ectoderm and mesoderm development, and is itself influenced by the canonic WNT/β-catenin pathway (Zorn and Wells, 2009). High-level Nodal signalling supports endoderm development and lower activity specifies mesoderm identity. The activity of the Nodal pathway is controlled by an auto-regulatory loop. Several genes involved in the pathway, such as Nodal, xnr1 and sqt, have conserved Foxh1 DNA-binding sites in their first introns, sustaining the high activity essential for endoderm development. In contrast, in developing ectoderm, a negative feedback of Nodal activity is caused via the transcriptional target Lefty (Shen, 2007). Soon after gastrulation, the endoderm germ layer forms a primitive gut tube, which leads to organ specification, then to the formation of organ buds and finally to more specialized cell lineages. In some cases, as in hepatocyte and β-cell differentiation, the whole cascade of events and the genes involved were uncovered (Zorn and Wells, 2009).

Developmental events in mesoderm and ectoderm progress simultaneously but independently, with significant interactions between derivates from the germ layers. As is well known, many organs have cellular components originating from different germ layers. Certain genes play a key role in the earliest stages of germ layer development. For instance the Eed gene, initially identified in mice, is critical for embryonic ectoderm development (Sharan et al., 1991), as deletion of this gene prevents formation of ectoderm. A highly homologous gene was found in humans, and there are few doubts that a porcine homologue will be identified. The murine Brachyury (T) gene is crucial for mesoderm development. Mice homozygous for mutant alleles of the T gene do not generate enough mesoderm, and show severe disruption in morphogenesis of mesoderm-derived structures, in particular of the notochord (Wilkinson et al., 1990). One of the T-box genes, Tbx6, in mice is implicated in development of paraxial mesoderm (Chapman et al., 1996). Tbx6 transcripts are first detected in the gastrulation-stage embryo in the primitive streak and in newly recruited paraxial mesoderm.

**Development of segment identity and HOX genes**

Segmentation, observed in different groups of animals and particularly in vertebrates, has deep evolutionary roots. Segments with a common origin remain relatively separate during
development to create an opportunity for diversification and specialization. This evolutionary-developmental strategy has been commonly used for the creation of morphological structures or groups of cells with distinct features. For instance, the development of two major structures, the ectodermal neural tube and the paraxial mesoderm, depends on segmentation: the first is critical for development of the hindbrain, the head process and the spinal cord; the second is essential in generation of somites, which give rise to the axial skeleton and skeletal muscles. In developing porcine embryos, visible features of segmentation are apparent at D14 (Patten, 1948). The first somites can be seen in the middle of the closing neural groove, and their number increases anteriorly. The genetic and cellular processes driving segmentation depend on the expression patterns of \( \text{HOX} \) genes (Alexander et al., 2009).

The homeotic genes, which encode helix-loop-helix transcription factors, were first described in \( \text{Drosophila} \) as the primary determinants of segment identity. They all contain a similar 180-bp DNA sequence motif named the homeobox. Comparative analysis of the \( \text{Drosophila} \) homeotic gene complex, called HOM-C, and the mammalian homeobox genes, called the \( \text{HOX} \) complex, demonstrates a striking case of evolutionary conservation. The \( \text{HOX} \) gene family determines a set of transcription factors crucial for the development of axial identity in a wide range of animal species (Maconochie et al., 1996). Figure 12.7 shows the striking similarity and collinearity found in the molecular anatomy of the insect and mammalian \( \text{HOX} \) complexes. The main difference is the number of complexes per genome. In insects there is only one, while mammals and other higher vertebrates have four separate chromosome clusters (Alexander et al., 2009). There are 39 \( \text{HOX} \) genes in mammalian genomes, which belong to 13 paralogous groups.

The \( \text{HOX} \) genes are expressed in segmental fashion in the developing somites and central nervous system, and each \( \text{HOX} \) gene acts from a particular anterior limit in a posterior direction. The anterior and posterior limits are distinct for different \( \text{Hox} \) genes (Fig. 12.7). A hallmark of \( \text{HOX} \) genes is the correlation between their linear arrangement along the chromosome and the timing and AP limits of expression during development (Alexander et al., 2009). \( \text{HOX} \) genes determine AP positional identity within the paraxial and lateral mesoderm, neurectoderm, neural crest and endoderm.

Thus, the vertebrate body plan is, at least partially, a result of the interactions of \( \text{HOX} \) genes that provide cells with the essential positional and functional information. Signals from the \( \text{HOX} \) genes force embryonic cells to migrate to the appropriate destination and generate certain structures. Major signalling pathways such as the fibroblast growth factor (Fgf), Wnt and retinoic acid (RA) pathways play important roles in affecting the expression of different \( \text{HOX} \) genes in different developmental conditions. The expression of RA and its protein-binding ability, as well as its other functions in development of the porcine conceptus have been described (Yelich et al., 1997). RA can affect the expression of \( \text{HOX} \) genes and there is a 5′ to 3′ gradient in responsiveness of the genes to retinoids (Marshall et al., 1996). RA acts via its receptors, which comprise two families, RAR and RXR, which are members of the ligand-activated nuclear receptor superfamily. The receptors interact to form complexes that, in turn, regulate target gene binding to retinoic acid response elements (RAREs). These RAREs are found in the 5′ regulatory regions of the murine \( \text{Hox} \) genes. \( \text{HOX} \) genes have profound influence on the whole array of developmental process and on the establishment of segment identity in a variety of Metazoa.

**Organogenesis**

Hundreds, if not thousands, of genes are involved in organogenesis. Here we briefly describe only a few of them, which have numerous effects on the formation of organs during development. The T-box family of transcription factor genes is a good example of such massive involvement in organogenesis owing to their contribution to embryonic cell fate decisions, the control of extra-embryonic structures, embryonic patterning and numerous aspects of organogenesis (Naiche et al., 2005). Some of the T-box genes are involved
Fig. 12.7. (a) Alignment of the four mouse (listed as vertebrate) Hox gene complexes with the HOM-C (ANT-C) gene complex from Drosophila. The vertical shaded boxes indicate related genes. The 13 paralogous groups are noted at the bottom of the alignment. The collinear properties of the Hox complexes with respect to timing of expression, antero-posterior (AP) level and retinoic acid (RA) response are also noted at the bottom. From Maconochie et al. (1996), with the authors’ permission.

(b) Summary of HOM-C and Hox-2 expression patterns. The upper part of the figure is a diagram of a 10-h Drosophila embryo with projections of the expression patterns of the different genes from the HOM-C gene complex to particular body segments. The lower part of the figure is a diagram of an 12-day mouse embryo with projections of expression patterns of different genes from the Hox-2 complex to particular body segments. From McGinnis and Krumlauf (1992), with the authors’ permission. int, intercalary; mx, maxillary; la, labial. Changes in genetic nomenclature made after the initial publication of this figure were not introduced here.
in limb morphogenesis and specification of forelimb/hindlimb identity. It has been shown that *Tbx5* and *Tbx4* expression is primarily restricted to the developing forelimb and hindlimb buds, respectively. These two genes appear to have been divergently selected in vertebrate evolution to play a role in the differential specification of fore- (pectoral) versus hind- (pelvic) limb identity (Gibson-Brown et al., 1998). Mutations in the human *TBX3* gene cause the ulnar–mammary syndrome, which is characterized by limb deficiencies or duplications, mammary gland dysfunction and genital abnormalities. It was suggested that *TBX3* and *TBX5* evolved from a common ancestral gene and each has acquired specific yet complementary roles in patterning the mammalian upper limb (Bamshad et al., 1997).

At least seven T-box genes, including *Tbx1, Tbx2, Tbx3, Tbx5, Tbx18* and *Tbx20*, are involved in heart development. Each T-box gene has a unique expression profile in specific heart regions (Naiche et al., 2005). The study of transcriptional activation and repression of several T-box genes in three-dimensional space provides unique knowledge about the cardiac progenitor fields, as well as about the development of heart form, function and pathology (Stennard and Harvey, 2005).

**Pax** genes are another family of developmental genes coding nuclear transcription factors. They contain the paired domain, a conserved amino acid motif with DNA-binding activity. **Pax** genes regulate the development of organs and structures such as the kidney, eye, ear, nose, limb muscles, vertebral column and brain. Vertebrate **Pax** genes are involved in pattern formation, possibly by determining the time and place of organ initiation or morphogenesis (Dahl et al., 1997). **Pax1**, for instance, is a mediator of notochord signals during the dorsoventral specification of vertebræ (Koseki et al., 1993). The **Pax3** gene may mediate activation of **MyoD** and **Myf5**, the myogenic regulatory factors, in response to muscle-inducing signals from either axial tissues or overlying ectoderm, and may act as a regulator of somitic myogenesis (Maroto et al., 1997). The **Pax2** gene is involved in optic nerve formation and **Pax6** is considered as a master gene for eye development as well as for the development of some other ectodermic tissues. Mutations in **Pax6** result in eye malformation, known as aniridia in humans and small eye syndrome in mice (Dahl et al., 1997). The eyes absent gene (**Eya2**), which is involved in eye development in several metazoan phyla, seems to be relevant to pig development. Like the **Pax6** gene family, **Eya2** was probably recruited for visual system formation some considerable time after its original function was established (Duncan et al., 1997). Several other genes, such as **Bmp4, Msx1** and **Msx2**, which encode bone morphogenetic proteins and are expressed before and after neural tube closure, interact with **Pax2** and **Pax3** (Monsoro-Burq et al., 1996). The expression of **Pax7** in the cells harvested from porcine embryos confirmed that the gene is involved in the development of myogenic satellite cells (Singh et al., 2007). **Pax8** is essential for the regulation of thyroid function (Kang et al., 2001).

**Gene regulation of muscle development**

Obviously, information on the development of muscle tissue in the pig has great practical importance. The genetic mechanisms of muscle development in mammals are complex (Firulli and Olson, 1997). Skeletal, cardiac and smooth muscle cells express overlapping sets of muscle-specific genes, although some genes are only expressed in one particular muscle type. So-called genetic modules or independent **cis**-regulatory regions are required to direct the complete developmental pattern of expression of individual muscle-specific genes within each muscle type, and the tempo-spatial specificity of these regulatory modules is established by unique combinations of transcription factors (Firulli and Olson, 1997).

The musculoskeletal system of the trunk and tail develops from the paraxial pre-somatic mesoderm (PSM) cells (Tam and Beddington, 1987). Once these cells reach a specified position, gene expression changes significantly and a segmentation process begins. In mice, new somites appear approximately every 2h, and they are separated from the anterior PSM (Dunty et al., 2008). The bHLH transcription factor encoded by the **Mesp2** gene and controlled by the Notch
signalling pathway is essential in the segmentation programme (Saga et al., 1997). Ripply2 is another identified gene involved in segment boundary regulation and this is also under the influence of Notch pathway genes, as well as the mesodermal transcription factors T and Tbx6 (Oginuma et al., 2008).

Somitogenesis is probably controlled by a segmentation clock, which consists of molecular oscillators in the Wnt3a, Fgf8 and Notch pathways (Pourquie, 2003; Aulehla and Herrmann, 2004; Rida et al., 2004). Alternatively directed gradients of fibroblast growth factor 8 (Fgf8) and/or Wnt3a and retinoic acid (RA) establish a boundary front in the anterior PSM. Dunty et al. (2008) demonstrated that the canonical Wnt3a/β-catenin pathway is necessary for molecular oscillations in all three signalling pathways, but does not function as an integral component of the oscillator. On the contrary, Notch pathway genes continue to oscillate in the presence of stabilized β-catenin and thus drive periodic expression of the target genes lunatic fringe (Lfng) and Hes7 (Bessho et al., 2001; Morimoto et al., 2005). Further investigations are necessary in order to reach a deeper understanding of this sophisticated dynamic system.

It is well established that mitogens inhibit the differentiation of skeletal muscle cells, but the IGFs, acting through a single receptor, stimulate both the proliferation and differentiation of myoblasts. For example, an inhibitor of mitogen-activated protein (MAP) kinase inhibits IGF-stimulated proliferation of L6A1 myoblasts and associated events, such as phosphorylation of the MAP kinases and elevation of c-fos mRNA and cyclin D protein. This inhibitor caused a dramatic enhancement of differentiation, evident at both a morphological and biochemical level. In sharp contrast, an inhibitor of phosphatidylinositol 3-kinase and p70 (S6k) completely abolished IGF stimulation of L6A1 differentiation. These data demonstrate that the MAP kinase pathway plays a primary role in the mitogenic response and is inhibitory to the myogenic response in L6A1 myoblasts, while activation of the phosphatidylinositol 3-kinase/p70 (S6k) pathway is essential for IGF-stimulated differentiation. Thus, it appears that signalling from the IGF-1 receptor utilizes two distinct pathways that lead either to proliferation or the differentiation of muscle cells (Coolican et al., 1997). IGF-I and II, their binding proteins and members of the transforming growth factor (TGF) superfamily (myostatin and TGFβ1) are key regulators of the proliferation and differentiation of porcine myogenic cells. Changes in the relative expression of IGF and TGFβ play a considerable role in regulating myogenesis in porcine embryonic myogenic cell (PEMC) cultures (Xi et al., 2007).

Members of the B-cell translocation gene family with anti-proliferative properties (BTG1, BTG2 and BTG3) are involved in muscle development and growth in pigs. The BTG2 gene is expressed at high levels in skeletal muscle and heart. The expression of BTG2 and BTG3 was significantly different in skeletal muscle among different developmental stages and between two studied breeds. Both genes were induced in differentiated C2C12 cells, suggesting their involvement in myogenic differentiation (Feng et al., 2007). Selection for greater and leaner muscle mass in pigs picked up a number of mutations affecting muscle development and functioning (see Chapter 15). This includes RYR, RN, MU and some other genes.

**Sex Differentiation**

**The major steps in gonad differentiation**

The earliest steps of gonadal development in mammals commence at a similar period in XX and XY embryos. Primordial germ cells, which differentiate relatively late in mammals, migrate into the gonad area of either presumptive sex indiscriminately, and may function even across species barriers (McLaren 1998, 1999). Anderson (2000) described sexual development in pig embryos in sufficient detail. To be functional, a gonad needs both germ cells and somatic cells. It seems that gonadal development in the pig does not significantly deviate from that in other mammals. A few dozen germ cells, originating from the proximal region of the embryonic ectoderm, start their journey inside the embryo, along with the invaginating hindgut. In the pig, primordial germ cells migrate from the dorsal mesentery...
of the hindgut (D18) to the primordium of the gonad (D23), and enter the genital ridge by D26. The number of these cells in the porcine embryo increases from D18 to D26 (Takagi et al., 1997). The first differences between male and female porcine embryos start to emerge at D26. Y chromosome-carrying embryos develop testicular cords and interstitium, and begin testosterone production, while undifferentiated gonadal blastema can be seen in XX embryos. Further events in males follow without delay; Sertoli cells can be observed at D27, production of anti-Müllerian hormone (AMH) starts at D29, and production of 3β-HSD (hydroxysteroid dehydrogenase) begins at D30–35, leading to the appearance of Leydig cells. Müllerian ducts disappear and Wolffian ducts are transformed into epididymes, vas deferens, seminal vesicles, the prostate and other structures. From approximately D100 of fetal development until D20 of postnatal development, germ cells show a constant rate of doubling (Anderson, 2000).

In female embryos, egg nests develop by D30–35, and meiosis resumes at D40. As previously mentioned, at the time of birth the population of germ cells is approximately 400,000. The pre-meiotic diplotene porcine oocyte appears by D50, and this process continues until about D20 after birth. In female embryos, Müllerian ducts are transformed into oviducts, uterus, cervix and the upper parts of the vagina, while Wolffian ducts disappear (Anderson, 2000). Primordial follicles are first observed in ovaries at D68 and primary follicles at about D75; secondary follicles appear near the time of birth. Starting from about D70 post coitum until D90 after birth, primordial follicles account for about 80% of the total follicular population (Oxender et al., 1979).

In mice, expression of the Bmp4 gene (bone morphogenetic protein 4) in the TE layer, which is in the closest contact with the epiblast, is responsible for the differentiation of both the primordial germ cells and the allantois (Lawson et al., 1999). If a similar mechanism operates in the pig, then BMP4 protein would be also produced, and thus precedes cellular migration. Owing to ongoing proliferation, a significant number of germ cells reach the genital ridge, which consists of a thin layer of mesenchymal cells located between the coelomic epithelium and the mesonephros. Two genes, Sf1 and Wt1, are particularly important in the development of the murine genital ridge (McLaren, 1998).

Eventually, four different cell lines comprise the genital ridge: primordial germ cells, somatic steroidogenic cells, supporting cells and connective tissue. The fate of each lineage depends on the sexual determination of the embryo in which they develop, and the patterns of genetic activity are quite different in the testes and the ovaries. The porcine DDX4 gene is a homologue to the vasa gene identified in other mammalian species (Lee et al., 2005). In adult tissues, DDX4 transcription was restricted only to the ovary and testis. In porcine fetuses, the transcript was found in all stages, except for D17–18. The DDX4 protein was observed in proliferating primordial germ cells but not in embryonic germ cells.

The genes involved in sex differentiation

Gonad development is a key element in establishing mammalian sex. The chromosomal constitution determines the migration of cells into the gonads and the final differentiation into a testis or an ovary (Hunter, 1995). It is well known that sex determination in pigs, as well as in many other mammalian species, depends on the presence or absence of the Y chromosome. Embryos without the Y chromosome develop as females, and those that carry this chromosome develop as males. The breakthrough in molecular understanding of sex determination and differentiation in the mouse and human (Goodfellow and Lovell-Badge, 1993) paved the way for that in other mammals, including the pig.

The study of porcine SRY revealed that the open reading frame (ORF) of the gene consists of 624 bp. The protein has a centrally located high mobility group (HMG) box domain as well as amino terminal and carboxy terminal regions. Structurally, porcine SRY is more similar to bovine and human SRY than to mouse Sry. The porcine SRY gene is expressed within the cells of the genital ridge of the developing male.
pig embryo between D21 and D26 of gestation. At this time the primitive gonads are still bipotential; however, after D31, testis determination is evident (Daneau et al., 1996). Comparative analysis of the gene order on the Y chromosome between pig, human and mouse showed a number of differences. There is greater order conservation with the murine Y than with the human Y chromosome. The porcine SRY is located on the p arm of the Y chromosome close to the centromere (Quilter et al., 2002).

In genetic males, porcine genital ridge cells express not only the Y chromosome-located SRY and ZFY genes and the X chromosome-located DAX1 (or NR0B1, which encodes a nuclear receptor protein) and DFX genes, but also the autosomal AMH, WT1, SOX2 and SOX9, SF1 (steroidogenic factor-1) and several other genes (Lahbib-Mansais et al., 1997; Parma et al., 1997). The SF1 protein transactivates the porcine SRY promoter (Pilon et al., 2003).

Figure 12.8 shows a simplified genetic model describing the major events in sex determination pathways. The activation of SRY in normal XY embryos shifts the balance in favour of testis development and male pathways through up-regulation of the SOX9 gene and signalling of FGF9 (fibroblast growth factor 9), which promotes the secretion of prostaglandin D₂ (PGD₂) (Nef and Vassilli, 2009). FGF9 and PGD₂ form a positive feedback loop and intensify SOX9 expression, thus directing the differentiation of supporting cells to Sertoli cells, which in turn down-regulate female signals such as WNT4 and FOXL2 (Fookhead transcription factor; Lamba et al., 2009). Thus, testicular development in mammals is triggered by SRY. In genetic males, this factor induces
the differentiation of Sertoli cells (reviewed by McLaren, 1991) and the secretion of AMH. AMH, which belongs to the transforming growth factor β family, causes regression of Müllerian ducts and promotes the development of Wolffian ducts and the differentiation of Leydig cells, which secrete the male steroid hormone, testosterone (Behringer, 1995). Testosterone binds to androgen receptors, which act as transcription factors.

In the absence of an SRY gene, which is typical for XX embryos, the gonads develop into ovaries. Two independent signalling pathways, the R-spondin1/WNT4/β-catenin pathway and FOXL2 transcription factor pathway, support this developmental sequence of events. R-spondin1 has been recently recognized as a key female-determining factor (Nef and Vassili, 2009). In females, no Leydig cells are formed, no testosterone is produced and gonad development steadily moves towards the female phenotype. Quite often the female developmental programme is considered as the ‘default’, while the male programme requires ‘switching on’ of the SRY gene, followed by activation of other genes. The comparison of four regulatory regions located upstream of SRY shows high conservation between the human, bovine, pig and goat regions. These regions of homology share transcription factor-binding sites that appear to be subject to strong evolutionary pressure for conservation and may, therefore, be important for the correct regulation of SRY (Ross et al., 2008). In contrast, the structure of the SRY region in the mouse is significantly different.

**Cycle of the X chromosome**

As proposed by Lyon (1961), and now uniformly accepted, one of the X chromosomes in eutherian females undergoes inactivation during early embryonic development. Numerous investigations shed light on different aspects of X chromosome behaviour, including preferential inactivation of the paternal X chromosome in the trophoblast, random inactivation in the inner cell mass and molecular mechanisms of inactivation (Goto and Monk, 1998).

These scenarios appear to be completely relevant to the cycle of the X chromosome in the pig; however, the supporting information is still lacking. Preferential inactivation of paternal X chromosome in porcine XX embryos probably occurs in trophoblastic cells around D7.5, and then random inactivation follows in the embryonic disc cells around D11.5. Females become natural mosaics, with one X chromosome randomly inactivated in each somatic cell. In post-meiotic oocytes, the X chromosome is active, as in other mammalian species. The paternal X chromosome, on the contrary, enters the zygote being inactive, but, soon after fertilization, it reactivates. In XX embryos, both X chromosomes are expected to be active until trophoblast differentiation. Then only one X chromosome remains active, regardless of the number of X chromosomes in a cell. This is an essential condition for gene dose compensation. The mechanisms of silencing the inactive X chromosome are complex. Several chromatin modifications are necessary in order to form stable facultative chromatin capable of propagating through numerous cell divisions. The so-called X-inactivation centre located on the X chromosome contains the Xist gene and cis regulatory genetic elements. The Xist gene encodes an RNA molecule that plays a key role in silencing the inactive X chromosome (Plath et al., 2002). Xist is negatively regulated by its antisense transcript, Tsix. It seems, however, that Tsix (the reverse spelling of Xist) is not the only regulator, and additional transcription factors are involved in this complex process (Senner and Brockdorff, 2009). It has been demonstrated that porcine Xist gene expression may be affected by maternal metabolic state at the time of ovulation (Vinsky et al., 2007). As this study shows, sows who were in a negative metabolic state during the week before ovulation and fertilization not only demonstrated greater than usual embryonic mortality by D30, but the mortality of female embryos was greater than that of the male embryos. This was attributed to aberrant Xist expression in female embryos, suggesting that maternally influenced epigenetic defects may contribute to sex-biased embryonic loss in the pig.
Anomalies in sex determination and differentiation

As indicated above, the SRY gene plays a critical role in sex differentiation, and usually only embryos carrying the Y chromosome possess SRY. However, SRY can become non-functional or be transferred from the Y chromosome to the X chromosome by a rare recombination event. Such events can cause sex reversal, whereby XY individuals become females and XX individuals become males. In either case, intersex individuals can arise (Cattanach et al., 1982). In humans, XY sex reversals are rather frequent (about 1 in 3000 newborns) and are genetically heterogeneous. XX sex reversals, on the contrary, are rare (about 1 in 20,000 newborns) and are usually caused by the translocation of SRY. Surprisingly, in the majority of human cases (~75%) the cause of sex reversal cannot be precisely identified (Nef and Vassilli, 2009).

A recent comparative study of several mammalian species, including the pig, revealed potentially important upstream-located SRY regulatory elements, mutations of which might lead to XY sex reversal (Ross et al., 2008). It seems, however, that currently available data do not point towards a significant importance of this phenomenon in the pig, as the majority of sex reversals found so far are mainly SRY-negative individuals with a 38 chromosome, XX (38,XX) karyotype. A number of pig intersexes with ovo-testes were examined. These animals were usually infertile without clear spermatogenetic or oogenic activity. In some of them, the presence of a properly developed uterus and ducti deferens was observed, but oviducts were not found. In many of these cases, the sex-reversal status was likely to have been caused by an autosomal recessive mutation (Hunter, 1996; Switoński et al., 2002). In an attempt to understand the aetiology of intersexuality in such pigs, the gonads of 38,XX (SRY-negative) female to male sex-reversed animals were studied during fetal life (50 and 70 days post coitum (dpc)), after birth (35 days post-partum (dpp)) and during adulthood. At the fetal stages, an elevated expression of SOX9, AMH and the P450 aromatase gene (CYP2E1) were observed. In addition to genes involved in the testicular pathway, the same gonads expressed an ovarian-specific factor FOXL2. The genes involved in this pathology pathway act early during gonadogenesis and affect the ovary-differentiating pathway. This process occurs with variable expressivity, producing embryos with abnormalities ranging from ovarian germ cell depletion through to trans-differentiation into testicular structures (Pailhoux et al., 2001).

The rare identification of XY pig females which carry a duplication of a certain X chromosome region (Xp21) led to the hypothesis that a double dose of a gene might cause sex reversal. The DAX1 gene (NR0B1), was isolated from this region and considered as a candidate. In porcine embryonic gonads, DAX1 expression starts in the urogenital ridges of both XX and XY porcine embryos between D20 and D22 of development, and continues until adulthood (Parma et al., 1997). A recent human study with a similar gonad disorder confirmed that indeed the NR0B1 (DAX1) gene was duplicated in such patients (Barbaro et al., 2008).

Aneuploidy and chimerism for sex chromosomes were also described in pigs. Depending on karyotype, such individuals may widely vary from nearly normal males to nearly normal females, with all kinds of intersexes in between (see Chapter 7).

Summary

The progress in developmental genetics of the pig has been significant in recent years, particularly in regard to conceptus growth, fetal–maternal interactions and placental attachment. However, the understanding of species-specific features of the earlier stages of development up to blastocyst formation and hatching has grown more slowly, and we have to rely on mammalian developmental genetics, which is mainly built on mouse data. One can anticipate that further advancement of molecular genetics methods will eventually improve our understanding of the gene expression profiles of cells and tissues involved in the critical and complex developmental stages in the pig. Through
continued gene discovery and functional genomics research on the establishment and maintenance of pregnancy in the pig, we will be able to manage key developmental events better, and in doing so improve conceptus development and survival rate in the pig.

Note

In this chapter, mouse genes follow the rules for murine genetic nomenclature, which is slightly different from that for other mammalian species: namely, only the first letter of a gene name is capitalized.

References


Maddox-Hyttel, P., Svarcova, O. and Laurincic, J. (2007) Ribosomal RNA and nucleolar proteins from the oocyte are to some degree used for embryonic nucleolar formation in cattle and pig. *Theriogenology* 68 (Suppl. 1), S63–S70.


Introduction

Genetic variation in the pig, as in other farm animal species, is known to be a resource of utmost importance. Pigs, indeed, account for about 40% of the world production of meat. Diversity of the genetic stock, understood as genetic variation, is a prerequisite for pig improvement, because variation is the raw material on which the breeder works (Lush, 1945). Genetic variation in pig populations is also needed to maintain fitness by countering inbreeding depression. Similarly, in natural populations, conservation of genetic variation is needed to allow evolution in response to environmental changes and to maintain population fitness (Reed and Frankham, 2003). Conservation of both agricultural and wild species genetic diversity is covered by the Convention on Biological Diversity (CBD), now ratified by 186 countries, following the Rio Conference of 1992 (FAO, 2007).

In this chapter we shall update the corresponding chapter by Hammond and Leitch in the 1998 (first) edition of The Genetics of the Pig. In the first section, what is known of the status of the world pig genetic resources and how this knowledge has been extended in the past decades will be reviewed. We shall then present the current methodologies for characterizing genetic resources by the use of porcine genetic markers, a field of research that has...
undergone considerable development since the 1980s. In the last section, potential links between molecular and quantitative trait diversity will be discussed, as well as information derived from molecular studies of diverse genetic stocks for elucidating the genetic basis of production and fitness traits, a field presently in considerable progress.

**Status of the World Pig Genetic Resources**

**Inventories**

A proper characterization of the available pig genetic resources is obviously needed for their management. Several levels of characterization exist. One of these can be achieved by a comprehensive inventory of the available breeds and populations. Pig genetic resources include a wide spectrum of populations – or breeds – and various methodologies have been proposed to classify breeds of farm animals. The first attempt was probably in Mason’s *World Dictionary of Livestock Breeds, Types and Varieties* (1951). Mason’s approach was later developed and clarified by Lauvergne (1982), who proposed classifying breeds into four categories on a domestication evolutionary scale, namely: (i) wild or feral populations; (ii) traditional or indigenous populations, usually not uniform and lacking any breeders’ organization; (iii) standardized breeds, usually managed within breeding organizations; and (iv) selected lines, generally part of crossbreeding systems. A milestone in the inventory of farm animal populations has been Mason’s *Dictionary*, mentioned above, which has remained a classic with many further editions until 1996, and a new (5th) edition revised by Porter that was published in 2002. In the 1988 edition, which presents a compilation of over 500 pig populations (Mason, 1988), one can find representatives of the above four categories, with a marked predominance of categories (ii) and (iii).

Extensive country-driven inventories of farm animal populations have subsequently been launched, first in Europe in the early 1980s under the aegis of the European Association for Animal Production (see Simon and Buchenauer, 1993, for details). Later, after the Rio Convention of 1992, in recognition of the importance of domestic animal genetic resources, FAO was ascribed a mandate for their management, and initiated a special action programme, the details of which are presented in Hammond and Leitch (1998). Within this framework, FAO launched an inventory and basic description of the breeds of domestic species, which can now be accessed from the Domestic Animal Diversity Information System (DAD-IS) (http://dad.fao.org/). An impressive amount of information has thus been gathered all over the world during the last 25 years, and this has recently been integrated into a global assessment of the world’s animal genetic resources (FAO, 2007). Table 13.1 shows the distribution of the pig breeds presently known (in 2009) over the regions of the world, and compares this with the situation in

<table>
<thead>
<tr>
<th>Region of the world</th>
<th>Africa</th>
<th>Latin America and the Caribbean</th>
<th>North America</th>
<th>Asia and the Pacific</th>
<th>Europe</th>
<th>Near East</th>
<th>Total pigs</th>
<th>Mammalian domesticated species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988 situation</td>
<td>7</td>
<td>56</td>
<td></td>
<td>152</td>
<td>160</td>
<td>0</td>
<td>375</td>
<td>3327</td>
</tr>
<tr>
<td>1995 situation</td>
<td>13</td>
<td>24</td>
<td>28</td>
<td>157</td>
<td>129</td>
<td>2</td>
<td>353</td>
<td>2442</td>
</tr>
<tr>
<td>2009 situation</td>
<td>143</td>
<td>241</td>
<td>64</td>
<td>422</td>
<td>495</td>
<td>12</td>
<td>1377</td>
<td>7835</td>
</tr>
<tr>
<td>Increase (fold)</td>
<td>20.4</td>
<td>5.4</td>
<td>2.4</td>
<td>2.8</td>
<td>3.1</td>
<td>nd</td>
<td>3.7</td>
<td>2.4</td>
</tr>
</tbody>
</table>

*FAO definition.
*Cattle + horses + goats + pigs + sheep.
nd, no data.

1995 and 1988. It reveals a nearly fourfold increase in global coverage over the last 21 years: 1377 breeds in 2009 as against 375 in Mason’s dictionary (Mason, 1988).

This enormous increase in the number of ‘breeds’ needs explanation. The first reason for the list enlargement is a growing awareness all over the world of the usefulness of identifying the amount of genetic diversity available in each country, and consequently an improved reporting under the incentives provided by FAO’s Global Programme for the Management of Farm Animal Genetic Resources (Hammond and Leitch, 1998). The second reason lies in the redefinition and broadening of the concept of genetic resource. Interest in animal genetic resources arose in the 1960s from an awareness of the potential dangers created by a continuous decrease in the number of breeds commercially exploited, and emphasis was therefore initially put on the conservation of endangered breeds. Then a distinction was made more clearly between conservation per se and general livestock resource management (see Barker, 1986). It is now recognized that, as mentioned above, the concept of a genetic resource should include all identified commercial breeds or lines, whatever their size or importance, as well as more specific categories of animals, such as wild populations or highly selected lines developed for research purposes. A distinction is also now being made at the FAO level between local (or ‘regional transboundary’) breeds that occur only in one country (or region) and ‘international transboundary’ breeds that occur in more than one region. In pigs, there are relatively few regional transboundary breeds, and the simpler distinction between local and international breeds may therefore be retained for all practical purposes. A total of 33 international pig breeds have been reported, five of which are widely predominant: Duroc, Hampshire, Landrace, Large White (Yorkshire) and Piétrain. Several of these entities may themselves be divided into different country varieties that can be considered as truly different breeds, such as, for example, Belgian and Danish Landrace. For more details on breeds of pigs see Buchanan and Stalder (Chapter 18).

Use of breed information

An important use of breed information is for assessing the degree to which each breed is exposed to the risk of becoming extinct, i.e. its degree of endangerment (DE). Assessing DE properly, however, is a difficult task because it requires integrating genetic and demographic factors which interact in various ways to determine the survival prospect of any breed; this was stressed by Gandini et al. (2004), who discussed various systems used for classifying breeds based on their DEs.

The most used genetic criterion of DE is effective population size ($N_e$), which was defined by Wright (1931) as $N_e = 4MF/(M + F)$, which combines the number of breeding males ($M$) and females ($F$) and allows prediction of the rate of increase ($\Delta F$) in inbreeding as $\Delta F = 1/2N_e$. A critical effective size may then be defined as the one allowing a given level of inbreeding to be reached within a given period of time (e,g, 50 years), a system proposed by Simon and Buchenauer (1993), and now used by the European Farm Animal Biodiversity Information System (EFABIS) (http://efabis.tzv.fal.de/). In the pig, critical effective sizes for inbreeding values of $F = 0.25$ or 0.40 within 50 years are 87 and 49, respectively (Ollivier, 1998). It should, however, be noted that inbreeding is expected to increase faster than is assumed from effective size. Two main reasons explain this: the first is that selection by itself increases average kinship among breeding animals and tends to accelerate inbreeding, as first noted by Robertson (1961), and the second is that effective size derived from the number of breeding animals may be overestimated when generations overlap. The overestimation depends on age at first offspring and survival rate (or culling policy) of the breeding animals. In the pig, assuming geometric age distributions for boars and sows, the overestimation may range from 14 to 72% (see Ollivier and James, 2004, Table 1 and references therein).

Breed extinction may otherwise be modelled in demographic terms. Demography classically
Pig Genetic Resources 309

considers population growth and its variance as due to the combination of strictly demographic stochasticity (birth and survival processes), environmental stochasticity (unpredictable changes in the environment) and catastrophes such as disease outbreaks. As an example, rates of growth of 110 breeds of cattle over 13 years have been analysed by Gandini et al. (2004) and shown to be significantly affected by herd size and country. More data would be needed for predicting the variance of population growth rate in a given breed or over a set of similar breeds. Continuous reporting on breed structure and inbreeding is encouraged in DAD-IS and may, in the course of time, become a useful tool for monitoring DE. In Table 13.2 it is shown that, together with a marked increase in the number of breeds being reported, breed-related information remains far from being complete. More than one-third of all reported pig breeds lack population data and cannot be evaluated for DE; this is also the average situation for other domestic species. Another weakness of the current monitoring of breed erosion is that it cannot capture the effects of uncontrolled crossbreeding (FAO, 2007). Table 13.2 shows a tendency for an increase of risk status over the last decade. This overall picture, however, covers widely different situations according to the region of the world, which is seen when referring to DAD-IS.

Breed inventories also include information on the production characteristics of the various breeds, but these data are mostly confounded with breed environment and therefore do not tell much about real genetic diversity. Within a country, however, confounding may to some extent be avoided by expressing breed performance as a deviation from a standard breed, as suggested by Simon and Buchenauer (1993). Well-designed experiments, such as diallel crosses, would be too costly to implement in pigs because of the considerable testing facilities required as soon as the number of breeds exceeds three or four. But, when the number of breeds to compare is large, multi-breed designs would allow the testing of a large number of breeds using a few animals per breed (Taylor, 1976). This offers a suitable approach for testing diversity among a large number of rare breeds, but it has rarely been implemented in pigs, to our knowledge.

**Conservation**

Strategies for maintaining domestic animal genetic diversity depend on the degree of control humans have over the available livestock populations. In that respect, two broad categories of populations may be considered, namely commercial populations, in which the issue is to maintain internal diversity, and genetic stocks (or so-called gene banks), which may be used as stores for maintaining between-breed diversity.

As mentioned above, genetic variation in commercial populations tends to decrease more or less automatically under the combined effects of restricted population size and selection. When pedigrees are missing, effective size may be manipulated by changing the herd demography parameters, such as age at first breeding and culling policy. As an example, the ratio $N_e/N$ of effective size to herd size ($N$) is expected to increase to above one when male turnover is high. In fact, the balance between the positive effects on generation interval of high survival rate or late culling and their negative effects on genetic drift plays in favour of short breeding lifespan and early culling (Ollivier and James, 2004). When

### Table 13.2. Risk status of pig breeds. Sources: 1995 data: Hammond and Leitch (1998); 2006 data: from Fig. 12 in FAO (2007).

<table>
<thead>
<tr>
<th>Breed status</th>
<th>1995 situation</th>
<th>2006 situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of breeds on file (breeds extinct excluded)</td>
<td>353</td>
<td>599</td>
</tr>
<tr>
<td>Number of breeds with population data and percentage of total on file (in parentheses)</td>
<td>265 (75.0%)</td>
<td>374 (62.4%)</td>
</tr>
<tr>
<td>Number of breeds at risk and percentage of breeds with population data (in parentheses)</td>
<td>69 (26.0%)</td>
<td>133 (35.6%)</td>
</tr>
</tbody>
</table>
pedigrees are available, effective size may be easily monitored through the calculation of individual inbreeding coefficients. Calculations may, however, become critical for very large populations, as discussed by Colleau (2002), who has proposed an indirect method that is able to considerably reduce computation time. Various software is available for pedigree analysis, such as (among others), PEDIG, from Boichard (2002) – now updated to PEDIG2007 (Boichard, 2007). Pedigree analyses may reveal unexpected bottlenecks and subsequent episodes of rapid increase of inbreeding, even in quite large populations of pigs of the international breed category (Maignel and Labroue, 2001). Pedigree information allows the control of effective size very efficiently through mating schemes intended to minimize inbreeding/genetic drift (see Fernandez et al., 2003, and Sanchez et al., 2003, for a variety of schemes). Robertson (1960) has argued that such schemes, while aiming at increasing effective size for a given population size, generally lead to a reduction of the genetic variance exposed to selection, and then cannot increase the limit eventually reached by selection. In the short/medium term, however, a joint management of inbreeding and selection advance is recommended (Kinghorn et al., 2009). Compromises between selection advance and genetic drift may be needed in pig breeding schemes, but it should also be kept in mind that optimal mating schemes may considerably reduce inbreeding without any detrimental effect on selection response, as shown by Colleau and Tribout (2008) in a retrospective study referring to matings actually performed. The theory of genetic contributions by Woolliams and Thompson (1994) can be used to predict the rate of inbreeding in populations undergoing selection. Using this theory, Bijma et al. (2001) showed how inbreeding can be predicted in specific livestock improvement schemes.

Conservation of animal genetic resources, understood as maintenance of between-breed diversity, may be achieved either by live animals (in vivo conservation) or by cryogenic storage (in vitro conservation). Let us also note here that in vivo conservation itself may be achieved in situ, i.e. under commercial farm conditions, or ex situ. This distinction, however, often remains unclear. Live animal gene banks offer good opportunities for controlling genetic drift/inbreeding, particularly when a strict genetic management can be guaranteed by institutions such as farm parks, museums, zoos, national parks, universities or research organizations. Several examples of (long-term) conservation of farm animal stocks exist for various species (see review in FAO, 2007, p. 461), of which examples are given for sheep by the French (Rambouillet) Merino flock (Prod’Homme and Lauvergne, 1993) and for pigs by the Spanish Large White herd at Pontevedra, Spain (Solano, 1984). Although those projects differ in timespan and genetic management, they share the common feature that, in spite of the high levels of inbreeding generally reached, fitness traits show no major deterioration. This indicates that natural selection is able to prevent inbreeding depression when the rate of increase of inbreeding is moderate, probably by purging deleterious recessive genes. On the other hand, the contribution of mutations to quantitative trait variation may guarantee that a reasonable level of genetic variation be maintained over the long term (Hill, 1982). Such a theoretical prediction would particularly apply to traits not currently selected for, such as future breeding goals, which are by definition unpredictable and consequently under no direct selection pressure. The maintenance of fitness in artificial selection programmes has been recently reviewed by Hill and Zhang (2009).

Cryogenic storage of animal genetic resources (in vitro conservation) implies various harvesting, processing, monitoring and maintenance procedures applied to biological materials such as animal cells, semen, oocytes or embryos (for details of the techniques used in pigs, see the review of Pizzi et al., 2001, and also Ross and Prather, Chapter 11). Storing semen and embryos is probably the most secure way of preserving genetic stocks, and it may also be the cheapest. The costs of different methods of conservation of farm animals were evaluated for the first time by Smith (1984), who stressed the contrast between a high initial investment for establishing cryogenic banks, particularly when embryos are to be collected, and the low annual storage costs. Consequently, cryogenic storage becomes cheaper than live animal conservation in the
course of time. More recent papers on the design of cryopreservation in farm animals have been reviewed by Labroue et al. (2001), who gave particular attention to the pig, a species in which cryogenic storage must essentially rely on frozen semen, because embryo conservation is not yet fully operational, in spite of recent progress in ultra-rapid cooling techniques. Security constraints have to be emphasized, implying that semen should be stored in at least two locations, in order to be able to replace the stock accidentally lost in one location by using the quantities stored in the other. In addition, when only semen is stored, the renewal of a breed that is becoming extinct should be made possible in the absence of living females of the same breed. This requires grading up any available female population using the stored semen through a backcrossing system over a sufficient number of generations.

Labroue et al. (2001) have proposed the following general expression for the number of doses \((D)\) of semen to be stored per pig breed in each location, as a function of \(d\), the number of doses needed per insemination, \(N\), the number of fertile gilts aimed at generation \(n\) of the repeat backcross scheme, and \(r\), the expected number of inseminations needed to obtain a fertile daughter, i.e. the inverse of the number of fertile daughters per insemination: \(D = dN(r^n - 1)/(r - 1)\). The application of this formula to various situations has been discussed by Labroue et al. (2001). Assuming \(d = 2\), i.e. double-dose artificial insemination (AI), \(N = 25\), a conservative value of 1.5 for \(r\), and \(n = 6\), it can be seen that over 2000 doses are needed to accommodate two locations. Labroue et al. (2001) also evaluated the cost of a breed collection under the above assumptions, and assuming 25 boars collected, as recommended by Smith (1984). On the basis of 2000 doses (80 doses from each of the 25 boars), their evaluation of €30,000 in 2001 would presently amount to about €35,000 per breed. Annual storage costs represent about 1% of this amount.

Table 13.3 summarizes in vivo and in vitro conservation activities in various regions of the world, with the latter technique playing a predominant role. Apart from Asia and Europe, which harbour many local breeds and where about a quarter of these are being preserved, a large variation can be seen – from a nearly complete preservation in North America (though of a small number of local breeds), to a contrastingly much lower coverage in Latin America and the Caribbean, and no reported in vitro conservation in Africa.

### The Use of Genetic Markers for Characterizing Genetic Diversity

Although the number of breeds known to exist in a country, or in a region of the world, is a valid indicator of some diversity, it remains of limited value when comparisons are being made between different countries or regions, because the data presented very much reflect the information provided by each country. In spite of the efforts made to coordinate inventories across different countries or regions through DAD-IS, the information displayed should not be considered as necessarily representing a comprehensive and comparable coverage of...
Genetic diversity between breeds

One classical method of measuring genetic diversity is the calculation of fixation indices ($F$) as proposed by Wright in 1943 (see Nei, 1977). The $F_{ST}$ in particular measures the degree of genetic differentiation among a set of populations based on analyses of variance of allele frequencies (Weir and Cockerham, 1984). A different approach is that of Nei (1973), who has proposed a concept of gene diversity based on the actual heterozygosities observed at several loci, and shown that total gene diversity can be partitioned into intra-populational and inter-populational gene diversities. Wright’s $F_{ST}$ and Nei’s $G_{ST}$ both measure genetic differentiation between groups of individuals. A method for measuring genetic diversity based on Nei’s diversity parameters has been proposed by Petit et al. (1998).

Another method, advocated by an economist (Weitzman, 1992), is based on the principle of ‘monotonicity in species’, according to which the addition of any population $i$ (species, breed, ...) to a set of populations ($S$) should increase its diversity ($V$) by an amount at least equal to the distance $d(i, S)$ between breed $i$ and the original set $S$, as represented by the expression: $V(S∪i) ≥ V(S) + d(i, S)$. Here, $d(i, S)$ is the so-called ‘Hausdorff’ distance, defined as the distance between $i$ and its closest neighbour in $S$. In fact, a distance in the mathematical sense is not necessary; a dissimilarity measure will do, and in genetics we can take a classical distance such as the Nei-standard, Nei-DA, Cavalli-Sforza, Reynolds or Sanghvi (Foulley and Hill, 1999; Ollivier et al., 2005). This principle led Weitzman to define the diversity of the set $S$ as the maximum, over all members of the set, of the distance of a member from its closest neighbour in the set augmented by the diversity of the set without that member. Taking $S\setminus i$ to define the set without member $i$, $V(S) = \max_{i} d(i, S\setminus i) + V(S\setminus i)$ measures the diversity of $S$. The maximization procedure is detailed in Weitzman (1992), and illustrated on a set of cattle breeds by Thaon d’Arnoldi et al. (1998). The method provides a way of measuring the contribution of any individual breed to the diversity of a given set of breeds, because the $V$ function can obviously be applied to any subset of $S$. The contribution of any given breed $i$ to between-breed diversity ($CB$) may be expressed as $CB_i = 1 – V(S\setminus i)/V(S)$. This can also be viewed as an estimate of the breed’s genetic uniqueness. In addition, the method generates a rooted tree, which may be interpreted as a taxonomic tree (see below), on which diversity can be ‘read’, as each branch length represents the contribution of the corresponding breed to total breed diversity (see examples in the following subsection on taxonomy and clustering). A software implementation of the Weitzman method, down to tree drawing, is available (Derban et al., 2002, updated in 2005).

As an illustration of the measurement of genetic diversity, a partitioning of MS diversity among 22 breeds of pigs is presented in Fig. 13.1; the analysis is more completely discussed in Ollivier and Foulley (2009). The data are drawn from a European investigation described by Ollivier (2009). In this subsample, as in the complete study, which involved 68 European domestic breeds (see Fig. 13.2), more than half of the breed diversity could be assigned to the 11
Fig. 13.1. Partition of diversity among 22 breeds of pigs. Breed contributions relative to total between-breed diversity (3.6) were computed over 50 microsatellite loci, using the diversity function of Weitzman (1992), applied to pairwise Reynolds genetic distances (Reynolds et al., 1983). Breed codes are: MS – Meishan (China); PR – Presticke (Czech Republic); CR, L1, Y2 – Créole, Landrace, Yorkshire Large White (France); AS, BB, H2, L2, P1, Y1 – Angler Sattelschwein, Bunte Benheimer, Hampshire, Landrace, Piétrain, Yorkshire Large White (Germany); CS, D1, NS – Cinta Senese, Duroc, Nera Siciliana (Italy); PU – Pulawska (Poland); BI – Bisaro (Portugal); RE – Retinto (Spain); D2, GO, H1, P2, TA – Duroc, Gloucester Old Spot, Hampshire, Piétrain, Tamworth (UK). Adapted from Ollivier and Foulley (2009).

Fig. 13.2. Partition of European pig diversity among three categories of breeds based on microsatellite (MS) and amplified fragment length polymorphism (AFLP) molecular markers, as described in Ollivier et al. (2005). The three breed categories were: Loc, local; Int, international; Com, commercial line. CB, category contribution (%) to between-breed diversity, applying the method described for Fig. 13.1. CW, category contribution (%) to within-breed diversity, based on expected heterozygosities. MS, data from 50 microsatellite loci on 68 European domestic breeds (total between-breed diversity 11.6 on the Reynolds distance scale, average expected heterozygosity 0.56). AFLP, data from 58 European domestic breeds (total between-breed diversity 4.5 on the Reynolds distance scale, average expected heterozygosity 0.11). Reproduced from Ollivier (2009) with the kind permission of Cambridge University Press.
local breeds. The method has already been applied to several farm animal species (see references in Ollivier and Foulley, 2009), and, as in this pig example, large contributions of ‘native’ breeds have been found in several other species.

Other methods for analysing diversity have been proposed, essentially based on the concept of co-ancestry (Caballero and Toro, 2002; Eding et al., 2002). Those methods offer the additional possibility of adjusting the contribution of each breed to a ‘core set’ in order to maximize diversity. This refinement may be of interest to in vitro conservation with limited budgets. An application of the method of Caballero and Toro (2002) to Iberian pig diversity has been made by Fabuel et al. (2004). The method has been shown to give in practice results very similar to the Weitzman method in terms of between-breed diversity, though it is liable to yield some negative breed contributions (Ollivier and Foulley, 2005).

**Genetic diversity within breeds**

The need to incorporate within-breed variability into diversity assessments has often been emphasized in plant as well as animal genetic resources analyses (e.g. Petit et al., 1998; Barker, 2001; Caballero and Toro, 2002). By the way, this may be seen as an equivalent for farm animals of the ‘principle of complementarity’ that is invoked in species conservation, whereby species are valued not only according to their dissimilarity from other species, but also according to their intraspecific diversity (Bonin et al., 2007). The contribution of breed i to within-breed variability may be defined, in a way similar to CB, as $CW_i = 1 - \frac{H(S\setminus i)}{H(S)}$, where $H(S)$ is the mean heterozygosity of the set S, and $H(S\setminus i) = \frac{(\sum_{i=1}^{n} H(i))}{(n-1)}$ is the mean heterozygosity of the set after excluding breed i, assuming n breeds in S, where l represents each breed of set S different from i. As $\sum_{i=1}^{n} H(S\setminus i) = \sum_{i=1}^{n} H(i) = nH(S)$, one can see that the contributions to within-breed diversity add up to zero over breeds. Negative CWs are thus to be expected for highly homozygous breeds, as their extinction raises the average heterozygosity of the remaining breeds. An example of within-breed diversity breakdown is given in Fig. 13.2.

The emphasis on within-breed relative to between-breed variability may vary according to the conservation objective being pursued. Flexibility can be obtained by a global diversity index that gives appropriate weights to the CB and CW of the breeds (Ollivier and Foulley, 2005). This general Weitzman-type approach thus offers a way to capture the maximum diversity in situations as diverse as, for instance, the design of synthetic lines for purposes of selection in farm animals such as the pig, or in the creation of experimental lines of laboratory animals for medical research.

**Allelic richness and allelic diversity**

Allelic richness, defined as the number of alleles per locus, is a diversity measure of obvious interest. The observed allelic richness, however, needs correction for sample size, as the chances of discovering a new allele increase each time a new individual is observed. Several methods have been proposed for taking sample size into account. Basically, allelic richness can be estimated either by rarefaction (El Mousadik and Petit, 1996) or by extrapolation (Foulley and Ollivier, 2006).

In the first case (rarefaction), allelic richness ($r$) is defined as the number of alleles expected to be seen in a sample of specified size $g$, which is the smallest sample of all breeds examined at a given locus. Each allele $k$, observed $N_{ik}$ times in a sample of size $N_i$ from population $i$, has a probability $(P_{ik})$ of not being seen in a sample of size $g$ equal to $P_{ik} = C(N_i - N_{ik}, g) / C(N_i, g)$, where $C(N,g)$ represents the number of combinations of $N$ objects taken $g$ at a time, so that allele $k$ has a probability of $1 - P_{ik}^{(g)}$ of being seen in a sample of size $g$. Then, the allelic richness of population $i$ at any given locus is obtained as $r_i = \sum_{k=1}^{K_i} (1 - P_{ik}^{(g)}) = K_i - \sum_{k=1}^{K_i} P_{ik}^{(g)}$, by adding the $K_i$ alleles observed in the population.

In the second case (extrapolation), the idea is to estimate allelic richness ($R_i$) by adding to the number of alleles actually observed ($K_i$) in the population sample the number expected to be missing owing to sampling. Thus, $R_i = K_i + \Sigma_{m=1}^{N_i} (1 - \prod_{k=1}^{\pi} P_{ik}^{(g)})$, where $\pi_1$, $\pi_2$, ..., $\pi_g$ are the overall allelic frequencies in the whole...
sample of populations considered, and \( km \) means \( k \in A^{(m)} \), the set of alleles actually missing in the \( i \) sample (Foulley and Ollivier, 2006). Notice the contrast between the two formulae. In the rarefaction situation, we subtract from the number of alleles actually observed (\( K \)) the expected number of alleles not seen in a sample of smaller size \( g \), whereas in the extrapolation case, we add to \( K \) the expected number of alleles missing owing to the sampling of a finite specimen \( N_i \) of the population.

The concept of allelic richness leads to the slightly different concept of allelic diversity, which refers to the existence of alleles specific to some breeds. One has indeed to admit that a high number of alleles in a breed does not automatically guarantee their originality. A parallel to the measure of diversity may be drawn with the approach previously described for genetic diversity. Based on the total number of alleles observed in any set of breeds, which is the diversity function of interest, the contribution of breed \( i \) to allelic diversity, equivalent to the \( CB_i \) defined for genetic diversity, is the number of alleles present in population \( i \) and absent in all others, also called ‘private alleles’ (Ollivier and Foulley, 2009). The parallel between genetic and allelic diversity cannot, however, be pursued further, because allelic richness is a property attached to each population, and there cannot be any variation in within-breed allelic diversity.

Here again, correction for sample size is needed to take into account sample size and the number of copies of each private allele. This can be done either through rarefaction or extrapolation, as explained in detail by Ollivier and Foulley (2009), who compare the two methods on the same sample of breeds. The diversity functions based on rarefaction and extrapolation will usually give similar results, at least in terms of breed contributions to diversity. Extrapolation, however, is recommended when sample sizes are, on average, small or highly unbalanced among populations (Foulley and Ollivier, 2006).

**Distinctiveness and conservation potential**

More generality may be given to the diversity functions considered above by taking into account the risks of extinction of each breed. Risks of extinction can be evaluated as indicated previously, using various criteria, examples of which can be found for African cattle in Reist-Marti et al. (2003) or for European pigs in Ollivier et al. (2005). Based on the survival probabilities, \( P_i \), of each of \( n \) breeds over a given period of time, \( 2^n \) extinction-survival patterns may occur with given probabilities and, assuming those events to be independent, an expected diversity can be defined, possibly discounted to present (Weitzman, 1993): \( E(V) = \sum P(Q)V(Q) \), where \( V(Q) \) is the diversity of subset \( Q \) (as previously defined for the diversity of set \( S \), \( V(S) \)), and \( P(Q) \) is its probability.

The distinctiveness \( (D) \) of breed \( i \) is obtained as the partial derivative of expected diversity with respect to \( P_i \), \( D = \frac{\partial E(V)}{\partial P_i} \), then represents the increase in expected diversity within a given period of time with respect to the increase in survival probability of breed \( i \) (Weitzman, 1993).

For instance, in the case of three breeds, the genetic distinctiveness of one breed (e.g. \( D_i^g \)) is \( D_i^g = P_i(1 - P_i) V_{12} + (1 - P_i) P_j V_{13} + P_i P_j V_{123} \), where \( P_i \) is the probability of survival of breed \( i \), and \( V_{12} \) is the diversity of the corresponding subset 12, \( V_{13} \) the diversity for subset 13, etc. A parallel expression holds for allelic distinctiveness, defined as \( D_i^a = P_i(1 - P_i) d(i, 2) + (1 - P_i) P_j d(i, 3) + P_i P_j d(1, 23) \), where \( d(i, Q) \) is the allelic distance (or dissimilarity) between breed \( i \) and subset \( Q \) (here either breed 2, breed 3 or subset 23). This distance is defined as the number of alleles present in breed \( i \) and absent in subset \( Q \), which is also obviously a ‘Haussdorff’ distance because it is equal to the number of alleles present in breed \( i \) and absent in its closest neighbour in \( Q \) (see Equation 23 in Weitzman, 1998). An example of comparative genetic and allelic distinctiveness of 11 endangered breeds of pigs is given in Ollivier and Foulley (2009), where a wide variation in breed distinctiveness is observed, ranging from zero to 24%. The product \( D_i E_n \), where \( E_n = 1 - P_i \) has been named the conservation potential by Weitzman (1993). This can be shown to be the optimal ranking criterion for establishing cryopreservation priorities under a budget constraint, in the particular situation (likely to apply to pigs) when cryopreservation can make a breed safe and cryogenic preservation costs are equal over different breeds (Weitzman, 1998).
**Taxonomy and clustering**

In addition to analyses of genetic diversity, genetic distances can also be used to draw trees, which are often termed phylogenetic trees. The term implicitly refers to evolutionary theory, where diversity arises from speciation, i.e. the division of one ancestor species into two different subspecies. Such a pattern of evolution cannot generally apply to farm animal breeds, except in particular short-term situations when one breed happens to be subdivided into two new ones. Domestic breeds do not remain as distinct as species do and the tree-like branching process in species evolution does not quite apply to farm animal breeds. The trees drawn must be considered as telling the evolutionary story that best fits the diversity observed, but not necessarily telling the ‘true’ story (Weitzman, 1992). The trees drawn are best viewed as classification tools, showing taxonomies rather than true phylogenies. Quite complex migration–admixture patterns usually prevail, as well as occasional unreported mixing of breeds. The resulting trees therefore often disagree with documented ‘old’ breed history and supposed breed proximities (e.g. in Porter, 1993), as discussed for European local breeds by Ollivier (2009).

An example of taxonomy is given in Fig. 13.3. This is the rooted tree generated by applying the Weitzman diversity function to the subset of breeds considered in Fig. 13.1. The longest branch is that of the Chinese Meishan, as expected for such an ‘exotic’ breed in the context of this European investigation. The graph also shows the non-additivity of individual breed contributions, as the joint contribution of Gloucester Old Spot and Tamworth on the left of the graph, represented by the ordinate of their node (below 0.4), is much less than the sum of their individual branch lengths (above 0.6), the reverse being seen for the two Duroc strains located just beside them.

Trees also show clustering phenomena, as exemplified in Fig. 13.3 for the two varieties of each of the five international breeds, Duroc, Hampshire, Landrace, Large White (Yorkshire) and Piétrain, a pattern that is confirmed when all varieties of each breed are considered. In contrast, no clustering of local breeds with international breeds appears, thus suggesting that the uniqueness of European local breeds reflects their having evolved apart from mainstream international breeds.

**Assignment of individuals to breeds and population structure**

Genetic markers can also be used to assign individuals of unknown origin to populations (e.g. breeds or lines). Two cases have to be distinguished here, usually reported in the literature as ‘supervised’ and ‘non-supervised’. In the first case, there is a reference set of populations out of which samples of individuals have been typed for genetic markers. The question then is to know which one of these populations an individual of unknown origin could be assigned to. In the second case, the structure of the populations (numbers and characteristics of each of them) is itself unknown and the problem is to draw inferences about that structure.

By using the estimated allele frequencies in each breed at several unlinked loci, one can compute the likelihood that a given genotype belongs to a given breed, and then assign each individual to the breed showing the largest likelihood (Paetkau et al., 1995). In a study of 11 breeds of pigs using 18 microsatellites, it was thus possible to correctly assign all animals (Laval et al., 2000). A potential advantage of AFLP over microsatellites is the possibility of selecting loci yielding ‘private fragments’ for assignment analyses. This has been shown to be feasible by Alves et al. (2002), who were able to discriminate Duroc from Iberian pigs, owing to 14 private Duroc fragments out of 139 fragments examined. Bayesian procedures can also be applied; these are especially well fitted to take into account missing data at some loci and uncertainty in the knowledge of gene frequencies (Baudouin et al., 2004).

When no objective information is available to define the set of populations examined, model-based clustering methods, as opposed to the standard distance-based methods presented above, can be used. Such clustering methods have been proposed by Pritchard et al. (2000) and Dawson and Belkhir (2001), and they have been widely applied to natural populations. These methods have also gained some popularity in studies of breeds of farm animals, though
in this case the populations are usually well defined and the usefulness of the method may be questioned. The method may, however, help to confirm the degree of genetic robustness of the actual breed structure and reveal possible admixtures. In fact, supervised methods of assignment may gain efficiency if some classification information among reference populations is also taken into account in the statistical procedure, as it is more difficult to assign individuals among close populations (e.g. among Large White lines) than among distant ones (e.g. Large White versus Meishan), as shown by Benezech (2008, unpublished results).

**Comparisons between types of markers**

Both microsatellites and AFLP have been extensively used in the pig, but rarely on the same set of breeds. When this was done, and meaningful comparisons could be made, it appeared that overall genetic diversity in AFLP was considerably below that in MS, both within-breed (0.12 versus 0.56 for expected heterozygosity) and between-breed (0.11 versus 0.23 for $F_{ST}$) (Foulley et al., 2006). In spite of these differences in total diversity, the individual breed contributions to both diversities were positively correlated between the two markers (see the similarity between the MS and AFLP graphs in Fig. 13.2). The correlations ($r = 0.5$), however, were moderate and somewhat lower than would be expected if the evolution of the two markers had been governed mainly by genetic drift. This suggests that MS and AFLP may carry different diversity information (Foulley et al., 2006), in accordance with what has been reported in plants (see the review by Nybom, 2004).
Relationships Between Molecular and Quantitative/Adaptive Trait Divergence Revealed by Genetic Resources Investigations

Tests of marker neutrality

Studies of DNA markers offer objective measures of diversity which can help decision makers to identify the most unique breeds, which may then be prioritized for conservation, as we have seen above. They also provide information on the history of domestication (e.g. Bruford et al., 2003). The relevance of molecular markers in studies of diversity, however, has been questioned on the grounds that neutral loci are being used, and these may not reflect differences between populations at the loci under selection. We know, however, that neutral genes can be affected by the selection applied to neighbouring genes, a phenomenon known as gene hitchhiking (Maynard Smith and Haig, 1974). Selection acts on the whole genome, and diversity should be viewed in the context of multi-locus systems. This situation is, in fact, exploited in reverse in marker-assisted selection procedures using markers close to QTLs. Quite extensive QTL maps are now available for several farm animals, including the pig (Hu and Reecy, 2007). Most of the microsatellites used in pig diversity studies are, indeed, linked to a large number of quantitative traits. This is the case in particular for the ISAG (International Society for Animal Genetics)/FAO panel of microsatellites listed in Laval et al. (2000), as shown in Table 13.4, established from the information available in the pig QTL database described by Hu et al. (2005). One would then a priori expect to find some correlation between marker and quantitative trait diversity, particularly for the markers most closely linked to QTLs. The testing of marker neutrality, however, is a challenging task. The tests are based on genome scans of DNA polymorphisms because the adaptation of domestic breeds to local conditions or to specific production objectives is expected to generate changes in within-breed and/or between-breed diversity at those loci underlying the traits under selection, and at nearby neutral marker loci. This is a subject also extensively investigated in natural populations (as reviewed, among others, by Storz, 2005).

The neutrality tests based on relative levels of diversity within populations exploit the reduction of variability around a selected locus owing to hitchhiking, a phenomenon called ‘selective sweep’. A test based on the variance of repeat number at microsatellite loci has been proposed by Schlötterer et al. (1997). Low variances for particular locus × population combinations reveal directional selection, whereas high variances indicate within-breed balanced polymorphisms. This test, however, is sensitive to the model assumed, and a more robust test, which compares two groups of populations for a large number of microsatellite loci, has been proposed by Schlötterer (2002). The test statistic is the log of the ratio of variance in repeat number. An application of this test to a set of European pig breeds typed for 50 microsatellite markers showed the existence of ‘outlier’ loci with contrasting allele size distributions between groups of breeds, indicative of selective sweeps (Ollivier and Foulley, 2009).

Differentiation between populations as measured by $F_{ST}$ is the basis of an early test of selective neutrality (Lewontin and Krakauer, 1973). The basic argument behind the Lewontin–Krakauer (LK) test is that, under the null hypothesis of neutrality, differentiation at all loci should be the same. The observed variance of $F_{ST}$ across marker loci can thus be tested against its expected value under the assumption of neutrality. Several improvements of the LK test have been proposed (reviewed by Ollivier and Foulley, 2009), which are essentially for taking into account the pattern of relationships among populations. After correction for this effect, the LK test showed a particularly large departure from neutrality for AFLP in a set of European pig breeds (Foulley et al., 2006).

Perspectives opened by single nucleotide polymorphisms (SNPs)

The lack of a sufficient number of genetic markers is recognized as a major limitation in the task of detecting marker–trait associations,
Table 13.4. Number of traits reported in the literature as being linked to the 27 ISAG (International Society for Animal Genetics)/FAO Advisory Group recommended microsatellites. Data on QTL flanking markers from the Pig Quantitative Trait Locus (QTL) database (Pig QTLdb) at http://www.animalgenome.org/cgi-bin/QTLdb/SS/index (as of December 2009).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome arm</th>
<th>Number of traits</th>
<th>Number of traits linked to</th>
<th>Trait classa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meat quality</td>
<td>Production</td>
</tr>
<tr>
<td>CGA</td>
<td>1p</td>
<td>11</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>S0155</td>
<td>1q</td>
<td>9</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>SW240</td>
<td>2p</td>
<td>27</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>S0226</td>
<td>2q</td>
<td>2</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>SW72</td>
<td>3p</td>
<td>40</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>S0002</td>
<td>3q</td>
<td>1</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>S0227</td>
<td>4p</td>
<td>14</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>S0005</td>
<td>5q</td>
<td>22</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>IGF1</td>
<td>5q</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>SW122</td>
<td>6q</td>
<td>27</td>
<td>26</td>
<td>–</td>
</tr>
<tr>
<td>S0228</td>
<td>6q</td>
<td>19</td>
<td>19</td>
<td>–</td>
</tr>
<tr>
<td>SW632</td>
<td>7q</td>
<td>8</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>S0101</td>
<td>7q</td>
<td>19</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>S0225</td>
<td>8q</td>
<td>9</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>S0178</td>
<td>8q</td>
<td>4</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>SW911</td>
<td>9p</td>
<td>20</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>SW951</td>
<td>10q</td>
<td>9</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>S0386</td>
<td>11q</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>S0090</td>
<td>12q</td>
<td>21</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>S0068</td>
<td>13q</td>
<td>24</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>S0215</td>
<td>13q</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>SW857</td>
<td>14q</td>
<td>12</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>S0355</td>
<td>15q</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>SW936</td>
<td>15q</td>
<td>27</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>S0026</td>
<td>16q</td>
<td>21</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>SW24</td>
<td>17q</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>S0218</td>
<td>Xq</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total number of trait-linked markers per trait class</td>
<td>25</td>
<td>17</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

*aAs defined on the above-mentioned QTL database web site; ‘meat quality’ includes carcass traits, ‘production’ mainly covers growth traits, ‘exterior’ includes behaviour.

In spite of the considerable numbers of loci from MS and AFLP already available. In recent years, new sequencing technologies have appeared and allowed the design of high-density genotyping assays. Large numbers of single nucleotide polymorphisms (SNPs) have thus been identified in several species of domestic animals (references in Ramos et al., 2009). This kind of genetic marker is especially appealing, as SNPs are biallelic, very common (one per 1000 bp) and easily assayed. SNPs have also been recently identified in the pig by Ramos et al. (2009). Using a ‘PorcineSNP60 Beadchip’, they were able to reliably score 62,621 loci, of which 58,994 were polymorphic in the 158 pigs they used. The availability of such a tool now makes it possible to more efficiently detect outlier loci by using a model-free approach along the line of the LK test previously described, as in the human example of Akey et al. (2002). Other methods for distinguishing loci under selection from neutral loci require assumptions on the demographic history of the populations and may be sensitive to the model implemented. Model-based
approaches using Bayesian hierarchical methods have recently been developed and applied to SNP data (Beaumont and Balding, 2004; Riebler et al., 2008; Gautier et al., 2009). Approximate Bayesian Computation (ABC) also allows the implementation of more sophisticated models of molecular evolution (Pritchard et al., 2000; Beaumont et al., 2002). SNPs thus open a new field of investigation in pig genetic diversity, already exemplified in cattle by the Bayesian scan of Gautier et al. (2009).

SNPs will also help to evaluate the extent of linkage disequilibrium (LD) in the various breeds of pigs. LD is a non-random association of genes at different loci, and is known to decrease rapidly with increasing map distance of the loci considered. Detecting significant LD, therefore, needs narrowly spaced genetic markers. One of the earliest studies of LD in pigs actually used 15MS, spaced 5cM on average, and was able to show significant LDs on two pig chromosomes (Nsengimana et al., 2004). The availability of SNPs now opens the way to more detailed investigations, as shown by a recent comparison by Amaral et al. (2008) bearing on ten European breeds, ten Chinese breeds and a European wild boar. This study of 371 SNPs revealed more extended LD in Europe than in China, with the wild boar in an intermediate position. Interestingly, this study also showed a very highly significant breed × genome region interaction for LD, which might reflect differential selection pressures across genomic regions among breeds. LD might then be an additional component of genetic diversity worth considering.

Conclusions
Interest in farm animal genetic resources, including the pig, emerged in the early 1960s, but, as noted by Barker (1986), for the following quarter of a century, there was ‘a great deal of talk and relatively little action’. The situation has since undoubtedly changed. As we have seen in this chapter, substantial advances have been made in the inventories of the pig genetic resources, in their management and conservation, and in the characterization of their genetic diversity. The need to properly evaluate diversity and to put its management into an economic context, integrating diversity and vulnerability, cannot be overemphasized. International cooperation is essential, as is funding, as has been recently exemplified by the enormous task of assessing the status and trends of animal genetic resources worldwide, based on 169 country reports (FAO, 2007). More research is certainly needed for objectively assessing fitness and adaptation traits among populations (van der Werf et al., 2009), and on the controversial issues around molecular and quantitative trait diversity. Molecular characterization of pig genetic resources should be encouraged, both as a tool for managing the resources and to throw more light on the genetic basis of adaptive divergence, so as to increase our knowledge of the pig genome to better preserve its diversity. Help will be provided in that respect by new molecular technologies, as exemplified by the PorcineSNP60 Beadchip recently developed (Ramos et al., 2009), as well as by new computing methods based on more complex models of population genetics (e.g. Beaumont et al., 2002) that are applicable to farm animal populations.

References


Ollivier, L. (1998) Results and prospects of reproductive technologies for (sustainable) breeding and conservation in farm animal species other than cattle. *Acta Agriculturae Scandinavica* Suppl. 29, 115–120.


Introduction

Performance of an animal destined for market in a pork production system is defined by the efficiency with which it develops saleable product. That efficiency is largely determined by costs associated with feed and time, and by the amount of quality lean tissue produced. Genetic merit for market animal performance is improved in seed-stock populations by performance testing and selection. For successful commercial pork production, a breeding system must be implemented that optimizes market pig performance and reproductive efficiency. Thus, the genetics of performance encompasses not only additive and non-additive genetic effects associated with feed intake and tissue growth, but also strategies for implementing performance testing and selection, and consideration of the genetic correlations between market pig performance and traits of reproduction.

Performance Testing and Selection Strategies

Classical index selection

A classical approach to multi-trait selection involves the derivation of a linear index of phenotypic measurements with weightings that maximize its correlation with a selection objective (Hazel, 1943). The selection objective is defined
as a linear combination of breeding values for traits considered of economic importance, and phenotypic measurements are chosen as criteria with which to most effectively estimate genetic merit for the selection objective. The traits that define the selection objective may or may not be among the phenotypic measurements chosen, depending on the difficulty of measurement and the availability of correlated indicator traits. Estimates of genetic and phenotypic variation and covariation associated with the measured traits and those in the objective, and the relative economic values of traits in the objective, are used to derive the optimum combination or index of the phenotypic selection criteria. Thus, this classical selection index is sometimes also referred to as an economic index of genetic merit.

Because the overriding objective of a pork enterprise is the efficient production of quality lean meat, typical selection indices applied in the industry have reflected this purpose by relating to a breeding objective that includes genetic merit for leanness, growth rate and feed efficiency. Phenotypic measurements have traditionally included post-weaning growth (days of age at an ideal market weight or rate of gain from around 25 kg to market weight) and live-animal backfat thickness at market weight. In recent decades, advancements in ultrasound technology have facilitated routine inclusion of loin muscle dimensions as well as backfat thickness of the live animal at the end of the testing period. Less frequently, phenotypic records may include individual feed intake and efficiency during the post-weaning period and carcass information from relatives. As a result, the corresponding phenotypic index usually includes an ultrasonic measurement of backfat depth and loin muscle dimensions at market weight and average daily gain for the test period, but often does not include direct measurements of feed intake and efficiency.

Hence, while one of the advantages of the classical index is the ability to select on an objective trait through correlated phenotypic measurements, the method requires reasonably accurate estimates of genetic correlations among all traits and of relative economic values. Also, an underlying assumption is often made that the trait measured during performance testing in the nucleus population (e.g. average daily gain) is perfectly correlated with the same trait in the commercial system being targeted (i.e. it has the same genetic control in both populations).

**Interactions with environments**

The general structure of a livestock industry, with respect to the creation and dissemination of genetic improvement, can be described as a pyramid with nucleus, multiplier and commercial levels. Although selection may be practised at all levels, it is selection at the nucleus that determines the rate of permanent genetic improvement in the industry. Thus, the selection objectives addressed in nucleus herds must accurately reflect production goals at the commercial level.

Testing methods in nucleus populations are designed to provide unbiased estimates of genetic potential, and generally result in relatively uniform performance and greater heritability. But the testing environment under which candidates for selection are evaluated in nucleus herds is often different from the commercial production environment. For example, extraordinary effort is made through rigorous biosecurity methods to prevent the exposure of nucleus animals to swine pathogens. Boars tested in nucleus populations are typically penned in small groups or even individually, and may receive feed *ad libitum*, or some form of restricted or semi-restricted feeding, such as scheduled feedings to appetite. Commercial market pigs, on the other hand, are usually penned in larger groups and are at greater risk of exposure to pathogens, and, while most commercial producers in the USA allow pigs free access to feed, restricted feeding in the latter part of the post-weaning period is more common in European commercial units. Consequently, the benefits of greater heritability from reduced phenotypic variance in nucleus testing may be offset by genetic correlations between performance in the nucleus testing environment and in the commercial production setting of less than unity (Brascamp et al., 1985; Mulder and Bijma, 2005), resulting in less genetic progress. In a summary of literature results (Merks and de Vries, 2002), genetic correlations between environments, or between purebreds and crossbreds, were close to unity.
for some traits but significantly less for others (e.g. for backfat, 0.5 to 0.8). Lutaaya et al. (2001) estimated genetic correlations between each of two purebred lines (A and B) with their reciprocal crosses (C). Estimates of genetic correlations were different in the two crosses for both lifetime gain (A-C: 0.99; B-C: 0.62) and backfat (A-C: 0.32; B-C: 0.70). Two strains of commercial Durocs (P1 and P2) were used to estimate genetic correlations for performance traits in nucleus pigs and their crossbred progeny (C1 and C2, respectively) (Zumbach et al., 2007). Estimates for weight at last measurement, weight per day of age, backfat and loin depth were 0.53 ± 0.08, 0.60 ± 0.07, 0.83 ± 0.09 and 0.78 ± 0.05, respectively, based on P1-C1, and 0.80 ± 0.10, 0.79 ± 0.07, 0.89 ± 0.05 and 0.80 ± 0.08, respectively, based on P2-C2. Although differences in the method and time of measurement in nucleus versus crossbred animals may have affected some of the genetic correlation estimates in these studies, the genetic correlation for some traits in some populations is likely to be different enough from unity to warrant consideration of breeding methods (such as the inclusion of commercial data in nucleus selection schemes) to account for genotype × environmental interactions.

Variation in testing environments can also change the effective selection objective for a given set of measurements. For example, single-trait selection for rate of gain among animals that have ad libitum access to feed puts emphasis on appetite and may result in increased fatness (Woltmann et al., 1992, 1995). However, selection solely for gain among animals limited to a standard amount of feed may emphasize lean gain by identifying those animals that partition the allotted feed to the relatively efficient process of lean deposition (Webster, 1977).

Several models of metabolizable energy utilization and resulting tissue growth have been described for the pig (e.g. Whitemore and Fawcett, 1976; Whitemore, 1986; Moughan et al., 1987). Performance testing environments and selection schemes based on these models have been proposed as alternatives with which to address selection objectives associated with post-weaning performance.

**Models of growth and alternative strategies**

Whitemore (1986) described a relatively simple model based on a deductive approach in which the objective is to determine the causal forces that result in animal tissue growth. The model is best summarized in terms of daily rates of lean and fat tissue gain in response to daily feed intake (Fig. 14.1). The animal’s genetic potential for maximum rate of lean tissue

![Fig. 14.1. A hypothesis for the relationship between daily gains of lean and fat and the daily feed supply. From Whitemore (1986, p. 619) with permission from Journal of Animal Science.](image-url)
deposition is relatively constant beginning at an early age, and over nearly the entire range of ages and weights associated with the post-weaning production period. Therefore, as daily feed intake is increased, rate of lean gain increases linearly until it plateaus when the animal’s maximum potential is reached. The area to the left of the point at which lean gain plateaus may be thought of as nutritionally limited, and the area to the right, in which maximum lean gain is realized, as nutritionally unlimited.

During the nutritionally limited phase of growth, it is assumed that the animal will partition most of the available energy to lean growth, while maintaining some physiologically normal, but minimal, level of fat gain. In the nutritionally unlimited phase of growth, most of the feed consumed beyond what is needed for maximum lean gain is partitioned to fat deposition. Thus, when feed intake is allowed to increase into this range the animal begins to deposit fat rapidly, and grows less efficiently and eventually more slowly owing to the relatively greater energy requirements of fat deposition.

The principles described in this model are the basis for post-weaning production environments in which ad libitum feed intake is restricted, a practice common in European systems. The objective of restricting access to feed by pigs in finishing facilities is to limit daily feed intake to the minimum level at which maximum lean gain is realized, thereby avoiding unnecessary and costly fat gains. In some populations, feed intake in the latter stages of the post-weaning period can be restricted to as much as 75% of ad libitum without inhibiting lean growth potential (Fowler et al., 1976). Restricting post-weaning feed intake is sometimes referred to as scale feeding because systems are typically scaled so that each pig’s daily allotment of feed is increased as the post-weaning period progresses according to either time or body weight.

As depicted in graphs a and b of Fig. 14.1, a given amount of daily feed intake can result in nutritionally unlimited growth in one animal, but in nutritionally limited growth of another, depending on their relative genetic merit for maximum lean gain. Differences in lean gain potential between sexes of pigs and between some genetic strains of pigs have been well quantified (Campbell and Taverner, 1988; Eissen, 2000). Maximum lean gain potential is also assumed to vary among individuals of the same sex within a breed or genetic strain, and thereby becomes a potential component of the breeding programme objectives. Consequently, the principles outlined here become useful in the development of hypotheses regarding expected response to selection for performance traits, and the design of testing and selection schemes to improve genetic potential for lean gain and the efficiency of lean gain.

Fowler et al. (1976) attempted to model interactions between selection (genetic) objectives, nutritional environments during performance testing and nutrition provided during commercial production. The model of growth that they considered (Fig. 14.2) was similar to that of Whittemore (1986). Metabolizable energy intake not lost as heat, i.e. energy retained as

![Graphs a and b of Fig. 14.1](image)
product (PE), is partitioned either into skeletal muscle and essential accessory tissues, categorized together as lean tissue (LT), or into fat tissue. In an environment in which the pig has *ad libitum* access to feed, energy will be partitioned to the deposition of lean tissue until the maximum potential for lean growth rate is met, after which energy will be deposited as fat. The nutritional environments considered, both for performance testing and for the commercial production system, were scale feeding and *ad libitum* access to feed. In the context of performance testing, the time- or weight-based feeding scale would be designed to limit intake to a proportion of *ad libitum*, such that variation in appetite is not expressed but restraint of lean growth is minimized. Discussion of *ad libitum* access also applies, to some extent, to systems in which animals are allowed to eat freely for a set period of time twice each day, i.e. semi-*ad libitum*.

As an alternative to the classical selection index objective, expressed as a linear function of breeding values for traits deemed to have economic importance, Fowler et al. (1976) suggested a more biological definition of the objective based on physiological factors related to the market value of the pig. Because the primary product of the industry is lean pork and the largest costs to the system are those associated with feed and time, the two ‘biological’ objectives suggested were lean tissue growth rate (LTGR) and the feed required per unit of lean produced or lean tissue feed conversion (LTFC).

The simplest of these biological indices would use either LTGR or LTFC as the selection objective and as the selection criterion. In practice, selection for the objective LTGR would be based on the difference between estimates of lean content at the onset of the test and at the end of the test, in each case based on a function of live weight and ultrasonic measurements of subcutaneous fat depth and loin muscle area or depth. In a testing environment in which candidates are given free access to feed, selection for the objective LTFC would require the direct measurement or estimation of individual feed intake. When performance testing is conducted with a time-scale feeding system, in which pigs start the test at a standard weight, are tested for a standard length of time and are given an amount of feed based on time of test, variation in feed intake and in days of test is zero by design. Consequently, the biological indices LTGR and LTFC are perfectly correlated under this system.

Three primary testing/selection scenarios were considered by Fowler et al. (1976): Case 1, testing environment – *ad libitum*, objective – LTFC; Case 2, testing environment – scale feeding, objective – LTFC (LTGR); and Case 3, testing environment – *ad libitum*, objective – LTGR.

**Case 1:** Response to selection for LTFC is expected to be through a reduction in genetic potential for feed intake (appetite) and an increase in lean tissue gain (Fig. 14.2a). The relative emphasis on decreased feed intake will be greater when full expression of appetite is allowed, and the emphasis may shift almost entirely towards reduced feed intake if the estimate of lean gain is poor, or if a greater amount of consideration is given to leanness rather than lean gain (Fig. 14.2b). Decreased appetite may presently be desirable in some industry populations in which lean gain is not limited by intake. However, as lean gain potential nears the limit presented by *ad libitum* energy intake, selection emphasis would need to be reversed to improve LTFC (Fig. 14.2c).

**Case 2:** In this case, the testing environment of scale feeding is aimed at removing variation in feed intake without limiting the expression of lean gain potential. Improvement of LTFC is entirely through selection emphasis on LTGR, with genetic potential for feed intake unconstrained (Fig. 14.2c). As lean gain potential is increased, additional improvements in LTGR and LTFC under this scenario will eventually require increased emphasis on appetite. However, unlike Case 1, a reversal in the selection pressure on feed intake would not be necessary.

**Case 3:** This is typical of many on-farm testing situations in which individual feed intake is not measured. Feed intake is expected to remain unchanged under this scenario (Fig. 14.2d) until lean gain potential becomes limited, after which improvements in LTGR must be accompanied by increased feed intake (Fig. 14.2c). This method avoids the cost of measuring individual feed intake, but only addresses one component of LTFC.
Based on the principles of the model, the authors also determined expectations for the performance of animals resulting from the testing/selection scenarios when used in commercial systems that provided either ad libitum or restricted access to feed. They concluded that pigs from Case 1 would be most likely to produce acceptable carcasses in systems allowing ad libitum feed intake. Pigs from Cases 2 and 3 would be expected to perform better in systems with restricted intake than with ad libitum intake, and in a lean-based marketing system.

Another alternative strategy for improving the efficiency of lean growth is more directly aimed at the reduction of residual feed intake, i.e. the amount by which actual ad libitum feed intake differs from that expected to be required for maintenance and maximum deposition of lean tissue. To obtain an estimate of residual feed intake in pigs, a statistical model that includes covariate adjustments for body weight and gain as well as ultrasonic measurements of body composition is typically applied to ad libitum feed intake records (e.g. Cai et al., 2008). When these methods are effectively implemented, the resulting estimates of residual feed intake are phenotypically independent of the performance traits, although underlying genetic covariances may exist. Optimal weighting of genetic merit for feed intake may also be achieved by defining a selection objective that includes as traits the parameters from the linear plateau model for feed intake and protein deposition maxima, as well as for the marginal ratio of fat and protein deposition (de Vries and Kanis, 1992). In this approach, feed intake can have either positive or negative value in the selection objective depending on the relative positions in the population for feed intake and protein deposition capacities. Application of this approach has been hindered by the difficulty in identifying practical traits for the index and adequate estimates of genetic covariances between those measured traits and the traits describing the linear plateau model (Hermesch et al., 2003).

**Parameter estimates**

There have been many reported estimates of heritabilities ($h^2$) and genetic correlations ($r_g$) associated with post-weaning performance traits based on covariation among relatives. A summary of estimates from several of these studies is presented in Tables 14.1 and 14.2 for environments of ad libitum or semi-ad libitum feed intake and restricted feed intake, respectively. The estimates are listed in the order of the citations in each case, along with the range and a simple average. It is important to mention that, in addition to differences in the way feed was provided, there were also differences in the breeds studied and some variation in the methods implemented, such as the testing interval used and location of backfat measurements. For some traits ($h^2$) or trait combinations ($r_g$), most of the estimates were fairly similar across experiments. But, for those cases in which there were conflicting results, possible interactions with experimental differences or fixed effects within experiments will be discussed.

In Table 14.1 is a summary of estimates from animals either allowed ad libitum access to feed or fed by hand to appetite (semi-ad libitum). Heritability estimates for average daily gain (ADG) and backfat thickness (BF) under those conditions were mostly moderate and clearly indicate that these traits would be expected to respond to selection. Estimates for backfat thickness tended to be higher than those for ADG. Heritability estimates for daily feed intake (FI), residual feed intake (RFI) and feed conversion rate (feed/gain, or FCR) were also generally moderate and suggest that selection would be successful. There were fewer estimates for LTGR and LTFC reported, each
Table 14.1. Estimates of heritabilities and genetic correlations for pigs with *ad libitum* or semi-*ad libitum* access to feed.

<table>
<thead>
<tr>
<th>References&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Trait(s)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Estimates&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritabilities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19</td>
<td>ADG</td>
<td>0.41&lt;sup&gt;c&lt;/sup&gt;; 0.41&lt;sup&gt;c&lt;/sup&gt;; 0.20&lt;sup&gt;c&lt;/sup&gt;; 0.40; 0.07; 0.36; 0.41; 0.41, 0.03; 0.28; 0.49; 0.43; 0.37; 0.14; 0.17, 0.33; 0.39; 0.24; 0.19; 0.23, 0.25; 0.44; 0.42</td>
<td>0.03–0.49</td>
<td>0.29</td>
</tr>
<tr>
<td>1, 2, 3, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23</td>
<td>BF</td>
<td>0.66&lt;sup&gt;c&lt;/sup&gt;; 0.74&lt;sup&gt;c&lt;/sup&gt;; 0.56&lt;sup&gt;c&lt;/sup&gt;; 0.47; 0.60; 0.12; 0.34; 0.57; 0.59; 0.46; 0.25; 0.56; 0.50; 0.43; 0.36; 0.45; 0.68; 0.56; 0.38; 0.48; 0.48, 0.44</td>
<td>0.12–0.74</td>
<td>0.49</td>
</tr>
<tr>
<td>1, 2, 3, 6, 7, 8, 10, 11, 12, 15, 17, 18, 24, 25</td>
<td>FI</td>
<td>0.34&lt;sup&gt;c&lt;/sup&gt;; 0.41&lt;sup&gt;c&lt;/sup&gt;; 0.13&lt;sup&gt;c&lt;/sup&gt;; 0.62; 0.23; 0.17, 0.26; 0.16; 0.45; 0.29, 0.19; 0.23; 0.14, 0.24; 0.53; 0.29; 0.18</td>
<td>0.13–0.62</td>
<td>0.29</td>
</tr>
<tr>
<td>1, 2, 3, 6, 7, 8, 10, 11, 12, 14, 15, 17, 18</td>
<td>FCR</td>
<td>0.58&lt;sup&gt;c&lt;/sup&gt;; 0.48&lt;sup&gt;c&lt;/sup&gt;; 0.23&lt;sup&gt;c&lt;/sup&gt;; 0.27; 0.18; 0.41, 0.12; 0.51; 0.28; 0.19; 0.15; 0.20; 0.16; 0.35; 0.37; 0.27</td>
<td>0.12–0.58</td>
<td>0.30</td>
</tr>
<tr>
<td>15, 17, 18, 19</td>
<td>RFI</td>
<td>0.17, 0.11, 0.15; 0.10; 0.14, 0.24; 0.34, 0.41, 0.42; 0.29</td>
<td>0.10–0.42</td>
<td>0.24</td>
</tr>
<tr>
<td>11, 24, 25, 22</td>
<td>LTGR</td>
<td>0.39; 0.44, 0.44, 0.46; 0.59; 0.42; 0.48, 0.18; 0.24; 0.29; 0.27; 0.37, 0.37</td>
<td>0.25–0.46</td>
<td>0.39</td>
</tr>
<tr>
<td>11, 24, 25</td>
<td>LTFC</td>
<td>0.34; 0.35; 0.25</td>
<td>0.25–0.35</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Genetic correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2, 3, 6, 7, 8, 11, 12, 15, 26</td>
<td>ADG/FI</td>
<td>0.32&lt;sup&gt;c&lt;/sup&gt;; 0.50&lt;sup&gt;c&lt;/sup&gt;; 0.37&lt;sup&gt;c&lt;/sup&gt;; 0.83; 0.89; 0.69; 0.73; 0.80; 0.76, 0.41; 0.82; 0.89</td>
<td>0.32–0.89</td>
<td>0.67</td>
</tr>
<tr>
<td>1, 2, 3, 6, 11, 12, 15</td>
<td>BF/FI</td>
<td>0.38&lt;sup&gt;c&lt;/sup&gt;; 0.08; 0.39&lt;sup&gt;c&lt;/sup&gt;; 0.59; 0.42; 0.51; 0.24; 0.64</td>
<td>0.08–0.64</td>
<td>0.40</td>
</tr>
<tr>
<td>1, 2, 3, 6, 7, 8, 11, 12, 18, 26</td>
<td>ADG/FCR</td>
<td>−0.57&lt;sup&gt;c&lt;/sup&gt;; −0.71&lt;sup&gt;c&lt;/sup&gt;; −0.56&lt;sup&gt;c&lt;/sup&gt;; 0.34; −0.05; −0.69; −1.24; −0.28; −0.52; −0.52; −0.10; 0.78</td>
<td>−1.24–0.34</td>
<td>−0.37</td>
</tr>
<tr>
<td>1, 2, 3, 6, 11, 12, 15, 26</td>
<td>BF/FCR</td>
<td>0.10&lt;sup&gt;c&lt;/sup&gt;; 0.44&lt;sup&gt;c&lt;/sup&gt;; 0.24; 0.28; 0.36; 0.40; 0.33</td>
<td>0.10–0.44</td>
<td>0.31</td>
</tr>
<tr>
<td>1, 2, 3, 6, 9, 11, 12, 13, 15, 26</td>
<td>ADG/BF</td>
<td>0.02&lt;sup&gt;c&lt;/sup&gt;; −0.26&lt;sup&gt;c&lt;/sup&gt;; −0.06&lt;sup&gt;c&lt;/sup&gt;; 0.55; 0.35; 0.32; 0.26; −0.25; 0.09; −0.12; 0.37; 0.37</td>
<td>−0.26–0.55</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>References&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Trait(s)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Estimates&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Range</th>
<th>Average&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>15, 17, 18, 19</td>
<td>RFI/ADG</td>
<td>0.11, 0.17, 0.12, 0.18; −0.35, 0.00; 0.28,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23, 0.15; 0.06</td>
<td>−0.35−0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>15</td>
<td>RFI/BF</td>
<td>0.67, 0.22, 0.67, 0.20</td>
<td>0.20−0.67</td>
<td>0.44</td>
</tr>
<tr>
<td>17, 18, 19</td>
<td>RFI/FCR</td>
<td>0.70, 0.71; 0.67, 0.86, 0.84; (−0.69 FE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ADG/LTGR</td>
<td>0.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ADG/LTFC</td>
<td>−0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>BF/LTGR</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>BF/LTFC</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24, 25</td>
<td>FI/LTGR</td>
<td>0.23; 0.31</td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>24, 25</td>
<td>FI/LTFC</td>
<td>−0.45; −0.36</td>
<td></td>
<td>−0.41</td>
</tr>
<tr>
<td>24, 25</td>
<td>LTGR/LTFC</td>
<td>0.76; 0.87</td>
<td></td>
<td>0.82</td>
</tr>
</tbody>
</table>

<sup>a</sup>References: (1) Smith et al. (1962); (2) Smith and Ross (1965); (3) Standal and Vangen (1985); (4) Nordskog et al. (1944); (5) Fahmy and Bernard (1970); (6) McPhee et al. (1979); (7) Wyllie et al. (1979); (8) Cameron et al. (1988), estimates presented for Large White and Landrace populations, estimates for backfat (BF) are an average of those for shoulder, mid-back and loin measurements; (9) McPhee et al. (1988); (10) de Haer and de Vries (1993); (11) Mrode and Kennedy (1993); (12) Cameron and Curran (1994a), estimates presented for Large White and Landrace populations; (13) Hetzer and Miller (1972a), estimates presented for Duroc and Yorkshire populations; (14) Cameron and Curran (1995a), restriction 75% of ad libitum, all estimates presented are averaged across selection lines in Large White population, estimates presented for BF are an average of those for shoulder, mid-back and loin measurement sites; (15) Johnson et al. (1999), estimates presented are based on alternative models for estimation of residual feed intake (RFI) – (i) initial test age and weight and test average daily gain (ADG), (ii) model (i) plus BF, (iii) model (i) plus loin eye area, (iv) model (i) plus BF and loin eye area; (16) van Wijk et al. (2005); (17) Gilbert et al. (2007), estimates presented are for boar candidates and for castrated males and females; (18) Hoque et al. (2007), estimates presented are from alternative models for estimation of RFI predicted by nutritional requirements, or by phenotypic or genetic regressions on metabolic body weight and ADG; (19) Cai et al. (2008) (20) Gray et al. (1968); (21) Berruecos et al. (1970); (22) Chen et al. (2003); (23) Newcom et al. (2005), estimates presented are for carcass and ultrasound BF; (24) Cameron (1994); (25) Cameron and Curran (1994b); (26) Robison and Berruecos (1973).

<sup>b</sup>Traits: ADG, average daily gain; BF, backfat thickness; FI, daily feed intake; FCR, feed conversion ratio or feed/gain; RFI, residual feed intake; LTGR, lean tissue growth rate; LTFC, lean tissue feed conversion.

<sup>c</sup>Semi-ad libitum fed (e.g. hand-fed to appetite).

<sup>d</sup>LTFC based on an index of BF and FCR; increasing values of LTFC correspond to greater lean efficiency (lower FCR).
based on an ultrasonic prediction of lean content of the animal at the completion of the test, but the reports were similar and indicate that these traits are also moderately heritable.

Estimates of the genetic correlations of ADG and backfat with FI were all positive, and most were moderate to high. Estimates tended to be higher with ad libitum than with semi-ad libitum consumption, probably reflecting less than full expression of appetite in the latter setting. Although, based on these estimates, FI would be expected to increase with greater genetic merit for ADG, the genetic correlation between ADG and feed conversion was favourable in most of the reports cited. The exceptions (McPhee et al., 1979 (0.34); Wyllie et al., 1979 (−0.05)) were from studies in which pigs had continuous access to feed. Thus, the majority of reports suggest that the greater FI accompanying increased ADG would be more than offset by fewer days on feed. Estimates of the genetic correlation between backfat thickness and feed conversion revealed that selection for less backfat should also improve efficiency.

Estimates of the genetic correlation between RFI and either feed conversion or efficiency indicate that selection for less RFI can improve the efficiency of weight gain. The estimated correlations of RFI with other production traits suggest that RFI may also result in changes in ADG and backfat thickness, but that correlated response may also depend on the model used to estimate RFI (e.g. Johnson et al., 1999).

Reports of the correlation between ADG and backfat thickness were quite variable, ranging from moderate and favourable (−0.26) to moderate and unfavourable (0.55). Differences among experiments in the methods and technicians used for backfat measurement, and the typically large sampling errors associated with estimates of genetic correlations in experimental populations (Koots and Gibson, 1994), make interpretation of the variation among estimates difficult. However, it appears that some of the variation may be due to breed differences. For example, Cameron and Curran (1994a) reported that the correlation was 0.26 in Large White selection lines, but −0.23 in Landrace selection lines. Perhaps the genetic correlation between ADG and backfat in a population depends on how tightly coupled the traits are with FI versus the ability to partition energy intake to lean tissue growth. In the report by Cameron and Curran (1994b), the estimated genetic correlations of FI with ADG and BF were 0.76 and 0.39, respectively, in the Large White lines, and 0.41 and 0.13, respectively, in the Landrace lines. The extent to which free access to feed is available may determine this link between growth and backfat with FI. The average estimate of the correlation between ADG and backfat was 0.22 and −0.10 under ad libitum and semi-ad libitum intake, respectively.

Mrode and Kennedy (1993) included LTGR and LTFC (feed/lean), based on ultrasonic predictions of lean content of Yorkshire, Landrace and Duroc boars at the completion of test, in their study of FI, ADG and backfat. They reported that genetic merit for ADG and LTGR were closely correlated (0.96), but that the correlation between ADG and LTFC was small (−0.09). Conversely, the genetic correlation between backfat and LTGR was near zero (0.02), but the relationship between backfat and LTFC was moderate and positive (0.52). Cameron and Curran (1994a) and Cameron et al. (1994) reported genetic correlations of FI with LTGR and LTFC. Notice that LTGR and LTFC were based on ultrasonic backfat measurements and ADG and feed/gain, respectively, and that decreasing values for LTFC corresponded with more feed per unit of lean. These estimates indicate that, under an environment of ad libitum feed intake, LTGR and LTFC are positively but not perfectly correlated, and selection for improved LTFC would be expected to result in decreased genetic potential for FI, both results consistent with the suggestions of Fowler et al. (1976) as discussed in the previous section.

Estimates of genetic parameters for pigs restricted in their access to feed are available from evaluations of data from Danish test stations and from investigations of selection based on performance testing with scale feeding (Table 14.2). For studies in which parameters from environments of ad libitum and restricted intake were compared directly, the comparable estimate from Table 14.1 is presented in parentheses in Table 14.2.

As might be expected, phenotypic variation in ADG tended to be less with restricted
than with *ad libitum* intake (Cameron et al., 1988; McPhee et al., 1988; Cameron and Curran, 1995a). Restricted feeding tended to reduce phenotypic variation in FCR (Cameron et al., 1988) and backfat thickness (McPhee et al., 1988) in some studies, but there have also been reports that observable variation in feed conversion (Cameron and Curran, 1995a) was unaffected by feed restriction. Restriction of feed would be expected to have the greatest effects on phenotypic variances in backfat and feed conversion in populations in which *ad libitum* energy intake greatly exceeds lean growth potential. The associated variation in fat deposition and partitioning of energy to lean versus fat would be reduced as feed restriction moves energy intake closer to lean growth potential.

Effects of feeding environment on heritabilities (Table 14.2) are difficult to summarize from literature reports. Average heritability estimates reported for *ad libitum* or semi-*ad libitum* (Table 14.1) and restricted (Table 14.2) feeding environments were very similar for ADG and feed conversion; the average estimate for backfat tended to be greater with *ad libitum* or semi-*ad libitum* feeding than with restricted intake (0.49 versus 0.31). Among the studies in which (semi-)*ad libitum* and restricted intake were directly compared, Standal and Vangen (1985), Cameron et al. (1988) and McPhee et al. (1988) reported that heritability of ADG was greater with restricted feeding, but Cameron and Curran (1995a) reported a tendency for the opposite effect. Heritability estimates for backfat reported by Standal and Vangen (1985), Cameron et al. (1988) and Cameron and Curran (1995a) suggested a reduction due to restricted feeding, but McPhee et al. (1988) reported that feed

### Table 14.2. Estimates of heritabilities and genetic correlations for pigs with restricted feed intake.

<table>
<thead>
<tr>
<th>References&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Trait(s)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Estimates&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Range</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritabilities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>ADG</td>
<td>0.24; 0.39(0.20); 0.22, 0.14; 0.76(0.41), 0.14(0.03); 0.41(0.28); 0.30, 0.35; 0.17, 0.16; 0.26(0.39)</td>
<td>0.14–0.76</td>
<td>0.30</td>
</tr>
<tr>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>BF</td>
<td>0.47; 0.41(0.56); 0.26, 0.29; 0.06(0.60), 0.12; 0.60(0.34); 0.32, 0.36; 0.29, 0.28; 0.35(0.43)</td>
<td>0–0.60</td>
<td>0.31</td>
</tr>
<tr>
<td>2</td>
<td>FI</td>
<td>0.20(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 3, 4, 8</td>
<td>FCR</td>
<td>0.35(0.23); 0.23, 0.19; 0.56(0.41), 0.16(0.12); 0.24(0.20)</td>
<td>0.16–0.56</td>
<td>0.29</td>
</tr>
<tr>
<td>7</td>
<td>LTGR</td>
<td>0.34, 0.28</td>
<td></td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Genetic correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ADG/FI</td>
<td>0.28(0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BF/FI</td>
<td>0.29(0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 3, 4</td>
<td>ADG/FCR</td>
<td>−0.93(−0.86); −1.07, −1.02; −1.03(−0.69); −0.99(−1.24)</td>
<td>−1.07−0.93</td>
<td>−1.0</td>
</tr>
<tr>
<td>2, 3</td>
<td>BF/FCR</td>
<td>0.30(0.24); 0.16; 0.23</td>
<td>0.16–0.30</td>
<td>0.23</td>
</tr>
<tr>
<td>2, 3, 5, 6, 7</td>
<td>ADG/BF</td>
<td>−0.21(−0.06); −0.07, −0.31; −0.22(0.35), 0.08, −0.39; −0.10; −0.02</td>
<td>−0.39–0.08</td>
<td>−0.16</td>
</tr>
</tbody>
</table>

<sup>a</sup>References: (1) Lush (1936); (2) Standal and Vangen (1985); (3) Merks (1987), estimates presented are for Landrace and Yorkshire boars; (4) Cameron et al. (1988), estimates presented are for Large White and Landrace populations, backfat (BF) estimates are an average of those for shoulder, mid-back and loin measurements; (5) McPhee et al. (1988); (6) Gu et al. (1989), fed to a time-based scale, male and female averages presented; (7) Cameron et al. (1994), restriction to 75% of *ad libitum*, estimates presented are for Large White and Landrace populations; (8) Cameron and Curran (1995a), restriction to 75% of *ad libitum*, all estimates presented are averaged across selection lines in Large White population, BF estimates are an average of those for shoulder, mid-back and loin measurements.

<sup>b</sup>Traits: ADG, average daily gain; BF, backfat thickness; FI, daily feed intake; FCR, feed conversion ratio or feed/gain; LTGR, lean tissue growth rate.

<sup>c</sup>Estimates in parentheses are for (semi-)*ad libitum* intake, and are from Table 14.1.
restriction tended to increase heritability for backfat. In none of these studies was precision adequate to conclude that the heritabilities were truly different.

Most estimates of the genetic correlation between ADG and backfat in pigs with restricted access to feed were negative (Table 14.2), consistent with the trend noted when comparing estimates of the same correlation in pigs with semi-ad libitum versus ad libitum access (Table 14.1). McPhee et al. (1988) reported that the correlation changed from 0.35 with ad libitum to −0.22 with restricted intake, and Standal and Vangen (1985) estimated that the correlation was −0.06 and −0.21 for pigs with semi-ad libitum and restricted intake, respectively. Although Gu et al. (1989) reported values of 0.08 and −0.39 in boars and gilts, respectively, under an environment of scale feeding, they noted that boars were fed more liberally than gilts and attributed the difference in the estimates to the relatively greater restriction of intake in the gilts. As pointed out by Gu et al. (1989), the change in sign of this correlation is also consistent with the suggestions of Fowler et al. (1976) and Webster (1977). Greater merit for ADG with ad libitum intake is associated with greater daily consumption and fat deposition, but superior merit for ADG when intake is restricted is due to the ability to partition available energy to the relatively efficient process of lean growth.

The genetic correlation between ADG and feed conversion also seems to be affected by restricted feeding. Estimates of the correlation under (semi-) unlimited intake are moderate to high and negative. With restricted intake, all of the estimates cited are near −1.0. This result is also consistent with genetic models of growth. With ad libitum intake, increased genetic merit for ADG is associated with greater FI, but also fewer days on feed, resulting in a net decrease in feed conversion (feed/gain). When intake is restricted, greater merit for ADG must be through greater efficiency without increased FI, and is thus more closely correlated with feed conversion.

**Results from selection experiments**

The potential to exploit the additive genetic variation and covariation associated with post-weaning performance traits of pigs has been studied in numerous selection experiments. As suggested by moderate heritability estimates based on covariation among relatives (Table 14.1), the fundamental post-weaning traits of body growth and fatness will respond to selection. In an early study (Krider et al., 1946), divergent selection on weight for age was applied in a Hampshire population. After nine generations of selection (Baird et al., 1952; Craig et al., 1956), average weight at 180 days of age was 27.7 kg greater in the high line than in the low line. Subsequently, there have been several additional reports of significant direct responses to single-trait selection for weight at a given age (Kuhlers and Jungst, 1990, 1991a,b) or for post-weaning ADG (Rahnefeld, 1971; Rahnefeld and Garnett, 1976; Fredeen and Mikami, 1986a,b; Woltmann et al., 1992, 1995; Clutter et al., 1995b).

Selection for total body growth with free access to feed has been accompanied by changes in FI. High-line Hampshires consumed 0.64 kg/day more feed than those from the low line during a 72-day post-weaning test that followed the nine generations of selection (Baird et al., 1952), and Clutter and Buchanan (1998) reported a 5 genetic standard deviation difference in daily feed intake between lines after ten generations of divergent selection for post-weaning ADG. When the divergent lines in the latter study were restricted to a standard amount of feed intake, differences in ADG between the lines were not significant (Woltmann et al., 1992). Thus, most of the response in ADG was attributed to changes in feed intake. Direct responses to selection for growth rate have been great enough to offset the correlated changes in feed intake, resulting in improved total body feed efficiency from upward selection for growth (Rahnefeld, 1973) and greater efficiency in upward than in downward divergent lines (Baird et al., 1952; Clutter and Buchanan, 1998).

Changes in body composition in response to single-trait selection for growth have varied somewhat among experiments. Divergent selection for ADG resulted in greater backfat thickness at 105 kg body weight, but also greater LTGR, in the fast line than in the slow line (Woltmann et al., 1992, 1995). Selection
for greater weight at 200 days of age in a Duroc population was accompanied by an increase in backfat thickness, a decrease in percentage of muscle and no change in LTGR (Kuhlers and Jungst, 1992b), but selection on the same criterion in a Landrace herd resulted in increased LTGR without significant changes in backfat and percentage of muscle (Kuhlers and Jungst, 1993). Differences in correlated responses may be due to the specific criterion applied, the selection differential achieved, random drift and the effects of sampling. However, the change in body composition from selection for growth may also depend on the genetic potential for FI relative to lean growth in the base populations in which selection is applied or, as depicted by the estimates of the genetic correlations between ADG and backfat (BF) in Tables 14.1 and 14.2, the real degree of access to feed. In general, it appears that direct response to selection for rate of total body growth with complete access to feed results largely through increased FI, and may be associated with an increased rate of fat as well as lean tissue growth.

Divergent, controlled selection for backfat in Duroc and Yorkshire populations resulted in greater than 4 SD of total direct phenotypic response (Hetzer and Harvey, 1967; Hetzer and Miller, 1972a,b). Similar direct responses were also reported from two independent studies in which five generations of downward selection for backfat were practiced (Gray et al., 1968; Berruecos et al., 1970). Berruecos et al. (1970) reported that there was not a correlated change in pig growth associated with selection for decreased fat, but Hetzer and Miller (1972a) concluded that the correlated response in growth rate may vary by breed. In Durocs, ADG increased significantly in both divergent lines; in Yorkshires, ADG did not change significantly in the high fat line, but decreased in the low fat line. The variable correlated responses in ADG or backfat to direct selection on the other trait are consistent with the estimates of the genetic correlation between the two traits (Table 14.1).

Although estimates of heritability for feed/gain suggest that selection would be effective (Table 14.1), results from selection experiments have been mostly discouraging. Dickerson and Grimes (1947) reported that divergent selection for feed/gain resulted in significant direct response and a corresponding realized heritability of approximately 24%. But in two other studies in which downward selection for feed/gain was practiced (Jungst et al., 1981; Webb and King, 1983), realized heritabilities were not significantly different from zero. Results of selection for RFI have been reported from two independent studies (Gilbert et al., 2007; Cai et al., 2008). After three generations of divergent selection on phenotypic estimates of RFI in a Large White population (Gilbert et al., 2007), the estimated difference in phenotypic standard deviation units between the lines (high to low) was approximately 0.3 and 0.2 for RFI and FI, respectively. The lines did not differ for weight of backfat; growth rates were not reported, but the high line had between 0.2 and 0.25 greater feed/gain than the low line. Four generations of downward selection on estimated breeding values for RFI in a similar breed background (Yorkshire) resulted in reductions in RFI and FI of 96 and 165 g daily, respectively, relative to an unselected control line (Cai et al., 2008). The selected line grew more slowly (33 g daily) but was more efficient (1.36% greater gain/feed) and had less backfat (1.99 mm) than the control line.

Leymaster et al. (1979a,b) reported significant direct responses to selection based on ultrasonic estimates of either LTGR (weight of lean cuts at 160 days of age) or percentage of lean carcass (percentage of lean cuts at 81.6 kg live weight), but only selection on the estimate of LTGR was successful in improving both traits simultaneously. While selection for LTGR resulted in less carcass fat and greater ADG, response to selection for a greater percentage of lean carcass was accompanied by decreased ADG. These results are consistent with some of the estimates of the genetic correlation between ADG and backfat presented in Table 14.1, and with the correlated decrease in ADG reported by Berruecos et al. (1970) when selection was for less backfat thickness.

Several experiments have been conducted to test responses to selection for a classical index of ADG and backfat thickness (e.g. Vangen, 1979, 1980a,b; Cleveland et al., 1982, 1983a,b, 1988; Fredeen and Mikami, 1986a,b,c,d,e; McKay, 1990, 1992), and
selection was generally successful in improving the component traits of ADG and backfat thickness. In a rather comprehensively designed experiment (Fredeen and Mikami, 1986a), index selection improved ADG and backfat as much as each component was improved by single-trait selection in two contemporary lines. However, consistent with the generally antagonistic relationship between these components observed under ad libitum intake (Table 14.1), neither of the single-trait lines was as effective as the index at improving overall merit for the two traits combined. Only in the experiment reported by McKay (1990) was there little response in ADG, but that result could be explained by the relatively greater selection emphasis applied to backfat.

The most detailed evaluation of correlated changes in LTGR and LTFC in response to the index selection was reported by Cleveland et al. (1983a). In a study of barrows from the select and control lines at three levels of feed intake, the select line was superior in rate of protein growth and feed required per unit of edible lean at each of the intake levels. Although specific responses in LTGR and LTFC in the other studies were not reported, the simultaneous improvements in ADG, backfat and total body feed conversion reported by Vangen (1980a) and by Sather and Fredeen (1978) indicate that correlated improvements were made. Even with selection against backfat thickness, improvements in efficiency were achieved without significant reductions in genetic potential for FI.

Selection indexes that included measurements of feed conversion, in addition to ADG and backfat, have also been evaluated as a means of improving LTFC (McPhee, 1981; Ellis et al., 1988). Both McPhee (1981) and Ellis et al. (1988) reported improvements in backfat and feed conversion, but not significant improvements in ADG, resulting from index selection. Both studies also reported that reduced ad libitum FI accompanied selection. The lack of improvement in ADG was explained by the relatively great negative emphasis on feed conversion (feed/gain) and backfat, and the tendency for an antagonistic genetic correlation between backfat and ADG, as feeding becomes more liberal (Table 14.1). Even though LTGR was increased by selection on the indexes, the improvements in LTFC were primarily through decreased appetite.

There have been a limited number of reports of experiments designed to evaluate restricted feeding as a performance testing environment and to test the theories of Fowler et al. (1976) regarding the interactions of LTGR and LTFC with testing regime. McPhee et al. (1988) reported that selection on an index that estimated LTGR in a testing environment of scale feeding successfully improved ADG, backfat thickness, feed conversion and lean ham, and that responses were greater when offspring from the line were allowed ad libitum intake than when they were restricted. Appetite (ad libitum FI) was also increased by selection on the index under scale feeding. Although the study did not include a line in which selection on the index was practised with ad libitum FI, the authors speculated that response would be greater when testing is with scale feeding, because it was under that environment that estimated heritabilities were greater and the genetic correlation between ADG and backfat was favourable (Table 14.2).

A comprehensive study was conducted in Landrace and Large White populations to test divergent, controlled selection for either FI, LTGR or LTFC in pigs allowed ad libitum feed intake and LTGR in pigs with restricted feed intake (Cameron, 1994; Cameron and Curran, 1994a; Cameron et al., 1994). When the testing environment allowed ad libitum feed intake and performance was measured in pigs given free access to feed, the greatest amount of improvement in LTGR was from direct selection. In addition, correlated response in LTFC from selection on LTGR was similar to (in Landrace) or greater than (in Large White) direct response from selection on LTFC. Correlated response in FI was zero or slightly positive from selection on LTGR, but selection on LTFC caused a reduction in FI. In the same set of studies, progeny from the line in which selection was for LTGR in a testing environment of scale feeding and from the lines in which selection was with ad libitum consumption were compared in both feeding environments to determine the best overall selection strategy (Cameron and Curran, 1995a). When
pigs from the Large White lines were allowed *ad libitum* intake, ADG was greatest in the line selected for high LTGR under scale feeding, followed, respectively, by those selected under *ad libitum* intake for high LTGR and improved LTFC. The lines ranked the same for FI as for ADG, but were similar to another in backfat and feed conversion. Performance in the lines was more similar when progeny were scale fed, but pigs from the line selected for LTGR with scale feeding still tended to have faster gains and less backfat than those from the lines in which selection was for LTGR or LTFC with *ad libitum* access. In Landrace pigs, response in the same traits was similar in the lines selected either for high LTGR or improved LTFC with *ad libitum* intake or for high LTGR with scale feeding, regardless of whether progeny were allowed *ad libitum* or restricted feed intake.

Cameron and Curran (1995b) also reported results from an evaluation of carcass traits in which boar and gilt progeny from the three lines (LTGR-*ad libitum* feeding, LTFC-*ad libitum* feeding and LTGR-scale feeding) were tested under the feeding conditions of selection in their respective lines. Selection for LTGR-*ad libitum* increased LTGR, but did not change rate of fat growth. Rate of fat growth was reduced in the line selected for LTFC-*ad libitum*, but the change in LTGR was not significant. Selection for LTGR-scale significantly increased LTGR and decreased rate of fat growth.

**Factors underlying genetic variation in performance traits**

The widespread development and availability of laboratory assays for hormonal molecules in the pig have provided opportunities for investigations of the physiological factors underlying variation in performance traits. In addition to a goal of greater understanding of the underlying mechanisms of quantitative genetic variation, many of these studies were also aimed at the discovery of physiological indicator traits that could be used to enhance the accuracy of selection for objectives such as LTFC, for which conventional methods are expensive and difficult to implement (Blair et al., 1990). Despite the inherent challenges in effective design of these studies owing to the general complexity of growth and development, an extensive body of information related to physiological pathways underlying these traits has led to studies of the somatotrophic axis and putative regulators of appetite.

Large performance differences between breeds were the focus of some of the earliest studies in livestock. Buonomo et al. (1987) reported that the levels of circulating IGF1 (insulin-like growth factor 1) in a cross of large, fast-growing breeds (¼ Large White × ¼ Landrace × ½ Duroc) were 24 and 105% greater, respectively, than levels in smaller, slower growing Yucatan micro and Hanford miniature pigs. While results like these probably reveal factors contributing to the observed differences in growth and development, breed comparisons are confounded with all genetic selection and drift that has occurred in the history of the populations. Experimental lines in which selection has been only for traits associated with growth and efficiency are free of confounding with other selection effects, and may reveal more directly indicator traits that can be used for within-line selection.

Lines selected divergently for post-weaning ADG (Woltmann et al., 1992) were used as a basis for a series of studies to test the potential roles of underlying physiological factors. Selection had resulted in 46% greater ADG, 60% greater average FI and 13% greater backfat in the fast than in the slow line (Clutter et al., 1998a). When serial blood samples were collected from gilts representing the fast and slow lines at approximately 55 kg body weight, there were no significant differences in mean GH (growth hormone) concentration or in pulsatility of GH secretion, but there was a greater mean concentration of IGF1 in the fast than in the slow line (Clutter et al., 1995b). In addition to greater circulating IGF1, there was less measured activity of IGF-binding proteins (IGFBP-2 and IGFBP-3) in pigs from the fast line, suggesting a greater availability of active IGF1.

Based on results of scale-feeding studies of these divergent lines in which most of the response in growth rate could be attributed to changes in daily feed intake (Woltmann et al., 1992), the role of cholecystokinin-8 (CCK) as
a putative satiety factor was the focus of additional experiments. When pigs from the lines were trained to eat twice a day to appetite and serial blood samples were collected during a 2-h feeding period that followed overnight feed deprivation, CCK per unit of feed intake was significantly greater in pigs from the slow line than in pigs from the fast line (Clutter et al., 1998a). In a second experiment, pigs sampled from the lines were infused with synthetic CCK during free access to feed (Clutter et al., 1995a). Feed intake of pigs from the slow line was reduced more than feed intake of pigs from the fast line in response to the synthetic CCK, suggesting a greater sensitivity to this putative satiety signal.

Genetic lines of pigs resulting from long-term divergent selection for components of LTFC have also been used as a basis for growth physiology studies (Cameron and Curran, 1994b; Cameron et al., 1994). Serum IGF1 concentrations were compared in lines created by allowing pigs ad libitum access to feed and selected divergently for either FI or LTGR (Cameron et al., 2001). As expected, pigs from the line selected for high FI ate more than those from the line selected for low FI, and also grew faster, had more backfat and were less efficient. Pigs from the line selected for high LTGR grew faster than those from the line selected for low LTGR, but did not eat more, had less backfat and were more efficient. A significant difference was detected in mean IGF1 concentration at 6 weeks of age in the FI lines (158 versus 104 µg/l in the high and low lines), but the difference declined in measurements made at 30 and 90 kg. There was not a difference in mean IGF1 between the LTGR lines at 6 weeks of age, but the high line had greater mean IGF1 than the low line by 30 kg of body weight (198 versus 153 µg/l in the high and low lines).

Because of their respective putative roles in appetite stimulation and inhibition, serum levels of neuropeptide Y (NPY) and leptin were measured in the FI and LTGR lines, as well as in lines created through providing free access to feed and selecting divergently for LTFC (Cameron et al., 2000, 2003). The FI lines did not differ in NPY concentration, but the line selected for high FI had a greater level of serum leptin than the line selected for low FI. The corresponding estimates of line means for performance traits were consistent with the phenotypic means reported previously by Cameron et al. (2001), in that upward selection in FI resulted in greater feed intake, more predicted lipid weight and less efficiency than downward selection. Serum concentration of NPY was greater in the high LTGR line than in the low LTGR line, but leptin levels did not differ significantly between the lines. Corresponding estimates of line mean differences were again similar to phenotypic means reported by Cameron et al. (2001), in that the high line grew faster, had more predicted protein weight and less predicted lipid weight and was more efficient than the low line; but in contrast to the earlier report, the high line was estimated to eat more per day than the low line. A lesser level of serum leptin in the high than in the low LTFC line was associated with less predicted lipid and more predicted protein and greater efficiency. Serum NPY was greater in the high than in the low LTFC line, but daily feed intake did not differ significantly between the lines. Correlations of leptin with FI and fatness indicated that the responses in leptin were largely due to greater fat deposition rather than to greater FI per se and, overall, the results suggest that response to selection for greater FI is more likely due in part to less sensitivity to leptin rather than less leptin production. Serum concentrations of NPY were not strongly associated with feed intake.

Correlated response in the circulating level of IGF1 at a target age of 35 days (referred to as juvenile IGF1) as a result of controlled selection for RFI was reported by Bunter et al. (2010). Serum IGF1 was measured in a total of 2570 animals sampled from the selected and control lines beginning at generation two. Genetic correlations were estimated for the population between IGF1 and the production traits, and genetic trends in the selected and control lines through generation five were used to estimate realized correlated responses. Estimated genetic correlations were significant between juvenile IGF1 and backfat, loin muscle area, RFI and feed/gain (0.52 ± 0.11, −0.35 ± 0.12, 0.63 ± 0.15 and 0.78 ± 0.14, respectively), but not different from zero between IGF1 and lifetime ADG, test ADG and FI.
(0.06 ± 0.14, −0.19 ± 0.14 and 0.26 ± 0.17, respectively). Correlated response in IGF1 to downward selection on RFI reached −1.07 SD by generation five, corresponding to a realized genetic correlation of 0.84.

Across the range of these experiments, significant results emerge that may begin to reveal some of the factors that underlie genetic variation and contribute to response to selection for performance traits in pigs. Each of the studies of IGF1 demonstrated that selection had affected some aspect of the IGF1 pathway, but the impact of selection observed probably depends on the timing of measurement of IGF1 and the selection emphasis on the components of post-weaning performance. Divergent selection for post-weaning ADG appears to have altered the production of IGF1 and associated binding proteins by several weeks after weaning, so that a greater amount of active IGF1 may be available in pigs selected for fast growth than in pigs selected for slow growth (Clutter et al., 1995b). Greater levels of circulating IGF1 were also observed beginning at 6 weeks of age in a line selected for greater feed intake than in a line selected for less feed intake (Cameron et al., 2001). A similar positive genetic relationship of IGF1 with feed intake and feed/gain was observed when selection was for less RFI, but the estimated genetic relationships between IGF1 and ADG were not significant even though the selection response in RFI was accompanied by a slight decrease in growth rate (Bunter et al., 2010).

The likely complexity in the physiological mechanisms that determine overall post-weaning performance begins to become even more apparent when considering that mean IGF1 at 30 kg body weight was greater in the divergent line that was selected for high LGA (lean growth rate with ad libitum access to feed), and also superior for total body gain and efficiency but not different from the low-LGA line for FI (Cameron et al., 2001).

Selection that resulted in relatively greater appetite may have both increased the amount of CCK circulating in response to feed intake and the sensitivity to the CCK satiety signal, the latter suggesting a change in CCK receptor characteristics (Clutter et al., 1995a, 1998a). Response to selection for FI and the correlated response in fat deposition were not due to insufficient leptin; therefore a decline in sensitivity to leptin may have contributed (Cameron et al., 2000). There was no evidence that genetic divergence in energy intake was due to increased appetite via greater circulating NPY (Cameron et al., 2003).

Each of the selection lines discussed here originated from a single base population and was subject to only a single selection criterion (post-weaning ADG, FI, lean tissue gain, lean efficiency or RFI) after establishment of the line. Consequently, physiological differences between the lines can be attributed to founder effects that may have occurred by chance in the derivation of the lines from the base population, random drift in gene frequency due to non-infinite line size and chance, or the selection applied. Although the effects of these factors are confounded in the comparisons made, there was clearly intense single-trait selection achieved in these experiments, and it is likely that the physiological differences observed are associated with the divergent genetic potential of the lines for FI, growth and tissue deposition.

Developments in tools related to molecular biology over the last two decades have allowed the extension of these types of studies to exploration for the genes or genomic regions that underlie variation in performance traits. Two distinct, but parallel, approaches that have been used in this exploration are comprehensive scans of the genome and the study of specific candidate genes.

A large number of genome scans for quantitative trait loci (QTLs) contributing to variation in performance traits in pigs have been reported, beginning with a few coarse scans using microsatellite markers in crosses of divergent lines, and accelerating in number as dense marker arrays have become available. The Pig QTL data base (PigQTLdb, at http://www.genome.iastate.edu/cgi-bin/QTLdb/SS/index) contains information for porcine QTLs reported in the literature. In Release 10 of the Pig QTLdb (30 December 2009) there are 429 QTLs for body weight or gain, 590 QTLs for leanness traits of the live pig or overall carcass composition and 65 QTLs for feed intake or feed conversion.

Some studies of candidate genes in pigs have been based on knowledge of how
physiological pathways may contribute to variation in performance traits. A polymorphism in the porcine gene encoding the cholecystokinin-A receptor (CCKAR; Clutter et al., 1998b) was targeted because of evidence that variation in sensitivity to the CCK satiety signal may contribute to variation in feed intake and growth in pigs (Clutter et al., 1995a, 1998a). Houston et al. (2006) conducted association studies of the polymorphism in CCKAR reported by Clutter et al. (1998b) and of a second polymorphism in the 5′ untranslated region (5′ UTR) of CCKAR that they had determined disrupts binding of the YY1 transcription factor. The association analyses were based on data from a Meishan × Large White F2 cross, from the divergent selection lines of Large White reported by Cameron and Curran (1994b) and from segregating commercial lines. Although the two markers were in strong linkage disequilibrium and were generally associated with performance traits across these populations, the 5′-UTR polymorphism was strongly and consistently associated with both feed intake and growth rate.

Other candidate gene studies in pigs have been initiated because of mutations reported in model species with large effects on traits related to pork performance traits. Kim et al. (2000) began investigation of the melanocortin-4 receptor gene (MC4R) in pigs after reports of the effects of a mutation in Mc4r in mice that resulted in large increases in feed intake and obesity, and demonstrated that a missense mutation (p.Asp298Asn) was associated with fatness and daily gain. At a more basic level, p.Asp298Asn displayed stimulatory effects on cAMP production (Kim et al., 2004), providing evidence that it may directly cause variation in these traits. Subsequent studies of p.Asp298Asn and other variants within porcine MC4R across multiple lines of pigs have revealed generally consistent associations with body composition and growth, but also that specific mutations within the gene may be more or less strongly associated with performance traits in individual lines (Fan et al., 2009).

The MC4R polymorphism (Kim et al., 2000) and the 5′-UTR polymorphism in CCKAR (Houston et al., 2006) are each examples of candidate genes being used in commercial pork production as tools by which to increase accuracy of selection for post-weaning performance traits. There have been many other reports of significant associations between markers in candidate genes and performance traits in pigs, and discoveries from comprehensive genome scans have begun to be developed and implemented in marker-assisted selection programmes designed to account for significant proportions of the additive variance in performance traits (Dekkers, 2004). As sequencing of the porcine genome is completed, the resolution of our knowledge of the basic factors contributing to genetic variation in performance traits will continue to improve and, along with it, the potential for increasingly powerful tools for selection and improvement.

**Implications**

Great opportunities exist for the genetic improvement of post-weaning performance through within-line selection. Various selection experiments have shown that predicted responses in components of LTGR and LTFC based on models of growth (Fowler et al., 1976) are remarkably accurate. In testing environments of ad libitum intake, selection for LTFC was successful primarily through reduction in appetite accompanied by a lesser rate of fat growth, rather than through increased LTGR. Selection for LTGR by testing with ad libitum intake improved LTGR and LTFC without reducing appetite, but did not reduce the rate of fat growth. Implementation of scale-fed testing to select for LTGR seems to combine the best features of selection for LTGR and LTFC under ad libitum intake, and may result in market pigs that are the most valuable under either commercial feeding environment (Fowler et al., 1976; Cameron and Curran, 1995b).

The potential benefit of any physiological factor as an indicator trait in selection for post-weaning performance is a function of the heritability of the indicator and the genetic correlation with the component(s) of post-weaning performance to be improved, relative to the expected response to selection without the indicator trait, i.e. the marginal improvement in accuracy. Ultimately, decisions to
implement physiological indicators in commercial selection programmes will be determined by not only the confirmed genetic associations of those indicators with components of breeding objectives for post-weaning performance (e.g. components of LTFC) within targeted populations, but also analyses of the cost of sample collection and assay completion relative to the marginal increase in accuracy and response. In addition, the possible opportunity cost or benefit of using physiological markers versus emerging genomics technologies to increase accuracy of selection will need to be assessed. Detailed updates and discussion of genomic approaches to selection in the pig are presented in Chapters 8 and 16.

**Genetic Relationships of Growth Performance with Reproduction**

To determine the optimum emphasis on lean gain and efficiency in selection objectives for each line contributing to a commercial breeding system, the genetic relationships (correlations) of the components of LTGR and LTFC (i.e. ADG, body composition and feed conversion) with traits of reproductive performance must be known. Expectations for changes in reproductive performance from selection for market animal performance can be derived from estimates of genetic correlations between the relevant traits and from estimates of correlated responses in selection experiments.

**Estimates of genetic correlations**

Two publications in 1981 included summaries of estimated genetic correlations between post-weaning performance and traits of puberty or reproduction based on covariances among relatives (Hutchens and Hintz, 1981; Johansson, 1981). A review by Brien (1986) also included a discussion of literature estimates of genetic relationships between growth and reproduction in the pig and other mammalian species. Many of the early reported estimates of genetic correlations that follow were included in more detail in those summaries. Vogt et al. (1963) first reported that the genetic correlation between post-weaning growth rate and litter size was near zero, but in subsequent studies in which first and second parities were considered separately (Morris, 1975; Johansson, 1981), the correlation tended to be positive for second litters. Conversely, in a study of more modern populations, Holm et al. (2004) reported that, in Norwegian Landrace pigs born between January 1990 and January 2000, there was an unfavourable genetic correlation between age at 100 kg and number born alive at first and second parities (0.60 ± 0.05 and 0.42 ± 0.06, respectively). Although estimated genetic correlations between production and reproduction traits were generally small in a study of Czech Landrace and the Slovak White Meaty breeds (Pekoviova et al., 2002), the estimate of greatest magnitude suggested an unfavourable relationship between ADG and number born alive in parity two and greater (−0.12 and −0.28 in the two breeds, respectively). Serenius et al. (2004) reported in Finnish Landrace and Large White populations a tendency for an unfavourable genetic correlation of ADG with total number born (−0.16 ± 0.13), but favourable relationships of ADG with number of still-born piglets (−0.25 ± 0.15) and piglet loss during the suckling period (−0.43 ± 0.16). Young et al. (1977) were unable to estimate genetic correlations between ADG and litter size because of a negative estimate of sire variance. However, they reported a positive genetic correlation between ADG and ovulation rate, and a tendency for a negative relationship between ADG and embryo survival (number of ovulations/number of embryos). These results for ADG and ovulation rate seem to be consistent with the results of selection for ovulation rate in the pig (Rosendo et al., 2007) and the positive genetic correlation between body size and ovulation rate observed in other species (Brien, 1986). Morris (1975) reported an unfavourable genetic correlation between killing-out percentage and litter size, but genetic correlations between backfat thickness and first and second parity litter sizes that were (negative) favourable and generally small to moderate in magnitude. Bereskin (1984) also reported a favourable correlation between backfat and litter size, as well as a favourable genetic correlation between loin muscle area and litter size. However, the
associated standard errors reported by Morris (1975) and Bereskin (1984) indicate that the estimates were not significantly different from zero. Johansson and Kennedy (1983) reported some tendency for unfavourable genetic correlations between backfat and litter size, but standard errors were not included and the estimates of unfavourable relationships ($r_g = 0.13$ to 0.22 for BF and number born alive) were probably not different from zero. More recent estimates of these genetic correlations have consistently revealed at least a tendency for an unfavourable genetic relationship of leanness with litter traits. Holm et al. (2004) reported that backfat thickness was uncorrelated with number born alive, but estimated that the genetic correlations of lean meat content with number born alive at the first and second parities were $-0.12 \pm 0.07$ and $-0.24 \pm 0.09$, respectively. Estimated genetic correlations of fat percentage with total number born and number of stillborns were $0.19 \pm 0.11$ and $-0.34 \pm 0.13$, respectively, in Finnish Large White pigs (Serenius et al., 2004); estimates of genetic correlations of lean percentage with total number born and number of stillborns were $0.19 \pm 0.11$ and $-0.34 \pm 0.13$, respectively, in Finnish Landrace and Large White (0.20 ± 0.13; Serenius et al., 2004), but there was a tendency for a desirable association between feed/gain and number of piglets lost during suckling (0.42 ± 0.17).

In attempts to draw overall conclusions regarding the genetic relationship between post-weaning performance and litter characteristics, some researchers have estimated genetic correlations between indexes of the relevant traits. Bereskin (1984) reported that genetic correlations of a performance index of ADG and backfat thickness with two indexes of sow productivity (litter sizes and weights) were not different from zero (average estimate $= 0.07$), indicating that selection for lean gain would not adversely affect litter characteristics. Those results support the findings of Morris (1975) that litter sizes and litter weight were not genetically correlated with a total point score that included ADG, feed conversion and carcass traits (estimates ranged from −0.04 to 0.02). However, in a more recent evaluation of field data from four US swine breeds, Chen et al. (2003) reported that, along with an unfavourable genetic correlation between number born alive and backfat (average estimate = 0.188), the genetic correlation between number born alive and lean gain was also slightly, but significantly, unfavourable (average estimate = −0.095). Similarly, Arango et al. (2005) concluded from parameter estimates in commercial Large White sows that intense selection for both faster growth and greater leanness would result in increased pre-weaning mortality in piglets born to first-parity sows.

It is more difficult to find estimates of the genetic relationships between performance traits and traits of reproduction that are not routinely measured. Favourable genetic correlations between ADG and age of puberty have been reported by Young et al. (1978) and by Hutchens et al. (1981). Results from these same studies indicated that the genetic correlation between backfat thickness and age of puberty was either favourable or not different from zero, suggesting that selection for lean gain would not be detrimental to age of puberty. Rydhmer et al. (1994) reported a tendency for
an unfavourable genetic correlation between lean growth and age of puberty; their estimates of genetic correlations from a population of Swedish Yorkshires selected for lean tissue growth rate were 0.40 (± 0.10) between lean percentage and age of puberty, and 0.21 (± 0.23) between lean gain and age of puberty. But Holm et al. (2004) estimated favourable genetic correlations of age at 100 kg with age at first service (0.68 ± 0.02), interval from first weaning to next service (0.16 ± 0.07) and interval from second weaning to next service (0.20 ± 0.10), and correlations of approximately zero between lean meat content and these same breeding traits.

Correlated responses to selection

Many studies of selection for post-weaning performance have included measurements of those traits of reproduction that are routinely available in the production system, primarily litter sizes and litter weights from birth to weaning. In a few studies, traits such as age of puberty, breeding performance and sow weight changes have also been recorded. Unfortunately, data on reproductive performance in experimental populations selected for traits of market pig performance typically have several limitations:

1. Data that are typically only from gilts and first litters. As noted earlier, correlations between performance and reproduction may differ by parity. If selection for lean gain and efficiency affects reproductive ability, it is possible that the effects are manifested after the gilt weans her first litter.

2. Measurements of correlated responses are often based on the reproduction of selected females, and consequently are subject to bias.

3. Many important traits of reproduction are quite variable and inherently difficult to measure with precision. Thus, data collected from the relatively small number of females maintaining a given selection line in each generation are usually not adequate to allow meaningful conclusions.

4. Most experimental selection lines of pigs are not replicated and are of relatively small effective size. Therefore, measurements of correlated (and direct) responses to selection are likely to be confounded with random drift in gene frequencies.

Among experiments in which selection was for some measurement of growth, significant changes in litter weights or pig weaning weights reflecting changes in the pig’s direct genetic merit for pre-weaning growth and/or the maternal ability of the sow were only observed by Craig et al. (1956) with selection for weight at a post-weaning age, and by Garnett and Rahnefeld (1976) with selection for ADG. In both cases, a positive impact of upward selection for growth was indicated. Although trends in litter sizes were not significant in most of the studies in which selection was for growth, a positive effect of selection for weight at 70 days of age on 21-day litter size was reported by Kuhlers and Jungst (1992a). Conversely, selection for weight at 200 days of age in Durocs decreased total litter size at birth (Kuhlers and Jungst, 1992b), and pre-weaning mortality increased due to selection for post-weaning ADG (Garnett and Rahnefeld, 1976). Number born alive did not change significantly in the former study (Kuhlers and Jungst, 1992b).

Litter traits did not change significantly over ten generations of divergent selection for ADG (Clutter and Buchanan, 1998), but correlated response in appetite seems to be reflected in weight changes of gilts during gestation and first lactation. Gilts from the line selected for slow growth and expressing relatively less appetite weighed less at farrowing, but lost a significantly greater amount of weight during lactation than gilts from the line selected for fast growth. Although size and weight of the first litter were not affected, weight loss in the female was great enough to suggest that future reproduction might be inhibited. Unfortunately, females were marketed after the first litter, hence re-breeding performance and subsequent litter characteristics were not measured.

For studies in which selection was for divergent or decreased backfat thickness, only Berruecos et al. (1970) reported significant changes in reproduction. Selection for increased leanness resulted in a significant decline in litter size at birth and weaning but, surprisingly, by 130 days of age the difference in litter size was nearly gone. Although, based
on the difference of sire and dam components of variance for feed conversion, Dickerson and Grimes (1947) concluded that females with greater efficiency for total body weight gain expressed poorer nursing ability, correlated responses in reproduction were not reported from studies in which single-trait selection for feed conversion was practised.

As with most of the experiments implementing single-trait selection, many of the studies of direct or alternative criteria to improve LTGR or LTFC revealed no significant trends in traits of reproduction. However, some exceptions suggest the potential for antagonistic relationships. Although the regressions of first parity litter sizes on generation number in a line selected for weight of lean cuts at 160 days (LTGR) were negative but not different from zero (DeNise et al., 1983), second parity litter sizes and corresponding litter weights were significantly decreased by selection. Second parity litter characteristics were also significantly decreased by selection for percentage of lean cuts at approximately 80 kg body weight. Kerr and Cameron (1995) did not detect correlated responses in litter traits to selection for either LTGR with ad libitum intake or LTGR with scale feeding, but litter sizes and weights at birth and weaning were significantly decreased by selection for LTFC with ad libitum intake. This result is consistent with their estimate of positive genetic correlation between litter size traits and FI (Kerr and Cameron, 1996), given the decrease in FI when selection was for LTFC with ad libitum intake.

Knowledge of the genetic relationships between performance and reproduction traits may also be gained by measuring correlated responses in performance traits to selection for reproduction. Petry et al. (2004) summarized correlated responses in a line selected for 11 generations on an index of ovulation rate and embryonic survival, then selected through to generation 19 for number of fully formed piglets born (Nebraska index line). Direct genetic effects for backfat at 88.2 kg and days to 88.2 kg did not differ between the index and control lines, and even though index line pigs had a loin muscle area that was 1.58 cm² less than control line pigs, they concluded that the methods of selection for increased litter size in this line had little impact on performance traits. Two other studies were of lines that were derived at earlier points in time from the Nebraska index and control lines. Ruiz-Flores and Johnson (2001) reported that an increase in backfat accompanied eight generations of two-stage selection for ovulation rate and number of fully formed piglets born; the regressions of mean breeding value for backfat deviated from the control were 0.14 ± 0.06 and 0.34 ± 0.06 mm thickness/generation, respectively, for selection lines derived from the Nebraska index and control lines. Correlated response in weight at 178 days of age was different from zero only in the select line derived from the Nebraska index line (1.31 ± 0.20 kg/generation). Holl and Robison (2003) summarized results from nine generations of controlled selection for number born alive using lines derived from the Nebraska control line. Correlated response in breeding value for backfat was not different from zero, but for days to 104 kg it was unfavourable (0.93 ± 0.21 days/generation). In an independent study, Estany et al. (2002) reported selection for number born alive that generated a genetic superiority in the select line of 0.46 live piglets relative to an unselected control. As a correlated result, the select line had a greater genetic mean for backfat at 165 days of age than the control line (1.26 ± 0.23 mm from a midline ultrasound). Overall growth and feed efficiency did not differ, but the patterns varied during the post-weaning period (the select line grew faster and was more efficient early, but grew more slowly and was less efficient late in the period).

Implications

Available estimates of genetic correlations and correlated responses to experimental selection present in total a complex and somewhat inconsistent picture of the genetic relationships between performance and reproduction traits. There are several technical factors that probably contribute to this variability in results. Estimation of genetic correlations from covariances among relatives is typically associated with large standard errors owing to insufficient sample size, although some more recent studies have made use of field records to increase precision. The selection experiments discussed here encompass a variety of criteria, objectives
and methods. Even among those studies of a classical index including ADG and backfat, selection varied owing to differences in specific index weightings, selection differentials achieved and the population in which selection was applied. Based on these factors alone, it is not surprising that results for correlated changes in reproduction traits are mixed. In addition, it is difficult to generate the number of observations necessary for powerful tests of correlated responses in these inherently variable traits of reproduction. This lack of precision adds to the inconsistency in results and can lead to precarious interpretation of non-significant trends.

Those differences and challenges associated with experimental designs undoubtedly contribute to the fact that, across much of the entire range of seven decades of research results reviewed, there are reports of favourable or neutral (e.g. Vogt et al. 1963; Morris, 1975; Bereskin, 1984; Petry et al., 2004) and of unfavourable (e.g. Dickerson, 1947; Kerr and Cameron, 1995; Imboonta et al., 2007) expected responses in reproduction to selection for post-weaning performance traits. Even so, it should be possible to draw some conclusions regarding the genetic relationship of these trait groups as it relates to breeding strategies.

Several studies reported significant responses in components of LTGR or LTFC with small or non-significant changes observed in traits of reproduction, but some clearly unfavourable estimates of genetic correlations or correlated responses were reported. While some of the antagonism reported was between growth rate and reproduction (Pekoviova et al., 2002; Holm et al., 2004; Serenius et al., 2004), the potential for adverse effects seems most likely when selection is for increased leanness (Berruecos et al., 1970; DeNise et al., 1983; Serenius et al., 2004; Imboonta et al., 2007), LTGR (Chen et al., 2003; Arango et al., 2005) or LTFC (Kerr and Cameron, 1995). These adverse effects tended to be more significant for second parity litters (DeNise et al., 1983; Holm et al., 2004), consistent with relatively recent reports that similar unfavourable associations may exist with sow lifetime performance (Clutter, 2009). In addition, potentially antagonistic associations between performance and reproduction traits were somewhat more frequent in studies of more modern populations of pigs. It is possible that genetic correlations between components of LTFC and reproduction are not linear, and that antagonistic relationships may develop as performance reaches new thresholds. This may be particularly true for selection approaches that take body fat to extremely low levels and/or depress ad libitum energy intake. Thus, while the literature does not provide conclusive evidence that selection for performance traits will have detrimental effects on reproduction in all populations, antagonistic relationships or negative effects in some studies are causes for concern. Breeding strategies and objectives for maternal lines should be based on covariances between components of LTFC and, at least, litter traits, estimated specifically for the targeted population.

**Non-additive Genetic Effects and Breeding Systems**

The most efficient commercial production of pork occurs in a breeding system that optimizes selection on additive genetic effects for market pig performance and for reproduction in the contributing lines, and maximizes exploitation of non-additive genetic effects (dominance and epistasis) through heterosis. To determine the efficiency of a breeding system from an overall industry perspective, however, the cost of maintaining the pure-line and multiplier herds that supply the commercial system must also be considered. In this section, estimates of maximum heterosis available for traits of post-weaning performance are briefly reviewed, and effective breeding systems for the pork industry are discussed.

Heterosis estimates for post-weaning performance traits are included in early reviews of the extensive research projects that became the basis for the initiation of commercial crossbreeding in pigs (e.g. Sellier, 1976; Johnson, 1980, 1981; Buchanan, 1987), as well as reported from additional more recent studies of crosses in experimental herds (e.g. McLaren et al., 1987; Baas et al., 1992; Cassady et al., 2002). Individual, maternal and paternal heterosis, the difference in performance due to a crossbred versus purebred individual, dam or sire, respectively,
has been considered. Consistent with theoretical expectations, heterotic effects are greatest for those traits that have relatively low heritability, such as litter sizes and pre-weaning litter weights. Accordingly, for components of post-weaning performance that are generally moderately heritable (see Tables 14.1 and 14.2), relatively moderate to small proportions of heterosis have been reported.

Average estimates of individual heterosis for post-weaning ADG reported in the early reviews ranged from 6.0% (Sellier, 1976) to 8.8% (Johnson, 1981) of the respective purebred mean. The estimated percentage of individual heterosis for post-weaning ADG averaged across more recently reported experiments (McLaren et al., 1987; Baas et al., 1992; Cassady et al., 2002) was only slightly greater (10.5%). Favourable individual heterosis was reported in the reviews by Sellier (1976) and Johnson (1981): average estimates of −3.0% for feed/gain and 5.9% for gain/feed, respectively. However, Cassady et al. (2002) reported that individual heterosis effects for average FI adjusted for ADG ranged from 0.10 to 0.15 kg/day in two separate experiments. A slight unfavourable individual heterosis may exist for backfat. Johnson (1981) concluded that this unfavourable tendency for carcass traits was generally not significantly different from zero, a conclusion supported by the results of Cassady et al. (2002), but both McLaren et al. (1987) and Baas et al. (1992) reported estimates of direct heterosis for backfat that were significantly different from zero (3.2% and 7.2%, respectively). As Sellier (1976) points out, the amount of heterosis for performance traits, especially feed conversion, may depend on the feeding system (restricted versus ad libitum). Effects of maternal heterosis were generally small and unimportant for post-weaning performance (Johnson, 1981; Cassady et al., 2002). Paternal heterosis, i.e. the advantage of a crossbred boar, may be important for the breeding efficiency of young boars, but was also negligible for post-weaning performance (Buchanan, 1987).

All three types of heterosis (individual, maternal and paternal) can be maximized in a static (terminal) cross of F1 sires and dams of unrelated breed background (e.g. AB sires × CD dams). Static crosses also offer the best opportunity to match complementary strengths of lines as sires and dams to fill the paternal and maternal roles in the breeding system. As described by Sellier (1976), differences in maternal effects are the basis for most of the benefit from complementarity in a breeding system, and the appropriate choice of a maternal line is most obvious for traits of reproduction. However, reported differences in feed efficiency (Kuhlers et al., 1972; Johnson et al., 1973) and carcass composition (Bereskin et al., 1971; Johnson et al., 1973) between reciprocal crosses indicate that complementarity may also affect post-weaning performance traits in the system.

Fortunately, the static cross is also the highest ranking breeding system for pork production in terms of overall industry efficiency (Dickerson, 1973). The reproductive rate of the pig results in a relatively small proportion of herds needed for seedstock and multiplier production to maintain the static commercial cross and, thus, a large proportion of animals that realize the maximum benefits of heterosis and complementarity. This breeding design is also compatible with the biosecurity requirements of modern pork production and amenable to scenarios in which commercial multipliers are populated at the top of the pyramid, and subsequently closed to the introductions of outside animals (e.g. Clutter, 2009).

In some situations, the rate of genetic improvement in the industry can be maximized by implementing specialized selection objectives for each line based on its intended role in the breeding system. The objective in paternal lines might focus exclusively on LTFC. Because maternal lines not only produce the commercial sow herd, but also contribute half the genes of market pigs, maternal line selection objectives must give relatively equal emphasis to LTFC and traits of reproduction. Smith (1964) concluded that the pursuit of specialized objectives in sire and dam lines would result in at least as much overall genetic improvement as selection on a general objective, and significantly greater improvement if genetic antagonism exists between market pig performance and traits of reproduction. As discussed in the section on the genetic relationships of growth performance with reproduction, there is evidence in at least some modern pig populations of significant genetic antagonism between production and reproduction traits.
Conclusions

The future competitiveness of pork in the food market depends on continued genetic improvement in the efficiency of quality lean production. From the perspective of post-weaning performance of the market pig, great opportunities exist for improvement through selection within seed-stock lines. Market pig performance can be characterized by several component traits that respond to selection, but the biological index LTFC has been proposed as the most appropriate expression of the industry’s objective for this phase of production.

Investigations of models describing tissue growth have yielded strategies for performance testing and selection to improve LTFC. Results from selection experiments have been reported which support the hypothesis that LTGR and LTFC are perfectly correlated in a time-based scale-feeding system, and that direct response to selection for LTGR (LTFC) among scale-fed candidates will be through increased LTGR and decreased fat growth without depressed ad libitum feed intake. Thus, this or similar testing and selection schemes may be effective for seed-stock lines supplying commercial systems that allow either ad libitum or restricted feed intake.

Knowledge of the genetic relationships of LTFC with the animal’s ability to reproduce is required to determine the optimum selection emphasis for LTFC in the various lines of a breeding system. Although significant correlated responses in reproductive traits have not been detected in some studies of selection for post-weaning performance, detrimental effects on reproduction have also been reported. The greatest risk for negative effects on reproduction from selection for LTFC probably exists for approaches that result in increased leanness through decreased genetic potential for ad libitum feed intake, and unfavourable effects may become more evident after the first parity and reflected in the lifetime performance of the sow.

This chapter has focused on the genetics of efficiency in lean tissue growth. The economic importance of LTFC is largely driven by the modern consumer’s demand for lean cuts of meat, but the quality of tissue in the final product is also of increasing importance. As a result, the genetics of tissue quality is an area of great interest (see Chapter 15), and genetic correlations of components of LTFC with quality characteristics must be well understood in order to effectively design breeding programmes.

As an extension to the information described herein regarding physiological variation contributing to performance traits, the completion of sequencing of the porcine genome and dense-marker association studies of performance traits will continue to reveal greater detail of the underlying genetic architecture and provide tools for enhancing the accuracy of selection (see Chapters 8 and 16). Consequently, methods of selection for post-weaning performance will continue to evolve to optimize the use of genomic information and in response to changing global market demands.

Acknowledgements

The author acknowledges the significant contributions of Professor Pim Brascamp, co-author of the previous edition of this chapter. Thanks are extended to Zhi-Liang Hu for providing summary information from the Pig Quantitative Trait Loci Database (PigQTLdb). The author is grateful for permission to reproduce the figures from Fowler et al. (1976) and Whittemore (1986).

References


Introduction

Traditional quantitative methods have been used successfully in the genetic improvement of traits economically important for producers, including growth rate, feed conversion and backfat/carcass leanness. All these traits are relatively easy to measure on the live selection candidates, and the genetic response is large owing to relatively moderate-to-high heritabilities. While emphasis on production efficiency and lean growth continues to be critical for success, other traits related to processing functionality and consumer satisfaction are of growing importance. If carcass weight, leanness, proportion of quality cuts and processing yields are traits important for processors, consumers are interested in eating-quality traits...
such as juiciness, tenderness, flavour and colour. As a result, meat quality has become increasingly important for the pork industry, and the processors offer branded pork products that are associated with high quality (Huff-Lonergan et al., 1998). The high demand for these products triggered a wave of research from genetics to meat science to improve meat quality and to reduce the production of low-quality pork, such as pale, soft and exudative (PSE) and red, soft and exudative (RSE) meat. Selection, feeding, animal husbandry and processing are the major traditional methods available for increasing the value of pork quality (Barbut et al., 2008). In the last decade, advances in molecular genetics and computational biology, enhanced by the recent availability of the swine genome sequence, opened exciting possibilities for dissecting and understanding the genetic factors that influence pork quality. These new tools are of particular importance because they can be applied to generate information about selection candidates early in life without slaughtering the animals. Several major functional polymorphisms affecting a number of traits measured on slaughtered animals, such as post-mortem pH, colour and tenderness were also identified, and are currently used in selection. The availability of the next-generation methods for sequencing, genotyping and profiling gene expression significantly influence our ability to identify, understand and quantify the expression of genes and networks responsible for the variation of complex quantitative traits. The purpose of this chapter is to review current knowledge and the latest progress in the understanding of genetic influences on phenotypic variation of meat and carcass quality traits.

Traits that Characterize Meat Quality and Carcass Composition

Skeletal muscle is a unique and complex tissue composed mainly of myofibres with various contractile and metabolic characteristics. The main role of the muscle is the contraction and relaxation necessary for the maintenance of posture and locomotion. As a direct consequence, muscle consumes energy in the form of ATP, which is continuously generated in muscle fibres through glycolytic and oxidative pathways. Differences in the percentage of specific fibre types exist both between and within muscles of the same animal. These fibre types differ in contractile rate, amount of force generated and efficiency in using either glycolytic or oxidative metabolism pathways. Ratios of different fibre types vary in diverse muscles, and are directed by developmental stage, muscle activity and muscle function. Fibre type ratios are genetically defined, and substantial differences in fibre type exist between wild and domestic pigs (Ruusunen and Puolanne, 2004), providing evidence that domestication and selection for meat production have resulted in a phenotype consistent with more glycolytic fibres and rapid growth.

The conversion of muscle to meat is a complex cascade of metabolic events that dictate the ultimate quality of meat product by altering the solubility and functionality of native proteins. The following events direct the extent of the changes during the conversion of muscle to meat:

- a gradual depletion of available energy in the form of glycogen, glucose, glucose-6-phosphate and ATP, which leads to rigor mortis;
- a decline in muscle pH from near neutrality to 5.4–5.8 (under normal circumstances);
- a rise in ionic strength, in part, because of the inability of ATP-dependent calcium, sodium and potassium pumps to function;
- an increasing inability of the muscle cell to maintain an environment that protects against lipid and protein oxidation;
- a degradation of myofibrillar, cytoskeletal and intermediate filament proteins.

It has long been known that the metabolic state of muscle at the time of exsanguination is critical in controlling the processes of pH decline and rigor development in post-mortem tissue. Therefore, factors (genetic or environmental) that define muscle metabolism can affect meat quality. Variation in carcass, meat and quality traits is significantly influenced by the genetic differences between the breeds, but variation within the breed is also reported – this topic is discussed in detail in the section on breed variation and combinability.
Meat quality is explained by a group of fresh meat processing and sensory characteristics that are important for profitability and competitiveness of the swine industry in the food market. Processing meat quality traits consist of water-holding capacity, colour, firmness, cooking loss and other processing yields, such as the Napole yield (an indicator of the technological yield of cured-cooked ham processing: the ratio of cooked to raw weight after curing and cooking a 100g meat sample from the semimembranous muscle). Sensory traits are assessed by trained panels or by consumer surveys, and involve appearance (colour and marbling), eating quality as represented by traits that evaluate the texture of uncooked meat, and the tenderness, juiciness and flavour of cooked meat.

Glycogen metabolism in the skeletal muscle directly influences colour reflectance post-mortem. This correlates with several technological traits, such as water-holding capacity, drip loss (the amount of moisture or purge lost over a period of time), tenderness and cooking loss (Sellier, 1998). As a result, pH at 1h post-mortem (pH₁) and pH on the day after slaughter (ultimate post-mortem pH, or pHᵤ) are two meat quality indicators used broadly for predicting the technological and eating quality of pork (Sellier, 1998; Barbut et al., 2008). Composition and intramuscular lipid content can also have some influence on sensory quality. In general, as lipid content increases, star probe values (a measure of tenderness) as well as sensory chewiness scores decrease, especially in pork with intermediate pH values (Lonergan et al., 2007).

The traits most often used to characterize carcass composition are dressing percentage or the proportion of the hot carcass from the live-weight of the animal, carcass length, carcass lean percentage, loin muscle area and backfat thickness. Backfat depth is the preferred measure of carcass composition because: (i) it can be accurately measured on live pigs using ultrasound scanning instruments; and (ii) it represents a good predictive value of carcass fat percentage.

Heritability of Meat Quality Traits and Carcass Composition

The first heritability ($h^2$) estimates for meat quality were reported in the early 1960s (Duniec et al., 1961; Jonsson, 1963; Ollivier and Mesle, 1963). Since then, $h^2$ estimates have been reported for many meat quality traits, with colour and pHᵤ being the most studied. A comprehensive review of these reports was presented by Sellier (1998), and the most recent reported estimates are within the previously reported ranges. A summary of the average and range of $h^2$ values reported by Sellier (1998) is presented in Table 15.1 for most of the meat and fat quality traits. Meat quality traits have in general low-to-moderate heritability, as the average values of many studies fall into the range 0.10–0.30. In the group of sensory traits, the $h^2$ of tenderness measured by subjective and objective determinations ($h^2 = 0.25–0.30$) is more heritable than flavour and juiciness ($h^2 < 0.10$). In the group of technological traits, the average heritability of colour was one of the highest (0.28, range 0.15–0.57), compared with cooking and drip loss (0.16) or pHᵤ (0.21). The high heritability estimates for Napole yield and muscle glycolytic potential involved studies of lines that were informative for the Rendement Napole (RN) locus (one cause of the RSE condition in pork). The substantial effect of the RN allele is responsible for the high heritability estimate for Napole yield and glycolytic potential in a study by Le Roy et al. (1994a). In another study of glycolytic potential, in a Large White line where it is highly likely that the RN allele is absent, the value of heritability was considerably lower ($h^2 = 0.25$; Larzul et al., 1995).

Carcass composition traits are moderately-to-highly heritable. The $h^2$ values of dressing percentage vary from 0.30 to 0.35, and for carcass length they vary from 0.55 to 0.60 (Table 15.2). Backfat thickness was the most studied trait from this class of traits and the average of the $h^2$ values is very similar to that estimated for firmness (0.43). Similar average $h^2$ values were estimated for loin muscle area (0.47) and lean percentage (0.48). These high heritabilities would be expected because breeders have known for over a century that backfat depth is relatively easy to change by selection.

The water, stearic and linoleic acid contents in subcutaneous fat have $h^2$ estimates that
Table 15.1. Range of heritability for meat and fat quality traits (updated from Sellier, 1998).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number and range of heritability estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General quality traits</strong></td>
<td></td>
</tr>
<tr>
<td>pH&lt;sub&gt;1&lt;/sub&gt;</td>
<td>14 0.04–0.41</td>
</tr>
<tr>
<td>pH&lt;sub&gt;u&lt;/sub&gt;</td>
<td>33 0.07–0.39</td>
</tr>
<tr>
<td>Colour (light reflectance, CIE L* value)</td>
<td>29 0.15–0.57</td>
</tr>
<tr>
<td>Water-holding capacity</td>
<td>15 0.01–0.43</td>
</tr>
<tr>
<td>Drip loss</td>
<td>10 0.01–0.31</td>
</tr>
<tr>
<td>Cooking loss</td>
<td>9 0.00–0.51</td>
</tr>
<tr>
<td>Technological yield (cooked ham processing)</td>
<td>3 0.09–0.40</td>
</tr>
<tr>
<td>Napole yield&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5 0.26–0.78</td>
</tr>
<tr>
<td>Meat quality index</td>
<td>13 0.11–0.33</td>
</tr>
<tr>
<td>Visual score of meat quality</td>
<td>8 0.10–0.37</td>
</tr>
<tr>
<td><strong>Eating quality traits</strong></td>
<td></td>
</tr>
<tr>
<td>Tenderness (instrumental determination)</td>
<td>10 0.17–0.46</td>
</tr>
<tr>
<td>Tenderness (sensory panel score)</td>
<td>9 0.18–0.70</td>
</tr>
<tr>
<td>Flavour (sensory panel score)</td>
<td>6 0.01–0.16</td>
</tr>
<tr>
<td>Juiciness (sensory panel score)</td>
<td>8 0.00–0.28</td>
</tr>
<tr>
<td>Overall acceptability (sensory panel score)</td>
<td>2 0.16–0.34</td>
</tr>
<tr>
<td><strong>Muscle composition traits</strong></td>
<td></td>
</tr>
<tr>
<td>% water</td>
<td>7 0.14–0.52</td>
</tr>
<tr>
<td>% lipid</td>
<td>19 0.26–0.86</td>
</tr>
<tr>
<td>% protein</td>
<td>1 –</td>
</tr>
<tr>
<td>% glycogen (glycolytic potential)</td>
<td>3 0.25–0.90</td>
</tr>
<tr>
<td><strong>Fat composition traits (backfat)</strong></td>
<td></td>
</tr>
<tr>
<td>% water</td>
<td>2 0.27–0.42</td>
</tr>
<tr>
<td>% lipid</td>
<td>1 –</td>
</tr>
<tr>
<td>% stearic acid (C18:0)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3 0.30–0.57</td>
</tr>
<tr>
<td>% linoleic acid (C18:2)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3 0.59–0.67</td>
</tr>
<tr>
<td>Androstenone level (entire males)</td>
<td>5 0.25–0.88</td>
</tr>
<tr>
<td><strong>Other traits</strong></td>
<td></td>
</tr>
<tr>
<td>Intensity of boar taint (score)</td>
<td>1 –</td>
</tr>
<tr>
<td>Firmness of backfat (score)</td>
<td>1 –</td>
</tr>
</tbody>
</table>

<sup>a</sup>pH<sub>1</sub>, 1 h post-mortem pH.
<sup>b</sup>pH<sub>u</sub>, ultimate post-mortem pH.
<sup>c</sup>Napole yield is an indicator of the technological yield of cured-cooked ham processing: the ratio of cooked to raw weight after curing and cooking a 100 g meat sample from the semimembranosus muscle.
<sup>d</sup>Expressed in % of total fatty acids.

Table 15.2. Average heritability estimates for carcass composition traits (Sellier, 1998).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrasonic backfat thickness</td>
<td>0.41</td>
<td>0.45</td>
</tr>
<tr>
<td>Fat depth over the tenth rib</td>
<td>0.52</td>
<td>–</td>
</tr>
<tr>
<td>Loin muscle area</td>
<td>0.47</td>
<td>0.48</td>
</tr>
<tr>
<td>Lean percentage</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>Dressing percentage</td>
<td>0.30</td>
<td>0.36</td>
</tr>
<tr>
<td>Carcass length</td>
<td>0.56</td>
<td>0.57</td>
</tr>
</tbody>
</table>
range from 0.35 to 0.64, which are quite a lot higher than the rest of the meat quality traits and more in line with the $h^2$ for carcass fatness (Sellier, 1998). Similarly, the average heritability for fat content in muscle was considerably higher (0.50) than the content of water and protein percentage (0.22 and 0.20). These reports are not surprising since a lipid profile is generally affected by carcass fatness (Sellier, 1998).

Heritability of boar taint score, an unpleasant odour and flavour of the meat from intact male pigs, is moderate to high (0.54). High levels of androstenone and skatole in adipose tissue are considered the primary sources of boar taint. Androstenone is a steroid synthesized in the testicles that causes a pronounced urine-like odour and flavour in meat. The level of androstenone is highly heritable, with an average of 0.56 and ranging from 0.25 to 0.88 (Sellier, 1998; Sellier et al., 2000; Tambryrajah et al., 2004). An additional source of odour is indoles, which are produced by bacteria in the colon during the breakdown of the amino acid tryptophan (Jensen et al., 1995); these give meat a faecal-like odour and flavour (Weiler et al., 2000). Skatole is an indole and its level is characterized by a moderate heritability, ranging from 0.19 to 0.54 (Pedersen, 1998; Tajet et al., 2006). It is not yet fully understood why the high levels of skatole in fat tissue exist in boars but not in castrates and gilts.

The moderate-to-high heritability of carcass traits and the moderate heritability of some of the meat quality traits make them relatively ideal for genetic improvement using traditional selection methods. However, measurement of some of these traits is relatively expensive and needs to be performed post-mortem. Selection of purebred individuals based on the performance of crossbred commercial products is another option, but this will significantly increase the generation interval. Therefore, this group of traits is one for which the application of genomic tools is especially helpful. Training data of the marker effects generated on commercial terminal lines, and updated regularly, can be integrated into the breeding values and used in the successive generations.

Genetic Relationships Between Carcass Composition, Meat and Quality Traits

Genetic correlations between carcass composition traits

Genetic improvement of carcass quality has still been achieved for some of the carcass traits despite the disadvantages associated with phenotyping. For example, in the last 40 years, a strong emphasis on selection for increased lean-ness has utilized the strong negative genetic correlation between ultrasound measurement of backfat thickness on live animals and carcass lean content. The magnitude of this genetic correlation ranged from −0.70 to −0.90, and is not influenced by differences in carcass leanness observed in different breeds (Ducos et al., 1993; Bidanel and Ducos, 1995, 1996; Labroue et al., 1996). Average values of genetic correlations ($r_A$) among major body and carcass composition traits have been summarized by Sellier (1998) based on the reviews of Stewart and Schinckel (1989) and Ducos (1994), and are presented in Table 15.3. Carcass length is shown here to be negatively correlated with loin muscle area (−0.18) and tenth rib backfat (−0.21) but positively correlated with total lean content (0.18).

Genetic correlations between meat quality traits

Estimates of genetic correlations between meat quality traits are affected by the segregation of alleles from major loci that significantly affect the variation of these traits in certain populations. For example a near ‘−1’ (Le Roy et al., 1994b) genetic correlation was estimated between muscle glycolytic potential and Napole technological yield of cured-cooked ham in a commercial line segregating for the $RN^-/rn^+$ locus, while the same correlation estimated in a population presumably fixed for the $rn^+$ allele was considerably lower ($r_A = −0.50$; C. Larzul, personal communication, 1996, cited by Sellier, 1998). A review of the average values of the genetic correlations between the main indicators of pork quality, such as $pH_1$, $pH_u$, muscle glycolytic potential and
intramuscular fat content, and the main processing and sensory traits is presented in Table 15.4, based on Sellier (1998). These studies show that pHu is a good predictor of cooking loss ($r_A$ range $-0.82$ to $-0.45$), drip loss ($-0.99$ to $-0.50$), tenderness ($0.40$ to $0.68$) and overall acceptability ($r_A = 0.59$). In addition to high pHu, another good indicator of low drip loss is a smaller myofiber diameter ($r_A = 0.73$) (Dietl et al., 1993). Interestingly, intramuscular fat is a less accurate genetic predictor of objective tenderness ($r_A = 0.15$) than pHu ($r_A = 0.49$) (Sellier, 1998).

Genetic correlations between carcass composition traits with meat and fat quality traits

A moderate genetic antagonism was reported between the main indicators of meat quality, such as pHu, reflectance and water-holding capacity, and carcass lean-to-fat ratio ($-0.21$ to $0.26$) (Table 15.5). A higher pH, lower colour reflectance and higher water-holding capacity are associated with less lean carcasses. A similar trend was observed for cooking loss: an increase in cooking loss seems to be associated with higher fat-to-lean ratio (Sellier, 1998) (Table 15.5). Relatively stronger correlations were observed between eating quality traits and carcass composition: all eating quality traits (tenderness, juiciness, flavour and overall acceptability) are negatively correlated with carcass leanness ($-0.18$ to $-0.48$) and positively correlated with carcass fatness ($0.24$ to $0.34$) (Sellier, 1998). As expected, the carcass leanness and fatness are genetically correlated with intramuscular fat content, but the estimated values are lower than expected ($-0.34$ and $0.30$, respectively).

In contrast, compositional traits of the subcutaneous fat are moderately to strongly correlated with the lean-to-fat ratio of the carcass. A higher lean-to-fat ratio is genetically associated with a reduced lipid or stearic acid composition and higher linoleic acid content and moisture. For example, saturated fatty acid content (C18:2) is positively correlated with carcass leanness ($r_A = 0.60$) and negatively with carcass fatness ($r_A = -0.70$) (Table 15.6). Linoleic acid is derived from the diet. As pig fatness increases, there is a greater proportion of synthesized lipids, so the adipose fatty acid profile is directly linked to carcass fatness (Piedrafita et al., 2001).

Major Functional Alleles that Affect Carcass Composition and Meat Quality Traits

In the last decade, several major functional polymorphisms affecting a number of carcass meat and quality traits measured on slaughtered animals, such as post-mortem pH, colour and tenderness, were identified and are currently used in selection (Table 15.7). Genetic tests for most

| Table 15.3. Average values of genetic correlations among carcass composition traits (Sellier, 1998). |
|---|---|
| Traits | Reference |
| Lean percentage | Ultrasonic backfat thickness |
| Tenth rib fat depth | −0.87 − |
| Loin muscle area | 0.65 − |
| Killing-out percentage | −0.10 0.20 |
| Carcass length | 0.18 − |
| TenK rib fat depth | Loin muscle area −0.38 |
| | Dressing percentage 0.19 |
| | Carcass length −0.21 |
| Ultrasonic backfat thickness | Dressing percentage − |
| Loin muscle area | Dressing percentage 0.50 − |
| | Carcass length −0.18 − |
| Dressing percentage | Carcass length −0.32 − |
of these polymorphisms are commercially available to the swine breeding industry for genetic improvement of meat quality.

**Ryanodine receptor 1**

The introduction of molecular genetics in pig genetic improvement was initiated by the discovery of a recessive allele in the ryanodine receptor 1 (RYR1) gene that is responsible for porcine stress syndrome (PSS), or malignant hyperthermia syndrome (MH) (Fujii et al., 1991). Pigs homozygous for this allele are susceptible to PSS, and are likely to develop the pale, soft, exudative (PSE) meat condition (Briskey, 1964).

Christian (1972) hypothesized that PSS is a result of an autosomal variation that is recessively inherited. Eikelenboom and Minkema (1974) revealed that the anaesthetic halothane can trigger PSS or MHS in genetically susceptible pigs. Several reports of halothane challenge tests confirmed the recessive

<table>
<thead>
<tr>
<th>Table 15.4. Average values of genetic correlations among meat quality traits (Sellier, 1998).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traits</strong></td>
</tr>
<tr>
<td>Drip loss</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>WHC</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cooking loss</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Napole yield or technological yield</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reflectance</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tenderness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Overall acceptability</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*WHC, water-holding capacity; pH<sub>1</sub>, 1 h post-mortem pH; pH<sub>u</sub>, ultimate post-mortem pH; IMF, intramuscular fat content; GP, muscle glycolytic potential; Napole yield, an indicator of the technological yield of cured-cooked ham processing; tenderness, instrumental determination or sensory panel score; overall acceptability, sensory panel score.*
Table 15.5. Average values of genetic correlations of meat and eating quality traits with carcass composition traits (Sellier, 1998).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Carcass leanness Mean</th>
<th>Carcass leanness Range</th>
<th>Carcass fatness Mean</th>
<th>Carcass fatness Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH1</td>
<td>0.10</td>
<td></td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>pHu</td>
<td>-0.13</td>
<td>-0.50 to 0.08</td>
<td>0.15</td>
<td>-0.05 to 0.45</td>
</tr>
<tr>
<td>Reflectance</td>
<td>0.16</td>
<td>-0.16 to 0.42</td>
<td>-0.21</td>
<td>-0.48 to 0.07</td>
</tr>
<tr>
<td>Water-holding capacity</td>
<td>-0.19</td>
<td>-0.57 to 0.24</td>
<td>0.02</td>
<td>-0.25 to 0.24</td>
</tr>
<tr>
<td>Drip loss</td>
<td>0.05</td>
<td>-0.10 to 0.13</td>
<td>-0.10</td>
<td>-0.20 to -0.01</td>
</tr>
<tr>
<td>Cooking loss</td>
<td>-0.07</td>
<td>-0.16 to -0.06</td>
<td>0.12</td>
<td>-0.04 to 0.39</td>
</tr>
<tr>
<td>Napole yield</td>
<td>-</td>
<td>-</td>
<td>0.15</td>
<td>-0.12 to 0.41</td>
</tr>
<tr>
<td>Muscle glycolytic potential</td>
<td>0.40</td>
<td></td>
<td>-0.21</td>
<td>-0.34 to -0.10</td>
</tr>
<tr>
<td>Meat quality index (IQV)</td>
<td>-0.23</td>
<td>-0.44 to 0.06</td>
<td>0.18</td>
<td>0.01 to 0.39</td>
</tr>
<tr>
<td>Intramuscular fat content</td>
<td>-0.34</td>
<td>-0.55 to -0.07</td>
<td>0.30</td>
<td>0.04 to 0.60</td>
</tr>
<tr>
<td>Tenderness</td>
<td>-0.20</td>
<td>-0.48 to 0.12</td>
<td>0.24</td>
<td>0.11 to 0.48</td>
</tr>
<tr>
<td>Juiciness</td>
<td>-0.18</td>
<td>-0.47 to 0.08</td>
<td>0.29</td>
<td>-0.19 to 0.85</td>
</tr>
<tr>
<td>Pork flavour</td>
<td>-0.27</td>
<td>-0.60 to 0.02</td>
<td>0.35</td>
<td>-0.03 to 0.72</td>
</tr>
<tr>
<td>Overall acceptability</td>
<td>-0.48</td>
<td>-0.71 to -0.32</td>
<td>0.34</td>
<td>-0.04 to 0.70</td>
</tr>
</tbody>
</table>

*pH1, 1 h post-mortem pH; pHu, ultimate post-mortem pH; Napole yield, an indicator of the technological yield of cured-cooked ham processing.

Table 15.6. Average values of genetic correlations of qualitative characteristics of subcutaneous backfat with carcass composition traits (Sellier, 1998).

<table>
<thead>
<tr>
<th>Backfat characteristic</th>
<th>Genetic correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carcass leanness</td>
</tr>
<tr>
<td>% water</td>
<td>0.35</td>
</tr>
<tr>
<td>% lipid</td>
<td>-0.75</td>
</tr>
<tr>
<td>% C18:0</td>
<td>-0.40</td>
</tr>
<tr>
<td>% C18:2</td>
<td>0.60</td>
</tr>
<tr>
<td>Firmness score</td>
<td>-0.40</td>
</tr>
</tbody>
</table>

aData from Schwörer et al. (1988), Bout et al. (1989) and Cameron (1990).

Table 15.7. Functional polymorphisms that affect meat and/or carcass traits in the pig.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Trait</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYR1</td>
<td>Malignant hyperthermia syndrome (MHS), pale soft and exudative (PSE) meat</td>
<td>Fujii et al., 1991</td>
</tr>
<tr>
<td>PRKAG3</td>
<td>Glycogen metabolism, Napole yield, pH and colour</td>
<td>Milan et al., 2000; Ciobanu et al., 2001</td>
</tr>
<tr>
<td>IGF2</td>
<td>Carcass composition, backfat thickness</td>
<td>Van Laere et al., 2003</td>
</tr>
<tr>
<td>MC4R</td>
<td>Carcass composition, backfat thickness</td>
<td>Kim et al., 2000</td>
</tr>
</tbody>
</table>

*ryanodine receptor 1 gene; †protein kinase AMP-activated gamma 3-subunit gene; ‡insulin-like growth factor 2 gene; §melanocortin 4 receptor gene.

In situ hybridization of the linked glucosephosphate isomerase (GPI) gene assigned the HAL locus on SSC6 (Sus scrofa chromosome 6) (Davies et al., 1988).
In skeletal muscle, contraction and metabolism are regulated by the concentration of intracellular Ca\(^{2+}\) (O’Brien, 1987; Britt, 1991), and RYR1 was considered a good positional candidate gene for the observed effects because it modulates the release of Ca\(^{2+}\) by the endoplasmic reticulum (MacLennan and Phillips, 1992). Fujii et al. (1991) uncovered a single missense substitution in RYR1 (p.Arg615Cys, commercially known as HAL-1843) responsible for the variation of the PSS phenotype. The n allele was present at relatively high frequency in several populations as a result of selection for increased lean meat content. The HAL locus is expressed in an approximately additive manner for the carcass lean percentage, loin muscle area, dressing percentage, pH\(_1\), and meat colour score (light reflectance, the CIE L* value) (reviewed by Sellier, 1998). The discovery of this mutation enabled breeders to exploit the intermediate advantages in leanness and other meat quality traits of the heterozygotes (Nn) while controlling the negative effects of PSS in nn homozygotes (Sellier, 1998).

Otto et al. (2007) investigated the relationship between 12 DNA markers, including RYR1-p.Arg615Cys, and several meat quality traits measured on 1155 market pigs. As expected, the p.Arg615Cys substitution had the largest effect on meat quality. Heterozygous individuals were significantly inferior for all meat quality traits with the exception of pH\(_1\), initial conductivity and meat redness. Drip loss measured from 1 to 7 days post-mortem was 43% higher in heterozygotes than in animals of the stress-resistant genotype. Commercial testing of this mutation is available and the reduction of the deleterious allele is being practised worldwide.

### Protein kinase AMP-activated gamma 3-subunit

Early studies of the variation in the processing yield of ham uncovered a locus that segregates in Hampshire populations (Monin and Sellier 1985; LeRoy et al., 1990). The locus, called Rendement Napole (RN), is characterized by the presence of the dominant RN\(^+\) allele that is responsible for an increase of approximately 70% in muscle glycogen content, primarily in the skeletal muscle, compared with the recessive RN\(^-\) allele (Monin et al., 1992; Estrade et al., 1993). The distribution of glycogen content is bimodal, with almost no overlap between RN\(^+\) homozygotes and carriers. Individuals that carry at least one copy of the RN\(^+\) allele produce meat with a lower ultimate pH\(_1\) as a result of post-mortem degradation of the increased glycogen content. The RN\(^+\) allele is almost completely specific to the Hampshire breed and its frequency was increased as a result of selection for lean meat content (Enfalt et al., 1997; Andersson, 2003); although the allele may occur in other breeds, its frequency is negligible.

The RN locus was mapped on SSC15 (Loof et al., 1996; Mariani et al., 1996; Milan et al., 1996), and a mutation in codon 200 (p.Arg200Gln) of the protein kinase AMP-activated gamma 3-subunit gene (PRKAG3) was found to be responsible for the differences in muscle glycogen content (Milan et al., 2000). PRKAG3 is one of the regulatory subunits of AMP-activated protein kinase (AMPK), which plays a key role in regulating energy homeostasis in eukaryotes (Hardie et al., 1998). Environmental or nutritional stress factors could affect the AMP/ATP ratio, which triggers the ‘AMPK cascade,’ which, in turn, initiates measures to conserve energy (Thornton et al., 1998) and activate ATP synthetic pathways (Hardie et al., 1998). This mutation is located in an extremely well-conserved region that directly binds AMP and initiates the process of AMPK activation (Barnes et al., 2004).

Using a three-generation intercross between Berkshire and Yorkshire (B x Y) pigs, a QTL (quantitative trait locus) affecting muscle glycogen content and related meat quality traits (colour and pH\(_1\)) was also mapped to SSC15 (Malek et al., 2001b). The PRKAG3 gene was considered to be a good candidate for the observed effects. The RN\(^+\) mutation was not present in the cross, but three missense substitutions were identified (p.Thr30Asn, p.Gly52Ser, p.Ile199Val). To confirm the hypothesis that additional alleles in PRKAG3 were associated with differences in muscle glycogen content, 1800 animals from five commercial lines were analysed for associations between candidate polymorphisms and ham and loin pH\(_1\) and Minolta L and b colour scores. The results showed that the p.Ile199Val polymorphism is associated with the largest effects.
In a different study, Otto et al. (2007) showed that the homozygote genotype for the PRKAG3-Ile199 allele was associated with less drip loss, higher pHu, and darker meat than the rest of the genotypes. p.Ile199Val is a conservative substitution, and is located next to codon 200 where the RN+/rn− (p.Arg200Gln) polymorphism is present in Hampshire populations. The effect of the PRKAG3 p.Ile199Val_200Arg haplotype is opposite to the effect of 199Val_200Gln as it reduces muscle glycogen content and increases pH post-mortem. The favourable PRKAG3 Ile199 allele is prevalent in all common commercial breeds and the potential implication for improving pork quality is significant. The discovery that the mutation in PRKAG3 other than the classic RN genotype influences meat quality demonstrates that different changes (alleles) in key metabolic proteins have the potential to affect meat quality.

**Calpastatin**

Calpastatin (CAST) is a specific inhibitor of m- and m-calpain proteases and for many years was considered a good functional candidate gene that could affect tenderness. Calpain proteinases (specifically m-calpain) are known to catalyse protein degradation in post-mortem muscle and improve meat tenderness (Huff-Lonergan et al., 1996). Inhibition of calpains by calpastatin can affect the rate and extent of tenderization of meat (Koohmaraie et al., 1991; Sensky et al., 1998; Parr et al., 1999). The first genetic evidence showing that the CAST gene is not only a potential functional candidate but also harbours genetic variants that influence tenderness was reported by Ciobanu et al. (2004). Using the B × Y reference population, Malek et al. (2001b) mapped a suggestive QTL to SSC2 that affected raw firmness scores, average Instron force, tenderness, juiciness and chewiness of cooked meat. CAST was considered to be a good functional and positional candidate for the B × Y QTL, and cDNA (complementary DNA) sequencing of CAST in B × Y and commercial lines uncovered 12 missense and silent mutations (Ciobanu et al., 2004). Haplotypes covering most of the coding region were constructed, and two SNPs (single nucleotide polymorphisms) (p.Arg249Lys and p.Ser638Arg) that specified common haplotypes were used for association analyses with meat quality traits. Results demonstrated that one CAST haplotype (249Lys-638Arg) was significantly associated with lower Instron force (indicating greater tenderness) and higher juiciness sensory scores in B × Y, and lower cooking loss and higher juiciness in a commercial crossbred population (Ciobanu et al., 2004). In a different study, CAST p.Ser638Arg genotypes were associated with more efficient moisture loss in dry-cured hams (Stalder et al., 2005).

The presence of the SSC2 QTL responsible for variation in tenderness measures was subsequently identified in other populations (Stearns et al., 2005a; Rohrer et al., 2006; Meyers et al., 2007). Meyers and Beever (2008) fine mapped the QTL previously discovered (Stearns et al., 2005a; Meyers et al., 2007) and provided additional evidence that CAST is the most likely gene affecting the tenderness measurements. Lindholm-Perry et al. (2009) recently assessed the associations between sliced shear force (SSF) and the genotypes of 40 markers, including nine CAST polymorphisms that cover the SSC2 QTL region from 64.8 to 93 Mb. Only three of the CAST polymorphisms were significantly associated with SSF, including a novel intronic SNP responsible for the largest effect, and two missense SNPs from Ciobanu et al. (2004) (p.Ser66Asn and p.Arg249Lys). The same intronic mutation was significantly associated with variation in CAST gene expression in the longissimus lumborum muscle.

In the most recent study, Nonneman et al. (2010) detected 194 SNPs in the promoter region of calpastatin. Twenty-nine of the SNPs were located in transcription factor binding sites, and gel shift assays for five of the polymorphic sites demonstrated a shift when probes were incubated with nuclear extracts from muscle, heart or testes. Four of these polymorphisms showed allele specificity in binding nuclear proteins that could alter expression of CAST and affect tenderness by regulating calpain protease activity.

**Insulin-like growth factor 2**

A paternally expressed QTL influencing muscle growth, fat deposition and heart size was
identified in crosses between European wild boar and Large White (WB × LW) and between Piétrain and Large White (P × LW) (Jeon et al., 1999; Nezer et al., 1999). This QTL explained 15–30% of the phenotypic variation in lean mass and 10–20% of the variation in backfat thickness. The QTL position coincides with the position of the imprinted insulin-like growth factor 2 gene (IGF2). A substitution located in a conserved CpG island in intron 2 of IGF2 was found responsible for the phenotypic differences (Van Laere et al., 2003). The pigs inheriting the mutant allele from their sire have a threefold increase in IGF2 expression in postnatal muscle and an increase of 3–4% in muscle mass. It was estimated that approximately 50% of the differences in lean mass between Piétrain and Large White pigs are caused by RYR1 and IGF2 polymorphisms (Nezer et al., 1999).

Other genes and DNA sequence variants

The large number of defined markers for meat quality phenotypes is not surprising because genetic changes that affect muscle metabolism and cellular function have the potential to influence the key events during the conversion of muscle to meat and/or the protein and lipid profile of meat. In the last decade, variants of many candidate genes were analysed for their effects on several carcass and meat quality traits. Different alleles in the melanocortin 4 receptor gene (MC4R) were shown to be associated with carcass composition. A missense polymorphism in MC4R (p.Asp298Asn) was associated with differences in feed intake and average backfat thickness in commercial populations (Kim et al., 2000). In a new study, Fan et al. (2009) discovered additional SNPs in the MC4R of the B × Y resource and found that the new p.Arg236His SNP influenced backfat and growth, while the p.Asp298Asn was primarily associated with variation in growth rate. An interaction between the effects of the two SNPs found for average daily gain could explain some of the previous discrepancies reported for MC4R in different pig populations (Park et al., 2002; Stachowiak et al., 2006). SNPs in genes from the cathepsin family (CTSB, CTSD, CTSF, CTSH, CTSK, CTSL and CTSL) known to influence protein degradation were associated with meat production and carcass traits (Russo et al., 2002, 2008; Fontanesi et al., 2010a,b).

The microsomal triglyceride transfer protein (MTTP) gene plays an important role in the assembly of nascent lipoproteins, and was analysed as a positional candidate gene for a QTL on SSC8 affecting the palmitic, palmitoleic and oleic fatty acids in an Iberian × Landrace (I × LR) F2 cross (Clop et al., 2003). A conserved non-synonymous mutation (p.Phe840Leu) in the MTTP gene was found to be significantly associated with MTTP lipid transfer activity and fatty acid profile in porcine fat (Estelle et al., 2009).

Otto et al. (2007) analysed the associations between 12 DNA markers, including RYR1-pArg615Cys and PRKAG3-I199V, and meat quality traits measured on 1155 market pigs. While the RYR1-pArg615Cys and PRKAG3-p.Ile199Val polymorphisms affected many meat quality traits, some of the other markers were also found to influence the variation of some of the traits. SNPs in MC4R (p.Asp298Asn) and high-mobility group AT hook 1 (HMGA1) significantly affected drip loss, while SNPs in the lactate dehydrogenase A gene (LDHA), CAST (p.Arg249Lys) and ATPase Ca\(^{2+}\) transporting fast twitch (ATP2A1) influenced pH. A SNP located in ATP2A1 was associated with differences in intramuscular fat content in the longissimus dorsi muscle, while a SNP in glucose transporter 4 (GLUT4) affected temperature 45 min post-mortem. SNPs in MC4R (p.Asp298Asn), LDHA, GLUT4, HMGA1 and CAST (p.Arg249Lys and p.Ser638Arg) influenced meat colour.

QTL Mapping

Since 1994 a large number of QTL mapping studies have been initiated and many QTLs responsible for variation in growth, carcass composition and meat quality have been reported. In the first QTL study in pigs, Andersson et al. (1994) used the WB × LW F2 experimental population to identify QTLs responsible for the differences in growth and fat deposition. A significant QTL was identified
on SSC4 that was responsible for 20% of the phenotypic variation for abdominal and backfat. A considerable number of QTL studies followed, and most of them were based on resource populations produced by crossing breeds that were phenotypically divergent. Most of the carcass and meat quality traits are quantitative and are affected by many loci that have a wide degree of effects on phenotypic variation. Conflicting effects and implications of most of the chromosomes in the variation of the complex meat quality traits is not a surprise because allelic segregation of these loci can vary from one population resource or cross to another, and markers were spaced far apart. A summary of the populations developed is presented in Table 15.8. Some of the studies employed a full genome QTL scan, while others focus their analysis on a limited number of chromosomes where QTLs were previously discovered. The traits that were easily measured, such as growth, are well represented in these studies compared with the traits that are more difficult to measure, such as carcass composition and meat quality traits. Some of these studies were difficult to compare as they often employed different types of analyses and used different sets of markers.

The Pig QTL database (PigQTLdb; http://www.genome.iastate.edu/cgi-bin/QTLdb/SS/index) has integrated all the QTL mapping results and provided the visualization tools to compare swine QTL data (Hu and Reecy, 2007). The reference map used to display and to compare QTL positions was based on the map reported by Rohrer et al. (1996). Currently, there are 5621 pig QTL results in the PigQTLdb associated with 546 different traits that were reported in 237 publications. The richest traits in QTL reports are for drip loss ($n = 936$), average backfat thickness (149) and loin muscle area (110). In the following subsections we will provide a description of several examples of QTLs that have been mapped in different swine populations.

### QTLs for fatness traits

Significant QTLs for average backfat thickness were detected in all porcine chromosomes with the exception of SSC15, where only suggestive QTLs were located. Consistent QTL effects were detected on the same region of several swine chromosomes in different swine populations or generations. For example, the QTLs for abdominal and backfat revealed by Andersson et al. (1994) on SSC4 were later confirmed by Marklund et al. (1999) in subsequent generations of the same population resource, and later refined by Berg et al. (2006). A QTL for backfat thickness was revealed at the proximal end of SSC1 by Malek et al. (2001b) and by Knott et al. (1998). MC4R is located in the same region of SSC1 and MC4R alleles were found to be associated with feed intake and average backfat thickness in commercial populations. Park et al. (2002) confirmed the causative role of the p.Asp298Asn for any fatness-related traits in a WB × LW cross. Additional SNPs were identified in the B × Y resource and interactions between the effects of two SNPs explained some of the previous inconsistency reported for MC4R in different studies (Fan et al., 2009). Two studies of a Meishan-derived population at the US Meat Animal Research Center (USAMRC) identified QTLs that influence backfat depth on SSC1 and 7 (Rohrer and Keele, 1998a; Rohrer, 2000). The QTL effect from SSC7 was confirmed by a subsequent study in the F8 and F10 generations of the USAMRC Meishan-based population (Kuehn et al., 2007).

A region that was well characterized and affects backfat is located at the end of the short arm of SSC2 harbouring IGF2 locus (Knott et al., 1998; Jeon et al., 1999; Nezer et al., 1999; Rattink et al., 2000; de Koning et al. 2001). A QTL associated with variation in backfat thickness was localized on SSCX by Rohrer and Keele (1998a), Rohrer (2000), Harlizius et al. (2000) and Cepica et al. (2007). A potential candidate gene for this QTL is the thyroxine-binding globulin (TBG) gene. A polymorphism in TBG was associated with an additive effect of 1.2 to 1.6 mm of backfat (Kuehn et al., 2007). QTLs that affect fatness were detected on SSC6 in a Meishan × Large White (M × LW; Bidanel et al., 2001), Iberian × Landrace (I × LR; Ovilo et al., 2000, 2002), B × Y (Malek et al., 2001a), Duroc × Piétrain (D × P; Edwards et al., 2008a) and commercial crosses (Grindflek et al., 2001; Mohrmann...
<table>
<thead>
<tr>
<th>Population</th>
<th>No. of animals</th>
<th>Markers</th>
<th>Traits</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P × LW</td>
<td>1032</td>
<td>11 (SSC2)</td>
<td>GR, CC</td>
<td>Nezer et al., 1999; Van Laere et al., 2003</td>
</tr>
<tr>
<td>M × LW</td>
<td>1103</td>
<td>137</td>
<td>GR, CC, MQ</td>
<td>Bidanel et al., 2001; Milan et al., 2002; Sanchez et al., 2006</td>
</tr>
<tr>
<td>M × GMP</td>
<td>215</td>
<td>318</td>
<td>GR, CC</td>
<td>Wada et al., 2000</td>
</tr>
<tr>
<td>(D × LR) × LW</td>
<td>305</td>
<td>11 (SSC4); 9 (SSC 6); 9 (SSC7)</td>
<td>MQ</td>
<td>Grindflek et al., 2001</td>
</tr>
<tr>
<td>M × LW</td>
<td>390</td>
<td>9 (SSC4)</td>
<td>CC</td>
<td>Walling et al., 1998, 2000</td>
</tr>
<tr>
<td>I × LR</td>
<td>500</td>
<td>7 (SSC6)</td>
<td>CC, MQ</td>
<td>Ovilo et al., 2000; Pérez-Enciso et al., 2000</td>
</tr>
<tr>
<td></td>
<td>369</td>
<td>92 (All autosomes)</td>
<td>MQ, CC</td>
<td>Ovilo et al., 2002; Varona et al., 2002</td>
</tr>
<tr>
<td></td>
<td>139</td>
<td>139</td>
<td>MQ</td>
<td>Mercade et al., 2005, 2006; Estelle et al., 2009</td>
</tr>
<tr>
<td>WB × LW</td>
<td>191</td>
<td>117–236</td>
<td>CC, MQ</td>
<td>Andersson et al., 1994; Andersson-Eklund et al., 1998; Knott et al., 1998; Jeon et al., 1999</td>
</tr>
<tr>
<td>LW, LR, H, P, M, CoL, LW, D × LW, Y × LW, LW, LR</td>
<td>4400</td>
<td>25 (SSC1, 2, 3, 4, 6, 7, 8, 9, 10, 13)</td>
<td>GR, CC, MQ</td>
<td>Evans et al., 2003</td>
</tr>
<tr>
<td></td>
<td>2713</td>
<td>GR, CC</td>
<td></td>
<td>Nagamine et al., 2003, 2004, 2009</td>
</tr>
<tr>
<td>CH × Y</td>
<td>294</td>
<td>5 (SSC4); 10 (SSC7)</td>
<td>CC, MQ</td>
<td>Rothschild et al., 1995; Wang et al., 1998</td>
</tr>
<tr>
<td>B × Y</td>
<td>500</td>
<td>125</td>
<td>CC, MQ</td>
<td>Ciobanu et al., 2001, 2004; Malek et al., 2001a; Thomsen et al., 2004; Kim et al., 2005; Ramos et al., 2006, 2009a</td>
</tr>
<tr>
<td>M × SL</td>
<td>540</td>
<td>157</td>
<td>CC, GE</td>
<td>Rohrer and Keele, 1998a,b; Rohrer, 2000</td>
</tr>
<tr>
<td>M × Y</td>
<td>298</td>
<td>119 (all autosomes)</td>
<td>CC, MQ</td>
<td>Paszek et al., 1999</td>
</tr>
<tr>
<td></td>
<td>116</td>
<td>119 (SSC1, 6, 7, 8, 12)</td>
<td>CC, MQ</td>
<td>Paszek et al., 2001</td>
</tr>
<tr>
<td>M × D</td>
<td>865</td>
<td>180</td>
<td>GR, CC</td>
<td>Sato et al., 2003</td>
</tr>
<tr>
<td>LR</td>
<td>546</td>
<td>24 (SSC1, 2, 3, 4, 6, 7, 8, 9, 10, 13)</td>
<td>CC, MQ</td>
<td>Vidal et al., 2005</td>
</tr>
<tr>
<td>WB × LW</td>
<td>353</td>
<td>225</td>
<td>MQ</td>
<td>Nii et al., 2006</td>
</tr>
<tr>
<td>B × D</td>
<td>832</td>
<td>30 (SSC2, 6, 13, 18)</td>
<td>GR, CC, MQ</td>
<td>Stearns et al., 2005a,b; Meyers et al., 2007; Meyers and Beever, 2008</td>
</tr>
<tr>
<td>D × LR</td>
<td>370</td>
<td>182</td>
<td>CC, MQ</td>
<td>Rohrer et al., 2006</td>
</tr>
<tr>
<td>D × BMP</td>
<td>308</td>
<td>88</td>
<td>CC, MQ</td>
<td>Wimmers et al., 2006</td>
</tr>
<tr>
<td>(P × LW) × AN</td>
<td>715</td>
<td>73 (SSC1, 2, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15, 17, 18)</td>
<td>CC, MQ</td>
<td>van Wijk et al., 2006</td>
</tr>
<tr>
<td>CoL 4× cross</td>
<td>1855</td>
<td>26 (SSC2)</td>
<td>MQ</td>
<td>Heuven et al., 2009</td>
</tr>
<tr>
<td></td>
<td>1187</td>
<td>198</td>
<td>GR, CC, MQ</td>
<td>Harmegnies et al., 2006</td>
</tr>
</tbody>
</table>

Continued
Table 15.8. Continued.

<table>
<thead>
<tr>
<th>Population</th>
<th>No. of animals</th>
<th>Markers</th>
<th>Traits</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D × P</td>
<td>585</td>
<td>106</td>
<td>GR, CC, MQ</td>
<td>Liu et al., 2007, 2008</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td></td>
<td>GE</td>
<td>Ponsuksili et al., 2008</td>
</tr>
<tr>
<td>D × LW</td>
<td>775</td>
<td>91</td>
<td>MQ</td>
<td>Sanchez et al., 2007</td>
</tr>
<tr>
<td>D × P</td>
<td>510</td>
<td>124</td>
<td>CC, MQ</td>
<td>Edwards et al., 2008a,b</td>
</tr>
<tr>
<td>H × LR</td>
<td>450</td>
<td>120</td>
<td>CC, MQ</td>
<td>Markljung et al., 2008; Karlskov-Mortensen et al., 2006</td>
</tr>
<tr>
<td>D</td>
<td>1004</td>
<td>10 (SSC7)</td>
<td>CC, MQ</td>
<td>Uemoto et al., 2009</td>
</tr>
<tr>
<td>D × E</td>
<td>1028</td>
<td>183</td>
<td>MQ, CC</td>
<td>Li et al., 2009; Ma et al., 2009</td>
</tr>
<tr>
<td>M × Y</td>
<td>180</td>
<td>39</td>
<td>CC</td>
<td>Zhang et al., 2007</td>
</tr>
<tr>
<td>P × AN</td>
<td>315</td>
<td>51 (SSC2, 4, 8, 9, 10 and 14)</td>
<td>CC</td>
<td>Duthie et al., 2008</td>
</tr>
<tr>
<td>M × P</td>
<td>316</td>
<td>171</td>
<td>CC, MQ</td>
<td>Geldermann et al., 2003</td>
</tr>
<tr>
<td>WB × P</td>
<td>315</td>
<td>152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WB × M</td>
<td>335</td>
<td>165</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aAN, Anonymous; B, Berkshire; BMP, Berlin Miniature Pig; CH, Chinese breeds; CoL, commercial lines; D, Duroc; E, Erhualian; GMP, Göttingen Miniature Pig; H, Hampshire; I, Iberian breed; LR, Landrace; LW, Large White; M, Meishan; P, Piétrain; SL, Synthetic line; WB, wild boar; Y, Yorkshire.

bCC, carcass composition; GE, gene expression; GR, growth; MQ, meat quality.
et al., 2006). A non-synonymous substitution (pLeu663Phe) in the leptin receptor gene (LEPR) was significantly associated with backfat depth in I × M and I × LR and considered as a candidate for the QTL (Muñoz et al., 2009). Two QTLs mapped in the same region of SSC5 were associated with first-rib backfat in a D × P F2 cross (Edwards et al., 2008b) and with tenth-rib backfat in a Duroc × Landrace (D × LR) F2 cross (Rohrer et al., 2006). Similarly, two different studies mapped QTLs associated with backfat on SSC8 (Rohrer, 2000; Bidanel et al., 2001).

As the majority of the QTLs were revealed in crosses that involved non-commercial pigs, it would be important to determine whether the QTL information could be translated in the commercial lines breeding programmes. Several QTLs for backfat and growth rate previously detected in experimental crosses that usually involved non-commercial pigs, such as Meishan or wild boar, were analysed for segregation in ten different commercial populations (Evans et al., 2003). The study showed segregation of QTLs in most of the populations despite the strong selection pressure for backfat in all analysed populations. The most consistent QTL effects across populations were located on SSC3 and SSC4, and were associated with fatness and growth. In a different study, the segregation of QTLs associated with fatness traits and previously located on SSC4 and SSC7 (Rohrer and Keele, 1998a; Walling et al., 1998; de Koning et al., 1999; Andersson, 2001; Bidanel et al., 2001) were analysed in five commercial lines (Nagamine et al., 2003). The most consistent evidence of segregation was observed for the QTLs located on SSC7. Both studies showed that QTLs revealed in experimental crosses can be detected in highly selected commercial populations and that QTLs that account for variation between populations can also explain genetic variation within populations.

QTLs associated with carcass composition traits

QTLs associated with carcass composition traits have been identified in most of the swine chromosomes. All chromosomes with the exception of SSC15 and SSC16 harboured QTLs associated with carcass length. Two significant QTLs associated with carcass length were mapped on SSC6 and 7 in a D × P cross (Edwards et al., 2008b). The QTL from SSC6 was associated with a dominance effect, while the QTL from SSC7 that influences both carcass length and dressing percentage acted in an additive manner. The allele derived from Duroc was found to increase carcass length and decrease dressing percentage. The position of the QTL on SSC6 coincides with QTLs identified in B × Y (Malek et al., 2001a), Meishan × Duroc (M × D; Sato et al., 2003) and Berkshire × Duroc (B × D; Stearns et al., 2005a) crosses. A dominance effect similar to that reported by Edwards et al. (2008b) on SSC6 was identified by Malek et al. (2001a). A QTL located in a similar position on SSC7 was reported by Rohrer and Keele (1998a) and explained more than 15% of carcass length in a M × LW cross, with the Meishan allele increasing the trait. Malek et al. (2001a) also revealed suggestive QTLs on SSC6, 11 and X, with the Berkshire (B) alleles increasing the carcass length for two out of three QTLs. All these effects were responsible for approximately 10.6% of the variation. The halothane-sensitive n allele from RYR1 is known to positively influence carcass length and dressing percentage, while the RN allele from PRKAG3 positively influences carcass length (Sellier, 1998; Le Roy et al., 2000).

QTLs significantly associated with dressing percentage were identified on all chromosomes with the exception of SSC5, 9, 11, 12 and 17. For example, QTLs with significant effects on dressing percentage were detected on SSC1 in a D × P cross (Edwards et al., 2008b), SSC4 in a B × Y cross (Malek et al., 2001b), and in the same region of SSC7 in M × LW (Rohrer and Keele, 1998b) and D × P crosses (Edwards et al., 2008b). Additional QTLs have also been suggested by Rohrer and Keele (1998b) on SSC3, by Malek et al. (2001b) on SSC7, 8, 13 and 14, and by Milan et al. (2002) in a different region of SSC4. The effects of Large White alleles were always favourable compared with Meishan alleles, except for the QTL from SSC8, where the effects were unfavourable compared with Berkshire alleles. Significant dominance
(SSC7 and 8) and even overdominance (SSC13 and 14) effects were observed in the B × Y cross (Malek et al., 2001b), but not in the M × LW crosses. The QTLs explain in most cases a relatively small fraction (< 8%) of the phenotypic variation.

Significant QTLs that affect carcass yield were mapped on SSC4, 7, 10 and 12 (Sato et al., 2003; Thomsen et al., 2004; Harmegnies et al., 2006). A consistent QTL that influenced variation in carcass yield was located on SSC7 in B × Y crosses (Thomsen et al., 2004) and in a commercial four-way cross that involved Large White, Landrace and Piétrain lines.

Most of the chromosomes were detected to harbour QTLs that have significant or suggestive effects on the relative proportions of carcass lean and fat tissues. Information from the literature indicates that SSC2, 4, 8, 9, 10 and 14 are highly associated with lean and fat tissue growth (e.g. Andersson et al., 1994; Malek et al., 2001a,b; Geldermann et al., 2003). QTLs associated with traits such as loin, ham and shoulder weight or percentage, with or without fat, were identified on several chromosomes and have an important economic value. A QTL associated with entire ham weight was identified at the proximal end of SSC8 in Wild Boar × Piétrain (WB × P; Beeckmann et al., 2003) and Piétrain × a commercial cross dam F2 crosses (Duthie et al., 2008). A QTL associated with loin weight without external fat was identified at the proximal end of SSC2 in Wild Boar × Large White (WB × LW; Andersson-Eklund et al., 1998). Wild boar × Meishan (WB × M; Geldermann et al., 2003) and Pietrain × commercial cross dams (P × CoL; Duthie et al., 2008) in the same very proximal end of SSC2, paternally expressed QTLs with significant effects on muscle growth of important carcass cuts such as loin, ham and shoulder were identified in WB × LW (Jeon et al., 1999), P × LW (Nezer et al., 1999) and P × CoL (Duthie et al., 2008) crosses. It has already been shown that an SNP located in a conserved CpG island in intron 2 of IGF2 is the source of these phenotypic differences (Van Laere et al., 2003).

As expected, large variations in QTL effects were also observed between experiments. For example, the QTL located on SSC2 in the IGF2 region explained 5–6%, 3–7%, 15–30% and 30% of the phenotypic variance of carcass lean proportions in M × LW, P × CoL, WB × LW and P × LW crosses (Jeon et al., 1999; Nezer et al., 1999; Bidanel et al., 2001; Duthie et al., 2008). QTLs that influence loin muscle area were detected on all chromosomes with the exception of SSC16 and 18. One of the QTLs with the largest effects was detected on SSC6 and is responsible for 21% of lean eye area variance in the I × LR cross (Ovilo et al., 2000). In the same region, Edwards et al. (2008b) identified a QTL responsible for approximately 5% of the phenotypic variation of loin muscle area in a D × P cross.

Meat quality QTLs

Quality of meat is influenced by several important characteristics, including general appearance, colour, taste, marbling, texture and tenderness. QTLs affecting the variation of these complex quantitative traits were identified in several regions of most chromosomes. There are several overlapping QTLs associated with colour scores. For example, the position of a QTL on SSC14 associated with lightness values (Hunter L*) in the D × LR cross (Rohrer et al., 2006) coincides with a QTL for Hunter L* reflectance in the B × Y cross (Malek et al., 2001b) and the colour score in an F2 cross between Meishan and commercial pigs (M × CoL; de Koning et al., 2001). A significant QTL detected on SSC1 for Hunter L* by de Koning et al. (2001) was also reported by Rohrer et al. (2006).

One of the major QTLs affecting pHu and colour was initially identified on SSC15 in the B × Y cross (Malek et al., 2001b). This QTL explains 4–6% of pHu variance, and the Berkshire allele, partly recessive, is increasing the trait. The QTL, which also affects muscle glycolytic potential, is localized in the same region as the RN locus. The RN− allele uncovered by Milan et al. (2000) as an SNP (p.Arg200Gln) located in codon 200 of the PRKAG3 gene was absent in the population studied. Additional polymorphisms in PRKAG3 (p.Thr30Asn, p.Gly52Ser, p.Ile199Val) were
identified as the source of the variation (Ciobanu et al., 2001). An association analysis of the PRKAG3 haplotypes in commercial populations revealed that differences in pH between the four haplotypes may be as high as 0.1 pH units in all lines analysed, with the exception of Berkshires where the differences may exceed 0.2 units (Rothschild et al., 2004). The results also showed that pIle199Val polymorphism was associated with the largest effects. A suggestive QTL associated with change in pH was identified in the same region of SSC15 in a D × LR cross (Rohrer et al., 2006).

A suggestive QTL for tenderness score and Star Probe Force (a measure of tenderness) in the B × Y cross (Malek et al., 2001b) and SSF in the D × LR cross (Rohrer et al., 2006) was identified at about 20 cM away from the QTL located on SSC15 and associated with pH3 and colour scores. One of the most interesting and studied QTLs associated with tenderness was revealed on SSC2. Suggestive QTLs affecting raw firmness scores, average Instron force, tenderness, juiciness and chewiness on cooked meat were mapped on SSC2 in the B × Y cross (Malek et al., 2001b). Subsequent studies mapped QTLs associated with tenderness in the same regions of SSC2 in the B × D (Stearns et al., 2005b) and D × LR (Rohrer et al., 2006) crosses. Ciobanu et al. (2004) identified CAST as the main candidate for the QTL and provided the first genetic evidence that CAST is not only a potential functional candidate gene but also harbours genetic variants that influence tenderness.

Marbling represents one of the important appearance factors used by consumers to perceive quality, and significantly influences purchasing decision (Resurreccion, 2004). Significant QTLs for the subjective measurements of marbling were detected on SSC1, 2, 5, 6, 7, 8, 10, 12, 13, 14 and 17. QTLs for amounts of intramuscular fat determined by objectives methods were identified in most of the chromosomes with the exception of SSC3, 10, 11, 14 and 16. Chromosomal regions that harbour QTLs for marbling and intramuscular fat in different populations are located on SSC1, 2, 5, 6, 8, 13 and 17. A QTL for intramuscular fat or marbling was identified near the heart fatty-acid binding protein (H-FABP) gene on SSC6, in a commercial population (Grindflek et al., 2001) and in I × LR (Ovilo et al., 2000, 2002) and D × P crosses (Edwards et al., 2008b). This QTL seems to explain 15–20% of the variation in intramuscular fat, with Duroc and Iberian alleles increasing the trait. Previously, Gerbens et al. (1999, 2000) used the candidate gene approach and found significant associations between polymorphisms in the H-FABP gene and intramuscular fat in Duroc pigs. Two potential candidate genes, H-FABP and small heterodimer partner (SHP), were evaluated by Arnyasi et al. (2006) as potential candidates for the QTL. Significant associations were observed between H-FABP polymorphisms and intramuscular fat, which explained 30–35% of the variation and confirmed the Gerbens et al. (1999) study. However, other studies failed to show a clear association between H-FABP alleles and intramuscular fat (Nechtelberger et al., 2001; Ovilo et al., 2002; Urban et al., 2002).

A QTL with significant effects on marbling was detected on SSC1 (near MC4R) in a B × Y cross (Malek et al., 2001b). Yorkshire alleles have favourable effects compared with Berkshire alleles, but this QTL explains only 3–4% of the phenotypic variance of intramuscular fat. QTLs associated with intramuscular fat were detected on the same chromosomal arm but closer to the proximal end in D × LR (Rohrer et al., 2006) and Duroc × Large White (D × LW; Sanchez et al., 2007) crosses. The Duroc allele in D × LR (Rohrer et al., 2006) and the Large While allele (Sanchez et al., 2007) increase intramuscular fat. A significant QTL associated with intramuscular fat was uncovered on SSCX in M × LW (Harlizius et al., 2000) and Duroc × Erhualian (D × E; Ma et al., 2009) crosses. Alleles inherited from Meishan and Erhualian breeds increase intramuscular fat. Harlizius et al. (2000) found the estimated QTL effects to be smaller in females than in males, probably as a result of random inactivation of the SSCX in females. The long-chain acyl-CoA synthetase family member 4 (ACSL4) gene was considered as a potential positional candidate for the QTL detected on SSCX (Mercade et al., 2006).
The fatty acid composition is known to affect nutritional, technical and sensory qualities of fresh and processed pork product (Lawrence and Fowler, 1997). QTLs influencing fatty acid composition in the latissimus dorsi muscle and perirenal, abdominal and backfat have been identified in several experimental populations (Pérez-Enciso et al., 2000; Clop et al., 2003; Lee et al., 2003; Kim et al., 2006; Nii et al., 2006; Sanchez et al., 2007; Guo et al., 2009; Uemoto et al., 2009). For example, Clop et al. (2003) uncovered significant QTLs for fatty acid composition on SSC4, 6, 8, 10 and 12 in an I × LR intercross. The QTL located at SSC8 had significant effects on palmitic and palmitoleic fatty acids, and a suggestive effect on oleic fatty acid. A missense mutation in a conserved site of the MTTP gene was shown to be strongly associated with protein activity and fatty acid profile in pigs. A QTL associated with fatty acid composition has been uncovered in an I × LR cross (Muñoz et al., 2007) and in a Duroc commercial line (Quintanilla et al., 2007). Two linked SNPs in acetyl-coenzyme A carboxylase α (ACACA), a positional candidate gene for the discovered QTL, were significantly associated with carcass leanness, intramuscular fat, and monounsaturated, saturated (myristic, palmitic and stearic) and polyunsaturated (linoleic) fatty acids in the longissimus thoracis and longissimus lumbarum muscles (Gallardo et al., 2009).

Boar taint QTLs

Boar taint is an unpleasant odour and flavour of the pork from intact male pigs, primarily caused by high levels of androstenone and skatole in adipose tissue (Robic et al., 2008). It has been suggested that a major gene that controls the level of androstenone in fat is segregating in Large White populations (Fouilloux et al., 1997), and that another gene influences the skatole level (Lundström et al., 1994). Quintanilla et al. (2003) performed a QTL analysis of the androstenone levels in fat in a M × LW F2 cross at 100, 120, 140 and 160 days old, and on individual boars slaughtered at approximately 80 kg live weight. There were four robust QTLs detected on SSC3, 7 and 14 that individually explained up to 15% of the phenotypic variance at different ages. This study indicated that the variation of this trait is complex and affected by numerous genes. A population of a similar genetic background was analysed by Lee et al. (2005) and QTLs for the laboratory measurements of androstenone level in fat measured at slaughter (85 kg live weight) were mapped on SSC2, 4, 6, 7 and 9. These five significant QTLs explained 6.2–8.6% of the phenotypic variance. In the same region of SSC6, a QTL for boar taint flavour assessed by a sensory panel was located. QTLs were also detected on SSC14 for the laboratory estimates of the level of indole and skatole, sensory panel scores for skatole and the scores for boar taint flavour in lean and in fat. In both studies, with the exception of SSC7, Meishan QTL alleles were associated with a high level of androstenone in backfat, indicating an important breed effect on boar taint phenotype.

Varona et al. (2005) conducted the third mapping study of QTLs that affect variation in boar taint. Androstenone and skatole levels in fat were determined in 217 intact Landrace boars, and QTL mapping was performed on ten regions of the genome chosen on the basis of previously detected QTLs for growth and fatness. Boar taint QTLs have been mapped on SSC3, 4, 6 and 7. A QTL for skatole level was mapped on SSC6 in a region where a QTL had been previously detected for androstenone (Lee et al., 2005).

An association study that involved 275 SNPs in 121 genes and more than 2800 individuals from Duroc and Landrace boars revealed genes with significant SNPs and haplotypes associated with compounds related to boar taint (Moe et al., 2009). These genes include several cytochrome P450 members (CYP2E1, CYP21, CYP2D6 and CYP2C49), nuclear receptor NGFIB and catenin delta (CTNND1). Recently Grindflek et al. (2010) compared gene expression levels in testicle samples of 192 Duroc and Landrace boars with extreme levels of androstenone in fat. Ten out of the 15 candidate genes analysed were significantly up-regulated in high androstenone boars from both investigated breeds. Associations between SNPs from differentially
expressed genes and androstenone level were observed only for cytochrome b (SCYB5A), suggesting a mechanism of cis-regulation that modulates the expression of this gene.

Expression QTLs

Variation in gene expression contributes to phenotypic diversity and has an impact on variation of high-order phenotypes, including meat quality traits. The advent of high-throughput microarray technology in combination with linkage analysis has allowed simultaneous mapping of expression QTLs (eQTLs) that potentially control the expression of thousands of transcripts. Transcript abundance is an intermediate phenotype between DNA sequence variation and complex phenotypes. Gene expression traits are also affected by complex regulation and environmental noise, but they are considerably closer to the molecular level than the phenotypes they modulate.

Lobjois et al. (2008) analysed gene expression in 17 F2 individuals derived from a commercial cross that displayed a range of Warner–Bratzler Shear Force (WBSF) measurements of tenderness. From 63 genes with expression associated with WBSF, 22 were successfully mapped in the pig genome and 12 were located in the areas previously reported to harbour QTLs associated with tenderness (SSC2, 6 and 13). The first comprehensive genome-wide expression QTL mapping in a livestock species was recently reported by Ernst et al. (2010) in pigs. Expression profiles from loin muscle tissue of 176 F2 D × P pigs were combined with linkage analysis of 124 microsatellite markers. Physical localization of a preliminary genome assembly was achieved for 13,611 oligonucleotides out of 20,400 oligonucleotides present in the array, and for 77 of the 124 microsatellite markers. There were 62 unique eQTL detected, including 24 that mapped at the location of the source gene (cis eQTL). Comparison of the eQTL position with a set of 173 putative QTLs previously reported for growth and carcass traits uncovered 11 common linkage regions, including seven regions involving putative cis-acting eQTLs. For example, the location of a QTL for loin muscle area coincided with an eQTL that modulates dynein, light chain, Tctex-type 1 (DYNLT1) expression. As shown in this study and others (e.g. Chen et al., 2008; Farber et al., 2009; Ciobanu et al., 2010), differences in gene expression can be used in reverse genetic studies to generate well-defined hypotheses regarding downstream effects on molecular, cellular and functional networks, and finally at the phenotype level.

Genetic Traceability

Traceability is the process of identification of animals and animal products through the steps within the food chain from farm to the retailer. Conventional traceability consists of a labelling system and the management of processed food in batches (Dalvit et al., 2007). This approach can trace the provenance of meat products and is mandatory in the USA and in all EU member countries. Genetic traceability is an approach for the identification of animals and their products based on variation at the DNA level (Mackie et al., 1999). However, the cost of DNA testing means that its practical application, at least currently, is in the verification of traceability systems. In the case of swine, this technology can be applied for identification of a particular animal or breed using deterministic and probabilistic approaches.

Individual identification, or tracking back to an individual animal or to the system it has been produced in, helps to reinforce the safety of products for consumption and assist actions associated with the protection of the health of human and animal populations in the case of disease outbreaks. Typing highly polymorphic DNA markers, such as microsatellites, was proposed as a viable solution owing to its precision and durability (Dalvit et al., 2007). Match probability (MP) was used to evaluate the efficacy of a panel of markers. A MP estimate is the probability of finding, by chance, two individuals that share the same genotypes at the analysed loci (Weir, 1996). Goffaux et al. (2005) suggested that a panel of 21 SNPs would give an MP of 7 × 10−9, which is sufficiently significant for the size of the pig population of Belgium (7 × 106). Obviously,
it becomes increasingly difficult to trace individual animal products in mixtures such as ground meat. Shackell et al. (2005) showed that DNA microsatellites may be useful for the application of DNA traceability of ground beef mixtures prepared from fewer than ten individuals, but less accurate in mixtures from larger numbers.

Individual traceability is an approach that supports food safety, but it can also be used to verify production claims. An example of this is the use of specific genotypes for the production of valuable products, as exemplified by the products of the Iberian pig (in Spain) and the Berkshire (in Japan and the USA). These products have been differentiated for their special sensory properties, and genetic tests are available to certify their origin (Dalvit et al., 2007). In this case, breed traceability can be used to certify food products of higher value and quality from certain breeds. A genetic test that involved 25 microsatellites revealed that up to 20% of Iberian ham samples, which cost up to ten times more than ‘normal’ ham, have a genetic composition incompatible with the present Spanish legislation – either because the Duroc genome was present in a percentage greater than that permitted (>50%), or because of the significant presence (>25%) of white-coated pig genomes (García et al., 2006).

One of the main distinguishing characteristics of swine breeds is coat colour and pattern (see also Chapter 3). Much of the variation in colour is explained by allelic variants of the MC1R and KIT genes (Kijas et al., 1998; Marklund et al., 1998; Kijas and Andersson, 2001). Initial sequence analysis of MC1R from different breeds revealed four polymorphisms corresponding to five different E alleles, with one of the alleles, E<sup>p</sup>, being a composite of two of the polymorphisms (Kijas et al., 1998). Additional alleles have subsequently being describer (see Chapter 3). A combined duplication and splice mutation of the KIT I allele causes the dominant white-coat colour in breeds such as Large White and Landrace. The presence of the duplication and the absence of the splice polymorphism of the KIT I<sup>p</sup> allele is associated with spotted coat colour in the Piétrain breed. In addition, three KIT intronic polymorphisms have been used to differentiate Berkshire from Tamworth (Alderson and Plastow, 2004), or the Berkshire breed from other populations in Japan (Carrion et al., 2003).

While the recent development in DNA sequencing, polymorphism discovery and high-throughput genotyping has underlined the potential of molecular genetics in traceability applications, its current use is limited because of the management involved in collecting and storing individual samples and the cost of genotyping.

**Breed Variation and Combinability**

**Breed differences in carcass composition, meat and fat quality traits**

The genetic background of the breed has a significant effect on the variation of most carcass and meat quality traits known to be influenced by a large number of genes. There are commercial lines marketed as ‘meat quality lines’, but significant variation in meat quality traits exists within these (Gil et al., 2003; Plastow et al., 2005). Comprehensive reviews of the numerous studies that compared fat and carcass quality traits among pure breeds/lines and various crosses were published in several reports in the 1980s and 1990s (Sellier, 1983, 1988, 1998; Sutherland et al., 1985; Schwörer et al., 1989; Wood and Enser, 1989; Sellier and Monin, 1994). Since then, genetic progress, adjustments in the selection criteria and the introduction of molecular technologies have generated changes in the genetic background and variation of certain traits in swine populations. Important contributions to the differences among commercial lines are attributed to the frequencies of alleles from certain loci responsible for major effects on carcass composition, meat and fat quality traits. RYR1 (HAL) or PRKAG3 (RN) are classical examples of genes that harbour alleles with strong effects that drive important variation in many economically important traits. In early studies of the HAL locus (Guéblez et al., 1995; Hanset et al., 1995) a significant difference in allelic frequency of the stress-resistant N allele in Piétrain (q<sub>n</sub> >0.90) and Large White (q<sub>n</sub> <0.10) was
determined to be the main source of the significant difference in carcass (dressing and lean percentage, carcass length) and meat quality (pH) traits. The non-conserved R200Q substitution (RN+/rn−) in the PRKAG3 gene, which was discovered by Milan et al. (2000), explains a 70% increase in muscle glycogen in RN− homozygous and heterozygous animals. The high frequency of the RN− mutation in US Hampshire pigs (pRN− = 0.63; Miller et al., 2000) compared with Yorkshires, where the RN− allele is absent, results in lower muscle pH, reduced water-holding capacity and a much lower yield of a cured cooked ham product in Hampshire pigs (Monin and Sellier, 1985; LeRoy et al., 1990).

The discovery of functional mutations and the introduction of genetic tests that differentiate all genotypes have reduced the influence of these two major loci on the variation of meat and carcass quality traits: the RN− allele has been eliminated from most of the Hampshire populations, while the stress-susceptible n allele has a very limited use as a result of the advantages in leanness and other meat quality traits of the heterozygotes (Nn). In contrast, the new favourable PRKAG3 Ile199 allele is prevalent in all common commercial breeds, and the high frequency of this allele in Berkshire (0.87) compared with other lines (Duroc, 0.38; Large White, 0.22; Landrace, 0.14) is one of the genetic sources that leads to a higher pH, and better colour scores of ham and loin in the Berkshire (Ciobanu et al., 2001).

In recent years, there have been several comparison studies of the carcass composition, meat and fat quality traits between commercial lines with different genetic backgrounds. In the study of Gispert et al. (2007), carcass composition was analysed and compared between Large White, Landrace, Duroc, Piétrain and a Meishan synthetic developed from a cross with a Large White-based line. The first four lines represent a significant proportion of North American and European pig production (Gil et al., 2008). Carcasses from the Piétrain line (Halothane negative) had the highest dressing percentage (83.3%) and were the shortest (81.8 cm). The Piétrain line was associated with the highest proportion of ham (270.9 g/kg) and the lowest proportion of belly (98.0 g/kg). The Piétrains were the leanest overall (755.9 g/kg), and the leanest in all of the dissected cuts compared with the Meishan, which was the fattest (638.5 g/kg) and had the highest intermuscular fat. The Duroc had a relatively high level of intramuscular fat and intermediate carcass quality but, in the study of Guerrero et al. (1996), it showed the best characteristics for dry-cured hams compared with Piétrain, Belgian Landrace and Large White × Landrace (LW × LR) crosses.

Gil et al. (2008) characterized five commercial lines for meat quality parameters and muscle biochemical characteristics of the longissimus thoracis (LT), and semimembranosus (SM) muscles. The muscle fibre size was the main difference in the LT. The largest difference was found between Meishan and Piétrain lines, with the Piétrain muscle fibre size the largest. The Duroc line was differentiated from the rest of the breeds by muscle oxidative traits, and the Landrace line was differentiated by the high percentage of fast glycolytic fibres. Duroc and Piétrain were different from Landrace and Meishan according to the metabolic and contractile characteristics of the SM. The measured muscle characteristics were associated with differences in drip loss and marbling, and could thereby influence the eating quality of pork.

Brewer et al. (2002) analysed quality characteristics of pork derived from pigs of the following backgrounds: Duroc and Piétrain (Halothane negative, NN), Piétrain (Halothane positive, nn), Berkshire and Hampshire (rn+), Hampshire (RN−) and a synthetic line. A trained panel evaluated the visual appearance of uncooked lean and fat, and the flavour and texture of cooked chops. Colour scores (Hunter L*, a* and b* values), hue angle, cooking loss and Warner–Bratzler shear force were also measured. The genetic background affected the visual colour differences between lines, with the chops from Duroc, Berkshire and Piétrain (nn) being the least pink. Berkshire chops appeared to have the most marbling in the lean, and those from Piétrain (nn) appeared to have the least marbling. Chops derived from Piétrain (nn) and Hampshire (rn+rn+) had the highest a* values, while those from Duroc, Piétrain (NN) and Hampshire (RN−) had the lowest. Chops from Duroc carcasses were associated with the
lowest cooking loss (18.8%) compared with Hampshire \( (rn^+rn^-) \), Piétrain \( (NN) \), and Hampshire \( (rn^+rn^-) \), 6.0 kg carcasses. Pork derived from Hampshire \( (RN^-) \) was associated with the highest juiciness, followed by those from Hampshire \( (rn^+rn^-) \), Piétrain \( (NN) \), Berkshire and Duroc.

In the study by Ciobanu et al. (2001) of the influence of \( PRKAG3 \) alleles on meat quality traits in five commercial lines, Berkshire was associated with the highest loin pH\(_u\) (5.83) and ham pH\(_j\) (5.81) and the lowest ham Minolta b\(^*\) (3.46) and Minolta L\(^*\) (43.12) (colour/reflectance scores). Duroc was associated with the lowest drip percentage (0.14%), while Large White had the highest (2.32%). Plastow et al. (2005) studied variability for meat and fat quality traits, within and between five commercial breeds, and found Duroc to be associated with the lowest drip loss (2.49%), Minolta b\(^*\) (3.44) and L\(^*\) score (46.15) and, together with Piétrain \( (NN) \), the highest pH\(_u\) (5.64). The highest proportion of intramuscular fat in LT was found in Meishan (1.8%) and Duroc (1.9%) lines, followed by Piétrain (1.2%, the overall leanest line), Landrace (1.1%) and Large White (1.0%). Regarding the fatty acids, the percentage of stearic acid was higher in the Duroc line (13.4%) than in the rest of the lines, with Piétrain having the lowest (11.9%). The percentage of stearic acid was relatively low in Meishan (12.17%), even if the Meishan LT was associated with highest marbling. Interestingly, the abundance of linoleic acid was significantly lower in Duroc and Meishan than in Landrace, Large White and Piétrain lines. The report suggests that the metabolism of these C18 acids was different between lines, which could be significantly influenced by genetics.

**Heterosis and maternal effects in breed crosses**

Crossbreeding is used almost exclusively in the pig industry, and a better understanding of the best combinations of genetic backgrounds to achieve heterosis for a particular trait is critical in implementing optimal crossbreeding schemes (Sellier, 1998). It is well established that heterosis has a limited effect on carcass traits, as exemplified by the F\(_1\) crosses between breeds that display important differences for this group of traits, such as Large White (or Yorkshire) and Meishan (Poilvet et al., 1990; Serra et al., 1992; Lan et al., 1993). The direct and maternal heterosis effects on meat quality traits are also relatively small, and are described in several reports that employed most breed combinations and/or most meat and eating quality traits (e.g. Young et al., 1976; Schneider et al., 1982; McLaren et al., 1987; Lo et al., 1992; Serra et al., 1992; Bidanel et al., 1993; Lan et al., 1993; Ellis et al., 1995). Previous studies showed heterosis effects on pH\(_u\), pH\(_j\) and related traits in crosses informative at the \( RN \) and \( HAL \) loci, summarized by Sellier (1998).

As discussed above, the \( RN^- \) allele has been eliminated from most of the Hampshire populations, while the stress-susceptible \( n \) allele is maintained at low frequency in certain populations owing to advantages in leanness and other meat quality traits in heterozygotes \( (NN) \). For example, the crosses that involve the ‘stress-sensitive’ Piétrain showed favourable heterosis effects on pH\(_u\), with the F\(_1\) hybrids having a phenotype closer to the ‘stress-resistant’ parent (Sellier, 1987). Significant direct heterosis effects (−20% to −13% of the parental mean) were reported on muscle lipid content in the Duroc × Landrace cross (Wood et al., 1987; Lo et al., 1992). Heterosis was also observed for intramuscular fat in females of the crosses that involved Large White and Meishan (Serra et al., 1992) or for Large White and Meishan × Large White (Poilvet et al., 1990).

The genetic correlation between the performance in a purebred Genetic Nucleus environment and in a crossbred-commercial environment is less than one, owing to different degrees of sensitivity. For example, the estimates for backfat range from 0.21 to 0.88, while those for growth vary from 0.19 to 0.99 (Brandt and Taubert, 1998; Lutaaya et al., 2001). As a result, crossbred performance has begun to be used in combination with purebred
performance in an index to estimate the sensitivity of the genetics inherited from a boar within a range of environments. Combined crossbred and purebred selection approaches have been reported to be superior in certain situations to pure selection alone (Bijma and van Arendonk, 1998; Lutaaya et al., 2001; Perez et al., 2006). For example, the incorporation of and emphasis on commercial/crossbred records can increase the accuracy for breeding for pH by 0.04 units (A.A. Sosnicki, personal communication).

Future Directions in Breeding on Carcass and Meat Traits

Significant emphasis on production efficiency remains a priority in pork production. This emphasis is critical in the face of demand for the responsible use of natural resources in food production. Advances in quantity of production and efficiency of production should not come at the expense of processing quality, sensory quality or nutritive quality. Several major changes witnessed lately in the food industry, consumer choices and the recent revolution of molecular technologies could have an important impact on the future role of genetics in meat quality improvement. Industry consolidation in the last few years has led to the integration and coordination of major sectors of food production systems from animal breeding to packing and retail. Consumer demand for high-quality products has led to an important shift from ‘commodity pork’ to meat product differentiation and ‘value-added pork’, which is superior for eating quality characteristics such as tenderness, juiciness and flavour. Future emphasis on the quality of other pork cuts will provide important information regarding the use of valuable products such as ham and bacon.

A number of single DNA tests have become available to breeders for use in the past few years. While currently there is no premium associated with higher quality carcasses, some of these markers are used intensely in the genetic improvement of lines dedicated for the sector of the industry that manages both pig production and meat packing activities. The main traits targeted by genetic marker improvement are drip loss and meat pH. More excitingly, rapid progress has been achieved in the development of more efficient molecular technologies for high-throughput genotyping, DNA sequencing and transcriptome profiling. At the end of 2008, an International Consortium and Illumina, Inc. launched the PorcineSNP60 BeadChip, which contains 62,163 SNP assays that uniformly cover the entire swine genome (Ramos et al., 2009b). This SNP chip provides improved power to map and dissect QTLs responsible for phenotypic variation. The high-density genotypes generated by this array can also be utilized to predict genomic breeding value that can then be used in genomic selection, a concept introduced by Meuwissen et al. (2001). This represents a unique opportunity to integrate molecular data with traditional quantitative measurements and modern computational approaches to provide accurate genetic predictions for selection in livestock. Accurate prediction values have the potential to dramatically change selection in the swine industry and the quality of the pork products. Genetic prediction can also take advantage of the high-throughput sequence data provided by next-generation sequencing methods (Meuwissen and Goddard, 2010). Recent advances in massively parallel DNA sequencing have accelerated the pace and reduced the cost of genome re-sequencing, and provide steady-state quantitative and qualitative measurements of transcriptome diversity. All of these advancements will have an important impact on our ability to understand genome biology and to better characterize the components of molecular networks that are affected by genetic variations, and to induce changes in metabolic traits, including those that affect meat quality.

Acknowledgements

The authors would like to thank the reviewers for their comments and Autumn McKnite for her assistance in preparing the manuscript.
References


Jeon, J.T., Carlberg, O., Tornsten, A., Giuffra, E., Amarger, V., Chardon, P., Andersson-Eklund, L.,
affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nature Genetics 21,
157–158.
Jonsson, P. (1963) Danish pig progeny testing results. Zeitschrift für Tierzüchtung und Zuchtwirtschaft
78, 205–252.
Karlskov-Mortensen, P., Bruun, C.S., Braunschweig, M.H., Sawera, M., Markljung, E., Enfalt, A.C.,
Hedebro-Velander, I., Josell, A., Lindahl, G., Lundstrom, K., Von Seth, G., Jorgensen, C.B., Andersson,
the near-complete mtDNA genome. Journal of Molecular Evolution 52, 302–308.
1 (MC1R) mutations and coat color in pigs. Genetics 150, 1177–1185.
map of the pig chromosome 6q32 region associated with a QTL for fat deposition traits. Animal
Genetics 37, 522–523.
breed cross populations in pigs to improve detection and characterization of quantitative trait loci.
melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits.
Mammalian Genome 11, 131–135.
teolysis in Longissimus muscle from beef, lamb and pork carcasses. Journal of Animal Science 69,
617–624.
nucleotide polymorphisms associated with ultrasonic backfat depth in a segregating Meishan × White
tiques pour les critères de comportement alimentaire dans les races Large White et Landrace
Yorkshire, Meishan, Yorkshire × Meishan, Meishan × Yorkshire, Fengjing × Yorkshire, and Minzhu ×
Résultats de quatre générations de sélection sur le potentiel glycolytique musculaire mesuré in vivo.
trait loci for androstenone, skatole and boar taint in a cross between Large White and Meishan pigs.
Animal Genetics 36, 14–22.
major locus. In: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production,
7–12 August 1994, University of Guelph, Canada, 19, 473–476.
Le Roy, P., Przybylski, W., Burlot, T., Bazin, C., Lagant, H. and Monin, G. (1994b) Etude des relations entre
le potentiel glycolytique du muscle et les caractères de production dans les lignées Laconie et


Genetic improvement in pigs began several centuries ago leading to a clear transformation from the wild pig to the domestic pig. The process of domestication began about 10,000 years ago, according to the archaeological data (Chapter 2). This was followed by breeding for specific characters and the development of specialized breeds and lines. The genetic improvement at that time was based on empirical methods or due to 'unconscious
selection’, as termed by Darwin (1859). The methods became more scientific after the discovery of Mendel’s principles and the development of genetics as a scientific discipline. A historical perspective of the early genetic improvement methods in pigs was provided by Ollivier (1976).

During the early part of the last century, the focus was on breed characteristics and physical appearance. This included the maintenance of pedigree records, herd books and prizes for champion boars and sows. Later, starting in the 1950s, genetic improvement efforts were diverted towards reduction in backfat and improvement in growth rate or days to market. However, breed characteristics and traits of physical soundness were still of utmost importance. This resulted in remarkable genetic progress in lean meat production and growth rate, while very small or negligible gains were made in reproduction traits. A more detailed overview of the genetic changes in different breeds and traits during the past century was given by Merks (2000). The majority of these changes were as a result of improvements in performance recording and genetic evaluation methods.

Modern pig breeding is evolving as a technology-based industry, making use of advancements in computing, Internet communications, biotechnology and molecular biology. In this chapter on modern genetic improvement, updated from Ollivier (1998), the principles of genetic improvement and the organization of current genetic improvement programmes will first be reviewed. This will be followed by discussion of the definition of breeding objectives, choice of selection criteria, including traits with economic and non-economic values, and their use in breeding value estimation. Detailed information is then provided for the design of breeding programmes, including the selection of specialized sire and dam lines, transfer of genetic gains in commercial populations, and adaptation of breeding programmes to meet the emerging needs of society. In the final section, the assistance that genetic markers and genomic selection can provide for further enhancing genetic progress will be discussed.

### General Principles of Genetic Improvement

The first objective of a breeding programme is to produce the ‘most improvement per unit of time’, as stated by Dickerson and Hazel (1944). They showed that selection response depends on three parameters, which may differ between the sexes, i.e. selection accuracy (\(r\), defined as the correlation between the aggregate genotype and the selection criterion), selection intensity (\(i\)) or standardized selection differential, and generation interval (\(t\)). The expected annual response (\(R_a\)), expressed in genetic standard deviation units, is:

\[
R_a = \left( i_s \rho_s + i_d \rho_d \right)/\left( t_s + t_d \right)
\]

where the subscripts refer to sires (s) and dams (d), respectively. This formula can be rewritten as:

\[
R_a = \left( i_s \rho_{sb} + i_d \rho_{bd} \right) * \sigma_G/\left( t_s + t_d \right)
\]

where \(\rho_{sb}\) refers to the accuracy of the genetic evaluation or the correlation between the breeding goal and the observations recorded on the animals, and \(\sigma_G\) is the available genetic variation. Genetic gain increases if selection intensity increases, if accuracy of estimating the genetic value increases or if generation interval decreases. Genetic variation is considered to be a given constant, although it might be differentially expressed (e.g. in heat- or disease-challenged environments).

Accuracy of the genetic evaluation depends on the availability of observations and on the heritability of the trait under scrutiny. The main basis of observations is records on an animal’s own performance. In addition, other sources of information are performance records on parents, offspring and other relatives. In practice, a combination of these sources exists. The sire can have observations on a (large) number of offspring, i.e. full and half sibs of the animal, the dam has her own (reproduction) records, and the selection candidate has its own recording of finishing traits. In a basic situation when only own performance records are available, the accuracy equals the square root of the heritability. In other words, if the heritability of a trait is high, then the
availability of own performance observations will yield a high accuracy and higher response to selection.

The breeding objective should define the main goals of genetic improvement. It must define a production environment and traits relevant for genetic improvement. The goal is then to maximize genetic change over generations or over time. An important aspect to consider is the dissemination of the genetic changes from the nucleus breeding farms down to the multiplier and commercial farms (the breeding pyramid). This process creates genetic lags (Bichard, 1971), which can be minimized by acting on the structure of the breeding pyramid (using artificial insemination (AI), for instance), as well as on the genetic level of the individuals migrating from one tier to the next, using the gene flow techniques proposed by Elsen and Mocquot (1974) and Hill (1974). An economic optimization of the whole system may also be attempted by introducing cost–benefit considerations at all levels of the breeding pyramid, as Elsen and Sellier (1978) did for determining an optimal selection policy in dam lines. The breeding goal typically includes more than one trait in the breeding objective. Multi-trait index selection relates traits from the breeding objective through economic values to traits measured for the genetic evaluation.

The rate of genetic improvement also depends upon the amount of available genetic variation ($\sigma_0$). Maintenance of genetic variation and biodiversity is also an important element of sustainable animal breeding and reproduction. The loss of genetic diversity within a breed is related to the rate of inbreeding ($dF$). Factors influencing the maintenance of genetic variation in quantitative traits in populations undergoing selection were reviewed by Hill (2000).

Modelling genetic gain and rate of inbreeding can be very helpful. Stochastic models can take into account the variability expected in genetic responses, as well as the changes in genetic variances and covariances under selection in populations of limited size. Simulation models such as those of Belonsky and Kennedy (1988) and De Roo (1988) are now increasingly relied upon in the study of pig breeding schemes. These models simulate selection in populations on an individual basis. Wray and Thompson (1990) developed the theory of genetic contributions, Bijma and Woolliams (2000) extended this theory by supplying a set of deterministic formulae predicting rate of inbreeding and genetic progress in situations with best linear unbiased prediction (BLUP) selection and overlapping generations. This set of formulae was then included in the software package SELACTION (Rutten et al., 2002), which helps breeders develop total merit selection indexes and predict response to selection.

The two disadvantages of inbreeding are loss of genetic variation and the increase in the risk of homozygosity of deleterious alleles. Crossbreeding of genetically different lines reduces the chance of homozygosity and results in heterosis. Crossbreeding is key in the set-up of modern pig production pyramids. It opens the opportunity for specialized sire and dam lines, and it opens the opportunity for specific correction of the shortcomings of a line or cross. A given dam cross can be mated with a fast-growing sire line to produce heavily muscled lean animals for a low carcass weight market or to a slow growing synthetic fatter sire line for the production of Parma ham (a type of dry-cured ham originating from the Parma region of Italy).

Organization of Genetic Improvement

**Pork value chain**

In most parts of the world, pigs are raised to produce pork. The production of pork happens through a chain of events starting from the production of breeding stock and ending with the production of pork products sold in retail outlets. The pork value chain includes all players involved in pork production, from genetics suppliers and pig producers to the slaughterhouses and the further-processing and retail outlets that bring the final products to the pork consumers (Fig. 16.1). In addition, transport companies and feed suppliers also contribute to the pork value chain. Each link in this chain has its specific and important function and the chain is as strong as its weakest link. There is also a strong interdependence between the value-chain partners. Moving forward from genetics
to production, the efficiency of production depends partly upon the quality of the breeding stock. The quality of pork at the processing plant depends upon the quality of the pigs from the production level and the breeding stock. At the same time, there is dependence from the consumer side, e.g. if there is a food safety issue with pork purchased by the consumer, it needs to be traced back to the processing plant and the farm of origin until the source of the problem is identified.

In many countries, the partners in the pork value chain work together for the competitiveness of their pork industry in the world market, while in some other countries, individual companies take up that role. In Denmark, pork production is highly integrated, and breeding, production and processing are all operated by one organization. In Canada, a national pork value chain round table has been created as a forum for interaction among producers, processors and the government, while individual breeders and breeding companies operate on their own and interact with their value chain partners. In the USA and some countries in Europe, there is no such national forum of value chain partners, but the multinational companies have an integrated system of their own, starting from the suppliers of genetics, to producers and processors. Integrated pork value chains allow breeding stock suppliers to understand the future needs of their customers and target genetic improvement programmes accordingly to provide the right type of breeding stock at the right time for the competitiveness of their respective value chains. The value chains, whether national or multinational, typically develop their own competitive strategies. These include the target markets, specific products that will bring them more value in the future, and methods to reduce costs by the integration of processes and avoiding duplication. In addition to attention to the quality of the product, it is important for the value chains to develop their product to conform with the needs of society, such as food safety, animal welfare, animal health and reduction of impact on the environment.

Genetics is the first link in the value chain. The quality and price of the final pork product depend upon the quality of the breeding stock and its ability to produce pork with minimum input costs for producers and processors. Pork value chains that have a strong breeding component should therefore be the most successful in the world.

**Pig breeding pyramid**

In most countries, pig breeding programmes operate in a three-tier pyramidal structure (Fig. 16.2). At the peak of the pyramid are the *nucleus* breeding farms that actually generate the genetic changes, followed by the next tier of *multiplier* farms that carry out specific matings or crossbreeding for the production of large numbers of females. These females are then sold to *commercial producers* for *piglet production* and *finishing* to produce market pigs that are sent to slaughterhouses to produce the pork. In addition, there are a number of variations in the tiers of the pyramid, such as separate operations for piglet production and finishing, piglet production with own multiplication, multiplication with a nucleus module, etc. The nucleus farms
actually conduct breeding and selection for the genetic improvement of specific breeds or lines. They target their selection programmes to the needs of their customers, the pork producers and processors. However, there is a time delay between the genetic improvement in the nucleus farms and the transfer of genetic gains to commercial producers through multipliers. This delay is typically 3–5 years and is called \textit{genetic lag}. It is, therefore, very important for the nucleus breeders to evaluate the future needs of producers ahead of time and decide their breeding goals accordingly.

The nucleus breeding farms typically maintain specific breeds or breeding lines. The commonly used breeds are Duroc, Piétrain, Landrace and Yorkshire. In addition, breeding companies develop their own \textit{synthetic lines} by crossing two or more breeds and further line breeding within the crossbreds for several generations to combine and stabilize the desirable characteristics. Furthermore, within a breed, specific lines are developed to advance certain traits of interest, while keeping up the main breed characteristics. Among the purebreds, Landrace and Yorkshire (Large White) are typically selected for sow productivity, while Duroc and Piétrain breeds are selected for efficiency during the finishing phase, quality of the carcass and the resulting pork. Crosses between Landrace and Yorkshire (Large White) are used by multipliers to produce F\textsubscript{1} females that are then used by piglet producers or by farrow to finish operations to produce finishing pigs. These breeds and lines derived for production of the required females are therefore called \textit{dam lines}. The breeds and lines used for mating with the F\textsubscript{1} females to produce the finishing pigs are therefore called \textit{sire lines}. In practice, there are many variations of using specific breeds as sire or dam lines. In some specific cases, Yorkshire/Large White is used as a sire line while Piétrain and Duroc are used as dam lines. In all these cases, they are selected either as sire line or as dam line for several generations. Genetic improvement within sire or dam lines is then targeted towards improvement in specific characteristics that are important to the producers and consumers.

\textbf{Role of breed societies}

The maintenance of herd book and breed characteristics by breed societies has been a long-time trend in pig breeding and still continues to be so. Breed societies are either for a specific breed or a number of breeds. These societies (e.g. the Canadian Swine Breeders Association) define the breed characteristics and maintain
pedigree records and the rules of entry of animals into the herd book. Breed societies often issue certificates of breed purity that are especially used for the export of animals. Traditionally, breed purity has been mainly determined through pedigree records. However, some breed societies (e.g. the National Swine Registry in the USA) have introduced mandatory DNA sample collection that is used for breed purity tests in disputable cases. The roles of breed societies have evolved over time. Increasingly, they have moved to the maintenance of DNA banks, supporting research, education of members and promotional activities. Many breed societies provide services for performance recording and genetic evaluations, either directly or through collaboration with other institutions, and play an important role in genetic improvement through national improvement programmes in their respective countries.

Role of breeding companies and national improvement programmes

Increasingly, pig breeding has moved from purebred breeding and genetic improvement through breed societies to the breeding of specific lines and associated services by breeding companies. In addition to supplying breeding stock, breeding companies or professional breed societies provide a range of services – from breeding and multiplication to production. Breeding companies typically can be faster in the extensive adoption of new technology, such as use of ultrasound and electronic equipment, database technology, molecular genetics and complicated statistical methods for genetic evaluation and selection. Breeding companies have transitioned pig breeding to a technology-based industry and play an important role in supporting research at universities and public institutions (Knap et al., 2001).

Many breeding organizations, whether private companies or professionalized breed societies, operate in more than one country and even have nucleus breeding herds in several countries. An overview of the relative market shares of breeding companies and national programmes covering several European countries, Canada and Australia was given by Brascamp (1994). The review suggested that national breeding programmes still had an important role in many countries at that time, in contrast with the poultry situation, for instance, where hardly any national involvement existed. During the past decade, there has been further consolidation in the pig breeding industry, as in other major industries. The number of breeders has declined, while the total number of breeding pigs has remained relatively steady, and larger breeding companies have picked up higher market shares. A list of the main breeding companies and breed societies and their relative market shares was summarized by the Working Group on the ‘FABRE Technology Platform’ in their ‘Strategic Research Agenda’ published in 2008, and is given in Table 16.1. The table shows the important role that is played by breeding companies in Europe, followed by those in North America, with a diminishing role for breed societies and herd book organizations.

In addition to breeding companies, national improvement programmes continue to exist in several countries, following the early Danish example, where all three tiers of the pyramid are integrated into a national programme. In such structures, with a dispersed nucleus rather than a single company nucleus, animals from several breeding farms are evaluated at central testing facilities, and the selected animals may then be used in the national herd by AI. Various routes have been taken in countries that have tried to maintain national programmes. The extremes are on one side a closed selection system, where a breeding company system is, so to speak, extended to the whole country, as in Denmark for instance, and on the other end an open selection system, for example in Canada, where genetic evaluation is carried out across several breeding organizations while they act rather independently in using those evaluations for genetic improvement. In that respect, France may be an extreme case, where most breeding companies are incorporated into the national programme.

Role of reproduction technologies

The progress in genetic improvement over the past years has been strongly supported by
developments in assisted reproduction technologies (ART), as described in Chapter 11. One of the most used among these is AI, which allows easier dissemination of genetics from nucleus to commercial farms, with considerably lower risk of disease transmission and a smaller genetic lag. The use of AI is especially cost-effective for transmission of genetics between countries, while transport of live animals can be problematic. In addition to helping dissemination of superior genetics over a wider population base, AI also helps health control and herd management, and it improves connectedness for more accurate across-herd genetic selection (Haley, 1991; Mathur et al., 2002; Fouilloux et al., 2008). Before 1991, the use of AI remained at a low level (Sellier and Rothschild, 1991); however, within the following decade the use of fresh-semen AI fully matured and contributed significantly to across-herd genetic evaluation and selection in national and multinational breeding programmes (Knap et al., 2001).

Furthermore, new developments in sperm sexing technology allow for the production of either male or female offspring for sire or dam lines, and this can be very useful for the nucleus and multiplication tiers. The technique involves the separation of spermatozoa carrying either X or Y chromosomes (Maxwell et al., 2004; Garner, 2006). The sorted sperm can then be used for AI or embryo transfer. Practical application of this technology in pigs is still very limited, and depends upon the efficiency of the separation technique and the resulting conception rate. Current speed of separation requires the use of deep intrauterine insemination techniques, such that much lower doses can be applied.

Another such reproductive technology (Chapter 11), embryo transfer (ET), has been used in pigs for the last 60 years (Kvasnitsky, 1950). Since its initial application, further development of ET has been hampered by the availability of methods to effectively cryopreserve swine embryos, unlike the situation in cattle.

Table 16.1. Estimated market shares of different breeding organizations in different countries. Source: FABRE TP (2008).

<table>
<thead>
<tr>
<th>Organization</th>
<th>Developed countries (%)</th>
<th>Worldwide (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EU-based organizations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIC (= Genus), UK</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>TOPIGS, Netherlands</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Danbred, Denmark</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Hypor-Genex, Netherlandsa</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>JSR, UK</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Seghers Rattlerow, Belgium-UK (incl. Newsham, USA)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Herdbooks/Nucleus, France</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>ACMC, UK</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Herdbook, Poland</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Herdbooks, Italy</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Herdbooks, Germany</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>BHZP, Germany</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>France Hybrides, France</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Herdbooks, Eastern EU</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Total of EU based organizations</td>
<td>60</td>
<td>28.5</td>
</tr>
<tr>
<td><strong>Non-EU-based organizations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monsanto, USAa</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Smithfield Genetics, USA</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Genetiporc, Canada</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>National Swine Registry, USA</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Canadian National Breeders, Canada</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Total of non-EU based organizations</td>
<td>21</td>
<td>8</td>
</tr>
</tbody>
</table>

*aHendrix Genetics acquired the pig breeding part of Nutreco (Euribrid: Hypor–Genex) in June 2007.*

*bNewsham (USA) acquired the pig breeding part of Monsanto in September 2007.*
(Youngs, 2001). There are two main methods of ET, surgical and non-surgical. Surgically recovered embryos are typically too old for successful non-surgical implantation. The problem of non-surgical ET is the difficulty in recovering embryos from donor sows and crossing the cervical barrier after the fertile phase has passed. Recovery of embryos requires flushing, which is technically too demanding in sows, or requires slaughter of the donor sows. The procedure also actually requires sows that produce a sufficient number of viable embryos for ET. More recent developments in the technology, especially in non-surgical and endoscopic procedures, may overcome some of the problems for the future wider use of ET in commercial practice (Hazeleger and Kemp, 2001; Martinez et al., 2004); currently, the technique is mainly used for research. If fully successful, there are distinct advantages of the technology. Especially, embryo transfer allows for transfer of 100% of the genome across tiers in the pyramid, while AI transfers only 50% of it. In addition, ET has reduced transportation cost and there is a lower risk of the introduction of diseases, especially compared with the use of live animals. In view of these considerations, there is high demand for use of this technology by multinational breeding companies and for research on cloning.

The use of cloning technologies was initially problematic in pigs. However, a number of technical difficulties have been overcome to some extent, leading to higher success rates (Onishi et al., 2000; Polejaeva et al., 2000; Betthauser et al., 2001). Offspring produced by the new methods grow normally and are fertile (Onishi, 2002; Shibata et al., 2006). Furthermore, the use of parthenogenotes in place of embryos to establish pregnancy and promote the development of a single cotransferred embryo offers new opportunities (Kawarasaki et al., 2009). It has been postulated that cloning will allow faster transfer of genetically superior animals, even directly to commercial producers. In this way, the technology can offer the possibility of reducing the genetic lag between nucleus, multiplication and commercial tiers of the breeding pyramid, and even eliminate the need for one or two tiers (Visscher et al., 2000; Niemann et al., 2003).

A combination of reproduction technologies with advancements in molecular genetics is particularly expected to deliver a powerful impact in increasing the rate of genetic progress and to allow for efficient use of genetic variation and reduce genetic lag.

Definition of an Overall Breeding Objective

Choice of breeding goals

An essential step in any genetic improvement programme is the definition of the overall breeding objective or breeding goal. This includes traits of interest, direction of improvement and relative significance of each trait. Initially, the main breeding objective of breed societies was to maintain breed characteristics, and selection was mainly based on exterior traits. Over the years, breeding organizations have been focusing their breeding goals on lowering the cost price of pork production and the quality of pork in terms of processing and consumer appreciation. It is important to note that the breeding objective is actually implemented at the nucleus level, but it is targeted to the production level and to future markets.

In commercial breeding, breeding goals are defined separately for sire and dam lines. The sire lines are mainly selected for genetic improvement in production traits such as growth rate, carcass quality, meat quality, feed conversion and conformation. The breeding goal for dam lines typically includes the same production traits, but in addition female reproductive traits, such as litter size, piglet survival, age at puberty, conception rate, number weaned and weaning to oestrus interval. In more detail, traits like farrowing survival, lactation survival, number of teats, litter weight, gestation length, longevity, age at first service, number stillborn and litter mortality may also be considered in the breeding goal for dam lines. It is important to note that genetic progress in individual traits is inversely related to the number of traits included in the breeding goal, and that negative or positive genetic correlations between traits exist. Therefore, it is important to choose traits with great care.
The choice of breeding goals and successful breeding have led to the genetic improvement of economically important traits such as daily gain, backfat thickness, feed efficiency and litter size, especially during the last decade. However, this is not enough for the future. Breeding goals have been, or are presently being, set up more broadly to include traits that are important to society. There is increasing societal pressure and desire from several pig producers to include traits such as the safety and improved quality of pork, the health and welfare of pigs, and the environmental impact of pig production (Verbeke and Viaene, 2000; Kanis et al., 2005). Further, the health of the pigs is becoming more important owing to the concentration and increasing scale of pig production (Merks, 2000; Bishop and MacKenzie, 2003).

**Derivation of economic weights**

Once the breeding goal traits have been chosen, the question arises of giving each trait its proper weight in an overall objective, also termed aggregate genotype. A common practice is to use profit equations of the form:

\[ P \text{ (profit)} = R \text{ (returns)} - C \text{ (costs)} \quad (16.3) \]

By taking partial derivatives of \( P \) with respect to the \( n \) traits included as goal traits in \( R \) and \( C \), economic weights are obtained and an aggregate genotype (\( H \)) is established of the form:

\[ H = w_1A_1 + w_2A_2 + \ldots + w_nA_n \quad (16.4) \]

where \( A_1 - A_n \) are the breeding values for the traits included in the breeding goal and \( w_1 - w_n \) are their respective economic weights. This has been the approach in the classic paper of Hazel (1943) on selection index theory. However, as pointed out by Moav (1973), the economic weight so derived depends on the perspective taken in defining \( P \). This dilemma of different economic weights for different perspectives in production, under unchanged economic conditions, was resolved by Smith et al. (1986), who suggested imposing two conditions. The first is that fixed costs, incurred in running the production enterprise, should be included with other costs and expressed per unit of output. The second condition is that any extra profit from genetic change that could also be obtained by altering the size of the enterprise should not be counted. The application of these two conditions also shows that the economic weights are those obtained in considering cost per unit return, i.e. \( C/R \) instead of \( R - C \). More complex economic models as possible alternatives were given by Amer et al. (1994).

Economic weights can also be derived from a system of equations known as a bio-economic model, or profit function, which is a function of phenotypic traits and management variables contributing to revenue and/or costs (Tess et al., 1983), as well as from the model of De Vries (1989). The economic weight of a trait is calculated from the change in predicted profit based on a single unit change in that trait, holding all other traits constant, or directly from the partial derivatives evaluated at the relevant population mean (Moav and Hill, 1966). A similar profit function approach was used by Quinton et al. (2006) to derive economic weights for sow productivity traits. The advantage of using the profit function approach is that economic weights can be recalculated for different production systems, market requirements or population trait levels with relative ease. For example, as litter size increases, the relative economic value of increases in litter size decreases but that of piglet survival increases (Knol and Mathur, 2009). In this way, the economic values can be different for different population means at the commercial levels, and the profit function approach allows for adjusting them accordingly.

In some cases, genetic improvement is desired in traits that do not provide a direct economic return but are important for non-economic reasons. Such traits have non-economic values. For example, improvement in physical soundness leads to economic benefits such as reduction in veterinary costs and better sow longevity, but there are additional non-economic benefits, such as those for pig welfare and ethics. In this case, the trait has an economic as well as a non-economic value. The relative importance of economic and non-economic values for different traits was described by Kanis et al. (2005), as given in Table 16.2. Non-economic values are also derived in monetary units and included in
the selection index calculations. Their estimation is based on relative genetic progress in traits with economic and non-economic importance, as described for societally important traits by Kanis et al. (2005).

### Adaptation of breeding objectives to changing economic conditions

Trends in breeding objectives have been discussed in detail by Ollivier et al. (1990), Haley (1991) and Merks (2000). Past experience shows that continuous re-evaluation of the economic weights applied to each trait is needed. In spite of efforts made to foresee changes in economic conditions, this exercise can never be entirely satisfactory, because of inevitable discrepancies between present objectives and future production conditions. However, this should be of little concern as long as the evolution of these conditions is slow and gradual, because using slightly ‘false’ economic weights has a limited impact on overall selection efficiency in most cases (Vandepitte and Hazel, 1977).

### Choice of Selection Criteria and their Use in Breeding Value Estimation

A distinction is classically made between traits that are considered as objectives for improvement, and traits that are actually used in ranking the male and female candidates, which are termed criteria of selection. In pig breeding, the two sets only partly coincide, and each set includes a fairly large number of traits. It is the purpose of performance recording programmes to define the measurements to be used as selection criteria and to organize the collection and processing of the corresponding data. Such programmes will now be reviewed briefly, as more details can be found in several textbooks on pig breeding, e.g. in Sellier (1986) and Glodek (1992), among others. A historical presentation of performance recording and genetic evaluation since the beginning of the century has also been given by Sellier and Rothschild (1991) and Merks (2000).

Reproduction performance has for a long time been assessed through on-farm litter-recording systems, including litter size at birth and at weaning, sometimes completed by litter weights, these traits being viewed as the most important reproduction traits (see Chapter 10 for more details). Production traits, namely growth rate, feed efficiency and carcass measurements (see Chapter 14), were initially recorded almost exclusively in central testing stations, built on the Danish model, as progeny-testing stations became progressively available to pig breeders in most countries from the early 1920s to the late 1950s. An important step has been the advent of techniques allowing a fairly accurate evaluation of body composition on the live pig, starting with the metal ruler of Hazel and Kline (1952), and followed by the

---

**Table 16.2.** Example of the allocation of the relative economic and non-economic values of different traits. Source: Kanis et al. (2005).

<table>
<thead>
<tr>
<th>Trait/characteristic</th>
<th>Economic</th>
<th>Non-economic</th>
<th>Explanation of non-economic value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate</td>
<td>++++</td>
<td>+</td>
<td>Traits reflect pig welfare and health, and are related to a possible manure surplus</td>
</tr>
<tr>
<td>Feed efficiency</td>
<td>++++</td>
<td>+</td>
<td>Mainly economically important</td>
</tr>
<tr>
<td>Meat percentage</td>
<td>+++++</td>
<td></td>
<td>Better quality is desired but not paid for</td>
</tr>
<tr>
<td>Meat quality</td>
<td>++</td>
<td>+++</td>
<td>Mainly economically important</td>
</tr>
<tr>
<td>Litter size</td>
<td>++++</td>
<td></td>
<td>Replacement costs of sows are relatively low, but citizens attach value to longevity</td>
</tr>
<tr>
<td>Sow longevity</td>
<td>+</td>
<td>+++</td>
<td>Apart from the economic reasons, impaired animal health is a major concern for many citizens and farmers</td>
</tr>
<tr>
<td>Piglet vitality</td>
<td>+++</td>
<td>++</td>
<td>Deviant behaviour does not cost much, but it is undesired by many citizens and farmers</td>
</tr>
<tr>
<td>Leg condition</td>
<td>+++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>Disease resistance</td>
<td>+++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>Aggressiveness</td>
<td>+</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>Stereotypic behaviour</td>
<td>++++</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
development of ultrasonic machines and more sophisticated technologies such as computer tomography (Kolstad, 2001). This has opened the way to central performance testing stations for young boars, replacing progeny-testing stations, and also to on-farm testing programmes. Boar testing stations usually record average daily gain over a given live-weight interval, food conversion ratio over the same interval, and backfat thickness at the end of test, whereas on-farm tests generally include only daily gain (expressed as days to a given weight) and backfat thickness, lean depth and intramuscular fat at a given weight measured through ultrasound. More recently, meat and fat quality traits have been introduced into testing programmes, based on measurements made in slaughterhouses, such as ultimate pH, colour and water-holding capacity. More details on the meat and carcass traits of interest can be found in Chapter 15, and especially on advances to be expected regarding live animal measurements based on tissue biopsies.

As shown by Hazel in his classic paper of 1943, pig improvement is essentially a multiple-trait selection problem, which can be solved by combining performance records into a linear index (I). This index is a predictor of breeding value for the overall breeding goal, defined so as to maximize the correlation between I and the aggregate genotype (H), described in the previous section. This principle eventually became almost universally applied in pig breeding, use being made of estimates of the necessary genetic parameters (heritabilities and genetic correlations), as presented in other chapters of this book. The performance records entering the selection indices, collected either on farm or in central test stations, are generally expressed as deviations from contemporary group means. In the meantime, selection index theory was being refined through the development of mixed model methodologies, taking account of unequal information among candidates and unknown means, and providing best linear unbiased prediction (BLUP) of breeding values. Application of the theory then involved a two-step procedure, first the estimation of individual trait breeding values and, in a second step, the application of relative weights to those breeding values for deriving the estimation of H. As noted by Hazel et al. (1994), such a procedure brings more flexibility in the adaptation of the economic weights to any breeding system, without the need to recalculate individual breeding values.

The use of BLUP methodology in a national pig breeding programme was done for the first time in Canada in 1985 for growth and backfat on-farm genetic evaluations (Hudson and Kennedy, 1985), and was later extended to maternal traits. In the USA, a cooperative project between Purdue University and purebred associations, termed the Swine Testing and Genetic Evaluation System (STAGES), was also initiated in 1985, and reported its first evaluations in 1986 (Stewart et al., 1991). STAGES and the Canadian programme were designed as on-farm evaluation systems, with the added capability of across-herd analysis performed at longer time intervals. Progress in speed of data transmission and computing has now made it possible to shorten the computation intervals and to make centrally produced BLUP values available to breeding farms overnight. Such systems have been applied by several European countries, in most cases with the implementation of multivariate prediction-estimation software such as PEST (Groeneveld et al., 1990), ASREML (Gilmour et al., 2002) and PIGBLUP (Hermesch and Crump, 2006). Most breeding organizations also apply similar procedures. In addition, there has been increased recording of data, both in terms of number of traits and number of pigs recorded. In many cases, the data are now transferred over the Internet from the pig farms to the genetic evaluation centres, and the genetic evaluation procedures are increasingly automated to compute estimated breeding values (EBVs) using millions of records and release them in a few hours. Computer software and web applications have been made available for breeders to define their breeding goals, develop customized selection indices, predict expected genetic gains and identify selection candidates (e.g. Groeneveld, 2004).

**Design of Breeding Programmes**

**Breeding for sire lines**

The full breeding objectives are typically met by a commercial crossbred pig produced by crossing different sire and dam lines. The contribution
of the sire line is through: (i) maximized AI performance (libido, semen quantity and semen quality); (ii) improved pre-, peri- and post-natal survival; (iii) improved production; and (iv) improved carcass and meat quality traits. Differences between lines are exploited to adapt to changing market conditions, such as differences in robustness, muscularity and gain.

Production traits in pigs are measurable on both sexes before breeding (growth rate, food conversion and backfat thickness), or after slaughter (lean content, lean and fat characteristics). Individual and sib information therefore provide the essential criteria for selection in most breeding programmes, as additional culling on progeny performance has been known for a long time to be ineffective (Dickerson and Hazel, 1944; King, 1955). The expected annual response \( R_a \) in Eqn 16.2 is then maximized when \( i/t = (i_s + i_d)/(t_s + t_d) \), i.e. the annual selection intensity is at its maximum. With the common demographic parameters of the pig, such as first offspring at 1 year of age, six candidates per litter at breeding age and a mating ratio (sow:boar) of 15 under conditions of natural mating, it can be shown that the optimal \( t_s \) and \( t_d \) in individual selection are close to 1 year, which yields a maximum \( i/t \) ratio of 1.75 (Ollivier, 1974). For a trait of medium heritability \( h^2 \), \( h^2 = 0.30 \), a maximum annual response of nearly one genetic standard deviation (SD) can be obtained with such a method of selection. When information on slaughtered sibs is used in selection, the value of \( i/t \) has to be adapted to the ensuing reduction in the number of candidates remaining and the change of sex ratio among them (Ollivier, 1988a). The loss in \( i/t \) incurred in any sib- or combined sib–individual-testing scheme can then be set against the increased accuracy of evaluation.

It should be borne in mind that the responses predicted in such a theory refer to an idealized situation of a large population of sows, farrowing simultaneously at fixed intervals of 6 months, and that selection is assumed to be carried out among a large number of independent observations. In practice, however, those assumptions are not fulfilled, as farrowings generally occur quasi-continuously in breeding herds, and comparisons between candidates are made within ‘batches’ containing a limited number of full-sib groups. The resulting reduction in selection intensity for the nested family structure typical of the pig was worked out by Meuwissen (1991). It can be shown that in a herd producing 100 gilt litters per year in 17 batches, the expected maximum of \( i/t \) in individual selection is reduced by approximately 15%, to a value of 1.5, with the same demographic parameters as above (Ollivier, 1988b). Even the latter objective will rarely be achieved in practice because of incomplete testing or of culling for reasons other than performances. Actual values of \( i/t \) achieved in individual on-farm tests rarely exceed one. Considerably lower values have been reported in national programmes emphasizing either on-farm tests or central-station family selection (Table 16.3). It can be noted that the individual selection schemes simulated

<table>
<thead>
<tr>
<th>Breeding structure</th>
<th>Selection criterion</th>
<th>Selection intensity ((i))</th>
<th>Generation interval ((t))</th>
<th>(i/t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breeding company</td>
<td>Individual(b)</td>
<td>1.28</td>
<td>1.25</td>
<td>1.02</td>
</tr>
<tr>
<td>National programme</td>
<td>Individual(c)</td>
<td>0.48</td>
<td>1.92</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Sib and progeny(d)</td>
<td>1.46</td>
<td>2.00</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Progeny(e)</td>
<td>0.91</td>
<td>2.00</td>
<td>0.46</td>
</tr>
</tbody>
</table>

\(a\) For \(i/t = 0.5\), using an individual selection index, expected annual responses of 5 g in average daily gain and −0.4 mm in backfat thickness have been given by Sellier (1986, p. 201). The ranges of annual genetic trends reported by Sellier and Rothschild (1991) – 3 to 6 g and −0.1 to −0.4 mm, respectively, based on central testing records – indicate that the corresponding \(i/t\) would generally be below 0.5.


\(c\) Kennedy et al. (1986): 1977–1983; growth and backfat index in retrospect (Canadian Yorkshire and Landrace).

\(d\) Christensen et al. (1986): 1980–1985; index on growth, lean and meat quality (Danish Landrace and Yorkshire).

by De Roo (1988) and Belonsky and Kennedy (1988) do not fully exploit the potential of such schemes either, as the values of \( i/t \) realized are generally well below one.

With the advent of BLUP (see the previous section), records from all relatives, such as sibs, cousins and ancestors, can be used to predict breeding values. Genetic response is then expected to increase, following the increases in average selection accuracies (\( \rho_1 \) and \( \rho_2 \)) of Eqn 16.2, predicted from standard index selection theory, and knowing that those increases are partially offset by the decreases expected in selection intensities, because of the high correlation between BLUP values of relatives. A further increase of response, however, more difficult to predict because it depends on the breeding structure, results from better estimation of the fixed effects, which allows across-farm (or station) evaluation, and also from the fact that it takes into account genetic trends (Sorensen, 1988). Another advantage of BLUP is to allow sequential culling, whereby sows and boars are culled on the basis of their estimated breeding values, with the best remaining longest in the herd. The simulation of Belonsky and Kennedy (1988) shows that this culling scheme reduces length of generation intervals by 16–22%, relative to strictly individual selection. This advantage actually depends on the herd replacement policy, which in that study implied culling sows after a maximum of five litters. Hagenbuch and Hill (1978) have shown that, with a more stringent policy of keeping sows only for a maximum of two litters, the gain from sequential culling is much reduced, to about 2–3%. Overall, the advantage of BLUP evaluation over individual selection is in a range of 10–30% for traits of high-to-moderate heritability, i.e. for most production traits, as shown by various simulations (Belonsky and Kennedy, 1988; Sorensen, 1988; Röhe et al., 1990).

The gains from BLUP selection are to some extent counterbalanced in small populations by a higher increase in inbreeding than under individual selection, a tendency that is also enhanced by sequential culling, as shown by Belonsky and Kennedy (1988). Various methods to restrict inbreeding without significant loss of response were proposed by Toro and Pérez-Enciso (1990). Another strategy, recommended by Brisbane and Gibson (1995), is to include genetic relationships in selection decisions, assuming a given value of a unit of inbreeding relative to a unit of genetic gain. Meuwissen (1997) developed algorithms for selection that maximize response to selection with a constraint on the rate of inbreeding by limiting the average genetic relationship among selected individuals.

**Breeding for dam lines**

As shown in Table 16.2, reproduction traits have to be included in the breeding objective in crossbreeding systems with specialized lines, apart from the breeds or lines that only serve to produce the terminal boars. However, until 1995, little attention was paid to such traits in most breeding programmes. The situation has changed dramatically, as a combination of theoretical (see Table 16.4) and experimental results (see Chapter 10) has shown that litter size can be successfully improved by selection, and also that economic conditions have made such selection increasingly worthwhile (Haley et al., 1988; Ollivier, 1988b).

The efficiency of simultaneous selection for reproduction and production traits has been extensively investigated, either by using the index selection approach of Smith (1964) or by more elaborate methods based on the economic returns of the entire crossbreeding system (Elsen and Sellier, 1978), life-cycle economic efficiency (Smith et al., 1983) or stochastic models in closed dam lines (De Vries et al., 1989). The general conclusion of those studies is that some benefit is expected from

### Table 16.4. Predicted annual selection response in litter size (number born per litter).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual response</strong></td>
<td>0.25</td>
<td>0.47</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Selection criterion</strong></td>
<td>Dam</td>
<td>Dam and family</td>
<td>Dam and family</td>
<td>Dam and family</td>
</tr>
<tr>
<td><strong>Population size</strong></td>
<td>Large</td>
<td>Large</td>
<td>100 sows</td>
<td>400 sows</td>
</tr>
</tbody>
</table>
the inclusion of the reproduction objective ($H_1$) in addition to the production objective ($H_2$) in specialized dam lines. The benefit will increase with the relative importance ($a$) of $H_1$ in the overall breeding objective (as defined in Table 16.2), with the accuracy ($r_1$) of evaluating $H_1$ relative to $H_2$ ($r_2$), and with an increasingly unfavourable genetic correlation ($r$) between $H_1$ and $H_2$. A useful approximation of the relative benefit ($RB$) in terms of those parameters was proposed by W.G. Hill (in Webb and Bampton, 1987):

$$RB = \left[ (x^2 + xr + 0.5)^{1/5} + 0.5 \right]/(x^2 + 2rx + 1)$$

(16.5)

in which $x = ar_1/r_2$.

In the past 15 years most, if not all, breeding organizations have adopted selection on litter size (Rydhmer and Berglund, 2006). As the larger breeding organizations are competing breeding companies, exact estimates of genetic trends are difficult to obtain. Merks (2000) reported trends on the order of 0.20 piglets total born per year, while the breeding programme (Anonymous, 2009) reported 0.44 piglets alive at day 5. Genetic trends appear therefore to be in agreement with the estimates of Table 16.4, with the Danish programme close to the theoretical maximum of Avalos and Smith (1987). Selection, however, is not for a single trait, and should include other fertility traits, such as survival, mothering ability and interval weaning to oestrus. The decreasing economic value of litter size is interesting. The concept of reduction in economic value of litter size with increase in sow productivity was further demonstrated by Quinton et al. (2006), using a profit function approach. It has been shown that the relative importance of litter size for the finished pig market decreased from 64% of the total breeding value, when the average litter size was eight pigs, to 29%, when the average litter size was 20 pigs. At the same time, the economic value of perinatal survival increased from 17 to 42%, and that of survival to weaning increased from 7 to 18%. In this approach, using the Canadian data, it has been further shown that the relative importance of litter size for piglet weaning weight increased from 22 to 41% as the average litter size increased from eight to 20 pigs, whereas that of perinatal survival increased from 12 to 22% and that of survival to weaning increased from 5 to 9%. These results further highlight the significance of piglet survival in increasing gains from start to finishing.

Though most studies have so far concluded that litter size is genetically uncorrelated with growth and carcass traits (see the review of Haley et al., 1988), there are indications that this might not be a general rule, as shown by the indirect responses observed in the selection experiment analysed by Kerr and Cameron (1994), or by recent estimates of genetic correlations between reproduction and production traits (see Chapter 10). Another approach to this correlation is the observation that the number of ovulations has increased more than litter size, creating a situation of uterine crowding, with negative consequences for survival chances of the fetuses and possible negative consequences for the development of the surviving fetuses; this may lead to a reduction in muscle fibre development, lower birth weights and consequences for finishing gain and carcass and meat quality traits (Foxcroft et al., 2004).

### Selection based on crossbred performance

In addition to selection and performance recording within sire or dam lines, there are distinct advantages in recording data on crossbreds and selection on crossbred performance. The two main reasons are: (i) the breeding objective is based on the crossbred level, crossbred data will show additive and non-additive expression of genes and pure-line parents can be selected that maximize performance at this level; and (ii) in smaller dam lines it adds records to improve the accuracy of BLUP estimates.

Breeding programmes, therefore, need to combine purebred information with crossbred information. Predicted responses in a cross-breeding scheme implicitly assume a genetic correlation of 1 between purebred and crossbred performance. The use of crossbred information, as in reciprocal recurrent selection schemes, offers some advantages compared with pure-line selection, especially for traits showing large non-additive genetic variation, as reviewed by Wei and van der Steen (1991). Two specific parameters are needed
to evaluate the efficiency of such a type of selection, namely the genetic correlation between purebred and crossbred performance \( r_{pc} \) and the crossbred heritability \( h^2_c \). Purebred and crossbred information should in fact be combined in an optimal way in order to maximize genetic response in crossbreds, which can be done within the general framework of index selection theory (Wei and van der Werf, 1994). For instance, the crossbred paternal half-sib family mean may be combined in such a way with the purebred information on paternal half sibs, full sibs and own performance. As emphasized by Wei and van der Werf (1994), such a combined selection is always superior to pure-line selection when the testing of crossbreds is not at the expense of testing purebreds as, for instance, when a crossbreeding structure exists through which crossbred information is collected for management purposes. In such situations that apply to pig breeding (see for instance Fig. 16.2), crossbred information may indeed come as an addition to purebred data, and the increase that can be expected in genetic response will then depend on a proper evaluation of the two genetic parameters, \( r_{pc} \) and \( h^2_c \).

Bijma et al. (2001) added rate of inbreeding to this approach and analysed short-time response and long-term genetic contributions using a deterministic approach. For their set of parameters, combining purebred and crossbred performance yielded a better result than the use of purebred information only for the long term.

Except for the theoretical approach as outlined here, there are practical and quantitative advantages of using crossbred data and/or data from different environments. Automated data recording has increased dramatically over the past years. Large volumes of fertility data and slaughter-line data are readily at hand; the advantages are that the performance traits are as close to the commercial reality as possible and that the traits are observed under crossbred and field conditions. The main limiting factor is the availability of accurate pedigree information for this type of animal. Investment of breeding organizations in proper recording of sire and dam of a crossbred animal yields access to these crossbred field data; DNA technology can help here.

**Breeding for specific environments**

The cost of running an effective breeding programme dictates the necessity of a large sales volume and with that a global approach. The pork value chain shows the worldwide interest in lowering the cost price of meat, but it also shows regional differences in production environment, products and societal constraints. The net result is that the breeding goal should be defined at the crossbred level under local environment, whether that is large-scale production in the Americas or family farms in north-western Europe. The result might be line differentiation for different markets, or robust and flexible lines that have the ability to adapt to differing environments.

In the generalized situation, a trait is defined per environment. If the genetic correlation between two production environments is lower than 0.4–0.6, then separate lines have to be developed (Mulder and Bijma, 2006). Bloemhof et al. (2008) showed differences between sow lines in reaction to environmental temperature at time of insemination. One sow line loses farrowing rate and litter size at temperatures above the upper critical level of 20°C, while the other does not. The authors suggest that genetic variation for the slope above the upper critical temperature may exist. This approach can be extended towards the general adaptive ability of the animal; this ability can then be estimated and added to the selection index using, e.g. reaction norm models (Knap and Su, 2008).

**Transfer of genetic gains in commercial populations**

Genetic gain starts with the identification of single superior animals and should result in the presence of the superior alleles that are present in these animals in as many crossbreds as possible. The female reproductive step is quite slow. One sow can produce around nine productive female offspring per year. Assuming a replacement rate of 40%, one pure-line nucleus sow can maintain 22.5 pure-line multiplication sows, which use the semen of a second dam line to produce 202.5 crossbred gilts. Assuming the same 40% replacement, a sow herd of 500
commercial sows can be maintained to wean 15,000 finishers, assuming 30 piglets per sow per year. When \( F_2 \) production is applied, that is, the commercial sow is a three-way cross or a backcross or a rotational cross, one purebred sow can yield her genes to well over 300,000 finishers.

Let us assume the optimal situation for pork production to be one purebred sire line and a two-way cross on the dam side. This structure could be adapted by creating a crossbred terminal sire, with improved male reproductive performance (better libido, better semen quality) because of heterosis advantage. The production traits, which should be transferred to the offspring, will not benefit from the crossbreeding advantage. The negative side, the cost of an extra generation, will, in most situations, not be offset by the fertility advantage. A second argument in favour of using a crossbred boar could be that the traits of two existing sire lines are on both sides of a market optimum, e.g. one sire line is too muscled without enough gain and the other sire line is the reverse. A crossbred boar will then additively average the two sire lines.

Because breeding programmes are more and more international, the relevance of health status increases. Pure-line production tends to take place at high-health farms to facilitate dissemination of genetic trend.

**Breeding for particular objectives**

‘Licence to produce’ is increasing in relevance. Society starts to question the direction in which animal breeding changes the pig populations. Neeteson et al. (1999) discuss this field across species, and Kanis et al. (2005) discuss, for pigs, the following areas for inclusion of societal traits in the breeding objective: pig welfare and health, ecological effects and natural resources, and healthiness and the sensory quality of pork.

These society concerns lead to traits on animal integrity, e.g. castration, tail docking and teeth clipping. Non-castration may potentially lead to the presence of boar taint in entire males; the underlying substances of boar taint are androstenone, a pheromone, and skatole, a degradation product of tryptophan. Genetic selection against boar taint to completely eliminate the need for castration is possible (Merks et al., 2009). This would also take advantage of better feed efficiency from entire males. This should be done carefully to overcome the potential risks, such as: (i) that it will influence the hormonal balance of the animal, causing male and female reproduction reductions; (ii) that a production change towards non-castration will disrupt the economic value system, because entire males are far leaner than castrated animals and, therefore, production animals might suddenly be too lean for the average market; and (iii) that male behaviour (aggression and mounting) will disrupt the finishing phase. The latter problems are also the drivers behind tail docking and teeth clipping issues.

Theory on social behaviour was first developed by Griffing (1967). The principle is that the phenotypes of pen mates influence the expression of the trait of the selection candidate.

\[
P_i = A_{D,i} + E_{D,i} \quad (16.6)
\]

\[
P_i = A_{D,i} + E_{D,i} + \sum_{j \neq i} \frac{A_{S,j}}{n_{ij}} + \sum_{j \neq i} \frac{E_{S,j}}{n_{ij}} \quad (16.7)
\]

Equation 16.6 is the classic expression, where \( P \) denotes the phenotype of selection candidate \( i \), \( A_{D,i} \) is the direct genetic effect of \( i \) and \( E_{D,i} \) is the environmental effect for animal \( i \). Equation 16.7 extends this approach towards the social influence of group members by including \( A_{S,j} \) as the social genetic effect of group member \( j \) and \( E_{S,j} \) as the social environmental effect of group member \( j \). The total genetic merit of an animal is then the sum of its additive merit for the trait plus \( (n-1) \) times the genetic influence on its group members, where \( n \) equals average pen size (Bijma et al., 2007). First genetic parameter estimates show considerable genetic variation in social behaviour effects for the traits gain and feed intake (Bergsma et al., 2008). Parameter estimates can be found in Chapter 10. An open question is whether the estimated social effects link with behaviour, as encountered in tail biting or male sexual behaviour. Estimation of genetic parameters requires large well-defined data sets with genetically mixed pens. Animal behaviour will become more and more relevant, because labour spent per animal decreases for economic reasons, and group sizes increase because of societal concerns about animal welfare.
Another important trait is uniformity. Uniformity at slaughter is: (i) relevant to reduce the number of pulls and, with that, the average usage time of finishing rooms and uniformity in birth weight; and (ii) valued because it reduces the percentage of small piglets, which have an increased risk of mortality. The latter example can be seen as a trait of the sow and, with that, as the expression of the genes of the sow. The first example is more complicated: how to reduce variation in a group with selection on the genotype of individuals. The classical evolutionary answer, natural selection, is stabilizing selection, which removes some undesirable alleles and increases homeostasis of the developmental process. A possible way out for human-directed breeding programmes is the assumption of genetic heterogeneity of environmental variance or, in other words, to assume that part of the environmental factor in the normal genetic evaluation (Eqn 16.4) is heritable and identifies the ability of the animal to cope with its environment (Mulder et al., 2007). Implementation in the breeding programme requires an economic value. For variance (or uniformity), this economic value is much higher if the trait shows an economic optimum (Mulder et al., 2008).

Other examples of particular objectives are the elimination of genetic abnormalities, such as the existence of chromosomal aberrations responsible for drastic reductions in litter size, as in the case of reciprocal translocations, which makes it worthwhile to screen the boars of paternal and maternal lines for karyotype abnormalities. This screening is systematically done in France, yielding 0.44% of the animals as affected (Ducos et al., 2007).

The development of statistical tools allows the application of more complex genetic models and helps to discover genetic variation for traits such as behaviour and uniformity with large indirect influence on the breeding goal.

**Marker-assisted Breeding**

**Limitations of phenotype-based breeding programmes**

The genetic improvement programmes described in the previous sections are driven by measuring phenotypes for traits of interest on selection candidates in the nucleus or on close relatives of the selection candidates. These phenotypes are then used to estimate the breeding values of selection candidates for traits and incorporate these in a multi-trait selection index to identify individuals that best meet the breeding objective. Although this has led to impressive increases in performance for several traits, the phenotype-based approach to genetic improvement suffers from several important limitations, including:

- Several traits have low heritability, e.g. reproduction, disease resistance and survival traits. This limits the accuracy and efficiency of selection and, therefore, genetic improvement for these traits.
- Several traits can only be measured later in an animal’s life, e.g. sow lifetime productivity. This either increases generation intervals if the choice is made to wait until the phenotype can be observed on the selection candidates, or reduces accuracy of selection if the choice is made to make selection decisions before the phenotype is observed.
- Several traits cannot be measured directly on selection candidates. For example, many meat quality traits require slaughter of the animal. Thus, for these traits, genetic evaluation is based on phenotypes of relatives that are slaughtered, limiting the accuracy of selection as well as the selection intensity, as individuals that may be potential selection candidates are slaughtered. Another example is disease resistance, which can often not be measured directly on selection candidates because they must be kept under high-health environments that are free of most diseases that affect pigs in the production environment. This category also includes the problem of the genotype-by-environment interactions that exist between the high-health purebred nucleus environment and the crossbred production environment for many performance traits. As explained previously, this can be overcome by collecting phenotype on crossbred progeny in the production environment, but this is expensive and leads to limited accuracies of
selection, higher rates of inbreeding and/or longer generation intervals.

- Some traits are expensive to measure (e.g. disease resistance). Phenotype-based programmes require routine collection of large numbers of phenotypes on the selection candidates themselves or their close relatives, resulting in costs that may not outweigh the benefits.

- The increased emphasis that is placed on relatives in many of the above situations also leads to greater rates of inbreeding within the selection lines. The higher rates of inbreeding result from family members having very similar estimates of breeding values if these EBVs are heavily based on phenotypes of parents, full sibs and half sibs rather than on own performance or progeny performance. Thus, selection tends to be of families rather than individuals within a family, resulting in selected individuals being more closely related and higher rates of inbreeding.

- Whereas the main focus in pig breeding programmes has been on the genetic improvement of additive genetics through selection on EBV, most traits of interest exhibit non-additive effects, e.g. through heterosis when crossing breeds. Traditional quantitative genetics, however, offers limited opportunities to select for non-additive effects.

With the implementation of molecular genetics in the 1980s and 1990s, in particular the discovery of new classes of DNA polymorphisms, prospects to overcome these limitations of phenotype-based selection emerged with the potential to directly select on an individual’s genotype for genes or genetic markers that are associated with the trait. Progress in developing genetic markers and using these to detect regions of the genome associated with traits of interest (quantitative trait loci (QTLs)) has been described in the various chapters dealing with growth, meat quality and reproduction. The focus of this section will be on how such genetic tests can be used for genetic improvement and not on specific markers or genes. The use of genetic markers to enhance genetic progress in the pig has also been discussed by Visscher and Haley (1995). A recent review of the use of markers to map and use genes for complex traits in livestock is in Goddard and Hayes (2009).

**Principles of the use of genetic markers for genetic improvement**

The use of marker-assisted selection (MAS) requires knowledge of genes or markers that are associated with the traits of interest and quantitative estimates of these associations in the population of interest. The principle behind finding such markers is to compare the average trait phenotype of animals that have, e.g. genotype AA at a specific marker to the average phenotype of animals that have genotype AG or GG.

**Three types of genetic markers for within-breed genetic improvement**

For the purposes of the use of genetic markers for selection, three types of genetic tests can be distinguished (Dekkers, 2004):

1. Direct markers: genotype for loci that code for the functional mutation that affects the trait.
2. LD (linkage disequilibrium) markers: genotype for loci that are in population-wide linkage disequilibrium with the functional mutation.
3. LE (linkage equilibrium) markers: genotype for loci that are in population-wide linkage equilibrium with the functional mutation in outbred populations and need to be used on a within-family basis.

Methods for detecting these types of loci were described in Chapters 5 and 7. LE markers can be readily detected on a genome-wide basis by using breed crosses or the analysis of large half-sib families within a breed. Such genome scans require only sparse marker maps (150–300 markers across the genome, depending on marker informativeness and genotyping costs) to detect most QTLs of moderate-to-large effects. Many examples of successful applications of this methodology for detection of QTL regions are available in the literature (see Andersson, 2001). The LD markers are by necessity close to the functional mutation for sufficient population-wide LD between the marker and QTL to exist (typically
within 1 cM, depending on the extent of LD in the population, which depends on population structure and history). LD markers can be identified using candidate gene (Rothschild and Soller, 1997) and fine-mapping approaches (Andersson, 2001; Georges, 2007). Recently, high-density single nuclear polymorphism (SNP) panels have become available, which allow the entire genome to be scanned for genetic markers that are associated with the trait based on population-wide LD with QTLs. Direct markers, i.e. polymorphisms that code for the functional mutations, are most difficult to detect because causality is difficult to prove and, as a result, a limited number of examples are available and most are for genes associated with genetic defects, rather than with a quantitative trait.

Direct markers are preferred for use in genetic improvement because they directly evaluate genotype at the locus that affects the trait. The next preferred type of marker with regard to use in genetic improvement is LD markers, as they allow selection on genotype across the population because of the consistent association between genotype and phenotype. In contrast, use of LE markers must allow for different linkage phases between markers and QTLs from family to family, and effects must be estimated on a within-family basis. Effects of direct and LD markers can be estimated on a random sample of the population, in principle without the need for pedigree. Because of the complications associated with the use of LE markers, they have seen limited application in livestock breeding programmes (Dekkers, 2004), and the emphasis is on developing direct and LD markers.

Methods to incorporate marker data into genetic evaluation procedures were summarized by Dekkers and van der Werf (2007). They depend on the type of marker used (direct and LD versus LE markers), and whether a limited number or many markers are included, as in genomic selection (Meuwissen et al., 2001). These methods result in marker-based EBVs that estimate the breeding value of an individual for QTLs that are associated with markers, and a ‘residual’ or polygenic breeding value that captures the effects of all other QTLs, based on phenotype. A simplified example with genotypes for three markers is presented in Table 16.5.

Once markers associated with the trait have been identified and validated, they can be used for genetic improvement. In the following, we first describe the use of marker information for enhancing the process of integrating superior qualities of different breeds, and then discuss the use of markers to enhance within-breed selection.

### Exploiting between-breed variation

Crossing breeds results in extensive LD, which can be capitalized on using MAS in a number of ways. If a large proportion of breed differences in the trait(s) of interest are due to a small number of genes, gene introgression strategies can be used. If a larger number of genes are involved, MAS within a synthetic line is the preferred method of improvement.

<table>
<thead>
<tr>
<th>Pig</th>
<th>Marker 1 Genotype</th>
<th>Marker 2 Genotype</th>
<th>Marker 3 Genotype</th>
<th>Marker-based EBV</th>
<th>Own phenotype</th>
<th>Index value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>5</td>
<td>+35</td>
<td>+15.0</td>
</tr>
<tr>
<td>2</td>
<td>AA</td>
<td>AA</td>
<td>BB</td>
<td>25</td>
<td>−10</td>
<td>+13.3</td>
</tr>
<tr>
<td>3</td>
<td>AB</td>
<td>BB</td>
<td>AB</td>
<td>−15</td>
<td>−15</td>
<td>−8.3</td>
</tr>
<tr>
<td>4</td>
<td>AB</td>
<td>BB</td>
<td>AA</td>
<td>+15</td>
<td></td>
<td>−5.0</td>
</tr>
<tr>
<td>5</td>
<td>BB</td>
<td>AA</td>
<td>AB</td>
<td>+25</td>
<td></td>
<td>+5.0</td>
</tr>
</tbody>
</table>

*aMarker-based EBVs are based on three markers with allele substitution effects (allele A versus B) of +10, +5 and −10. The markers jointly explain 50% of the genetic variance for a trait with heritability (H²) 0.5. Resulting index weights on molecular score and phenotype are 2/3 and 1/3, respectively, based on Lande and Thompson (1990).*
MARKER-ASSISTED INTROGRESSION. Introgression of the desirable allele at a target gene from a donor breed to a recipient breed is accomplished by multiple backcrosses to the recipient breed, followed by one or more generations of intercrossing. The aim of the backcross generations is to generate individuals that carry one copy of the donor QTL allele, but that are similar to the recipient breed for the rest of the genome. The aim of the intercrossing phase is to fix the donor allele at the QTL. Marker information can enhance the effectiveness of the backcrossing phase of gene-introgression strategies: (i) by identifying carriers of the target gene(s) (foreground selection); and (ii) by enhancing recovery of the recipient genetic background (background selection). Effectiveness of the intercrossing phase can also be enhanced through foreground selection on the target gene(s). If the target gene cannot be directly genotyped, carrier individuals can be identified based on markers that flank the QTL at \(<10\,\text{cM}\), because of the extensive LD that exists in crosses. The markers must have breed-specific alleles, such that line origin can be identified. For the introgression of multiple target genes, gene pyramiding strategies can be used during the backcrossing phase to reduce the number of individuals required (Hospital and Charcosset, 1997; Koudandé et al., 2000). For background selection, markers are used that are spread over the genome at \(<20\,\text{cM}\) intervals, such that most genes that affect the trait will be within 10\,\text{cM} of a marker. Combining foreground and background selection, selection will be for the donor breed segment around the target locus, but for recipient breed segments in the rest of the genome. Foreground selection will result in selection for not only the target locus, but also for donor breed loci that are linked to this locus, some of which could have an unfavourable effect on performance. To reduce this so-called linkage drag around the target locus, greater emphasis can be given to markers that are in the neighbourhood of the target locus during background selection (apart from the flanking markers, which are used in foreground selection).

Most studies have considered marker-assisted introgression of a single QTL (e.g. Hospital and Charcosset, 1997), but often several QTLs must be introgressed simultaneously. Koudandé et al. (2000) showed that large population sizes are needed to obtain sufficient individuals that are heterozygous for all QTLs in the backcrossing phase. This would make marker-assisted introgression not feasible in livestock breeding programmes. In many cases, however, immediate fixation of introgressed QTL alleles may not be required. Instead, the objective of the backcrossing phase can be to enrich the recipient breed with the favourable donor QTL alleles at high enough frequency such that they can be selected on following backcrossing. The effectiveness of such strategies was demonstrated by Piyasatian et al. (2008).

MARKER-ASSISTED IMPROVEMENT OF SYNTHETIC LINES. With introgression, the aim is to recover the recipient breed genotype, except for the donor QTL, because of the superior general performance of the recipient breed. An alternative objective could be to simply aim for individuals with highest merit. Selection would then be for marker genotype as well as EBV, estimated across breeds or lines. This EBV selection would replace background selection, as recovery of the recipient genotype is achieved through selection on genetic merit rather than through selecting for breed of origin. This strategy would be more competitive if the original breeds overlap in merit, and, indeed, background selection based on anonymous markers would be less profitable, as was shown by Dominik et al. (2006).

Strategies for the use of markers to select within a hybrid population were first proposed by Lande and Thompson (1990). The strategy capitalizes on population-wide LD that initially exists in crosses between lines or breeds. Thus, marker–QTL associations identified in the \(F_2\) generation can be selected on for several generations, until the QTLs or markers are fixed or the disequilibrium disappears. Zhang and Smith (1992) evaluated the use of markers in such a situation with selection on BLUP EBV. Piyasatian et al. (2007) showed that the use of genomic selection procedures with marker effects fitted as random effects resulted in improved response and performance. Although all these studies considered the ideal situation of a cross with inbred lines, there will be opportunities to utilize a limited number of markers to
select for favourable QTL regions that are detected in crosses between breeds, thereby enhancing the development of superior synthetics.

**Exploiting within-breed variation**

Most genetic improvement in pigs is based on selection within pure breeds or lines. Within-breed selection can be improved with availability of genotype on markers that are associated with the trait(s) of interest. The benefit of and strategy for the use of marker data for within-breed selection depend on the proportion of genetic variance that is explained by the genetic markers and the accuracy with which the true association of a genetic marker with the trait is estimated. Variance explained by genetic markers depends on the proportion of QTLs that the genetic markers are associated with and on how tight the association is between marker genotype and QTL genotype, as quantified by the extent of LD for LD markers. The accuracy of the estimate of the marker–trait association depends primarily on the number of individuals with phenotype and marker genotype that are available to estimate the marker effects, along with the genetic architecture of the trait (heritability, number and effects of QTLs) and the statistical method used for estimation. Generally, data sets greater than 500 and into the thousands are needed to obtain sufficiently accurate estimates of LD-marker effects. As described previously, estimates of marker effects can be used to develop a marker-based EBV for selection candidates.

Unless the accuracy of the marker-based EBV is high, it usually is advantageous to combine selection on the marker data with selection on phenotype or phenotype-based BLUP EBV, because the latter captures the collective effect of all QTLs, including those not captured by markers. In general, four strategies can be distinguished for the use of the marker data in selection.

1. Selection on marker genotype or marker-based EBV alone.
2. Tandem selection, with selection of candidates on marker genotype or marker-based EBV, followed by selection on phenotype or standard phenotype-based BLUP EBV.
3. Simultaneous selection on a combination of marker data and phenotype or EBV (index selection).
4. Preselection on marker data at a young age, followed by selection on phenotype or EBV at a later age.

Selection on marker information alone (strategy 1) ignores information that is available on QTLs that are not captured by markers, and is expected to result in the lowest response to selection, unless most QTLs that affect the trait are included in the marker-based EBV, as is anticipated for genomic selection (see later). This strategy, however, does not require additional phenotypes, other than those that are needed to estimate marker effects, and can be attractive when phenotype is difficult or expensive to record (e.g. disease traits, meat quality, crossbred performance in the field). Tandem selection (strategy 2) will also not maximize response to selection, because selection on the marker data may eliminate individuals that have superior genotypes for QTLs that are not captured by the marker-based EBV. To prevent this, Smith (1967) and Lande and Thompson (1990) suggested combining the two sources of information in a selection index (strategy 3). Lande and Thompson (1990) showed how selection index methods could be used to optimize the weights to apply to the marker data and phenotype, depending on the proportion of genetic variance explained by the markers and the heritability of the trait, and the increase in accuracy of selection that could be expected. The resulting index provides the best estimate of breeding value based on marker data and phenotype. Extensions of selection index methods to include marker data to multiple-trait selection are in Lande and Thompson (1990) and Weller (2009). Dekkers (2007a) showed how selection on a marker-based EBV could be incorporated into standard selection index software such as SelAction (Rutten et al., 2002) by including the marker-based EBV as a correlated trait with $h^2 = 1$ and correlation equal to the proportion of genetic variance explained by the marker-based EBV. Methods that directly incorporate genetic markers in routine mixed-animal model BLUP genetic evaluation (e.g. Fernando and Grossman, 1989) also result in EBVs that optimally combine marker and phenotypic information.
Although index selection is expected to result in greater response than tandem selection, the choice between tandem and index selection (and other alternatives) also depends on other factors, such as market and cost considerations. For example, rapid fixation of the targeted gene (e.g. by tandem selection) will reduce the costs of genotyping over generations, and may be desirable from a marketing perspective. This can, however, also be achieved by increasing the weight on the molecular score in an index.

Tandem and index selection apply to the use of marker information within a given stage of selection. If selection is over multiple stages, selection on molecular score could be emphasized at an early age when limited or no phenotypic information is available to distinguish selection candidates (strategy 4). A prime example is in dairy cattle, with preselection among full-sib dairy bulls for entry into progeny testing programmes (Kashi et al., 1990; MacKinnon and Georges, 1998). Similar approaches could also be used in pig breeding when deciding which boars to put on, e.g. feed intake recording programmes.

The potential benefits of MAS in pig breeding programmes were evaluated by simulation by Meuwissen and Goddard (1996). Using LE markers, they demonstrated that MAS is mainly useful for traits where phenotypic measurement is less valuable because of: (i) low heritability; (ii) sex-limited expression; (iii) availability only after sexual maturity; and (iv) necessity to sacrifice the animal (e.g. slaughter traits). Benefits were greatest for the latter category of traits. To evaluate the benefit of MAS for meat quality traits, Meuwissen and Goddard (1996) considered two implementation strategies:

1. A random two of four members of each full-sib family are slaughtered to record meat quality data. The remaining individuals are selected on the basis of a marker-assisted EBV for meat quality, once data on their sibs are recorded.

2. Animals are selected on the basis a marker-assisted EBV, and non-selected animals are slaughtered to provide data for the next generation of selection.

Comparisons of these two strategies were made to a conventional selection based on strategy 1 – but without the availability of genetic markers. Strategy 1 gave 24% greater response than conventional selection. The benefit of strategy 2 was substantially greater, but declined over generations as favourable alleles at the QTLs were fixed. The greater response from strategy 2 than 1 was in large part the result of the greater selection intensity that was achieved with strategy 2 because half of the selection candidates were not slaughtered before selection.

Marker data can also be used to address the problem associated with the genetic improvement of crossbreed performance in commercial environments based on pure-line selection in nucleus environments by estimating effects on phenotypes and genotypes collected on crossbred individuals in the field, and using the resulting estimates to estimate breeding values of individuals in the purebred nucleus populations based on their marker genotypes, as illustrated in Fig. 16.4 (Dekkers, 2007b). By analysing data from crossbreds in the field, estimates of marker effects will be for performance of crossbreds under field conditions, rather than for purebred performance in nucleus herds. When using LD markers, estimation of effects can in principle be on an unpedigreed sample of individuals. This has the potential to substantially increase accuracy of selection for commercial crossbred performance, without the increases in generation intervals or rates of inbreeding that are associated with phenotype-based combined crossbred/purebred selection (Dekkers, 2007b).

Commercial application of MAS requires careful consideration of economic aspects and business risks. Economic analysis of MAS requires a comprehensive approach that aims to evaluate the economic feasibility and optimal implementation of MAS. An excellent example of such an analysis is in Hayes and Goddard (2003), who conducted a comprehensive economic analysis of the implementation of LE markers in the nucleus breeding programme of an integrated pig production enterprise. Detection of QTLs and MAS on identified QTL regions for a multi-trait breeding goal and associated genotyping costs and extra returns from the production phase of the integrated enterprise were considered in the economic assessment. Hayes and Goddard
concluded that the implementation of MAS using LE markers was feasible for the assumed cost and price parameters. They also found that, in particular, if QTL detection was based on small sample sizes, stringent thresholds should be set during the QTL detection phase such that genotyping costs during the implementation phase are reduced and selection of false positives is minimized. In a related study, Hayes and Goddard (2004) evaluated the break-even cost of using LD markers developed using the candidate gene approach.

### History of the development of the use of genetic markers in pig breeding

The assistance that can be provided by genetic markers was already illustrated in the 1980s by the use of biochemical markers closely linked to the halothane gene in order to eliminate the susceptibility allele from maternal lines (Mathur and Liu, 2003). The objective was to obtain a halothane-negative slaughter generation, protected from the deleterious effects of the recessive susceptibility gene on liveability and porcine stress syndrome (PSS)-related meat characteristics. This example enters a first category of marker-assisted breeding, whereby gene frequency at a locus of interest or with large deleterious effects can be acted upon more efficiently than through a direct approach based on the phenotypic effects of the gene itself. With the advent of molecular genetics in the 1980s, DNA markers have become the tool of choice to develop indirect tests that can be used to advance selection. Using molecular genetics, a large number of markers and genes associated with traits of interest have been identified in the pig. These are catalogued and summarized in a web-based database called PigQTLdb (Hu et al., 2005). As of March 2010, this database contains reports on 5621 QTLs on 546 different traits from 237 publications. A sizeable number of the identified markers and QTL have shown evidence of non-additive effects, including dominance, epistasis and genomic imprinting. The latter refers to parent-specific expression of genes in the progeny, which is a well-known phenomenon in mammals. Using a genome scan in a breed cross, de Koning et al. (2000) identified many imprinted QTLs in the pig, and several subsequent studies have confirmed these results. Jeon et al. (1999) and Nezer et al. (1999) found paternal expression for muscularity in the insulin-like growth factor 2 gene (IGF2) region of chromosome 2 (SSC2) in pigs. The causative mutation for this effect has now been identified and its imprinted effect confirmed (Jungerius et al., 2004).

Despite this wealth of information on genetic markers and QTLs associated with traits of interest, the application of genetic markers in commercial pig breeding has been limited. Potential reasons include: (i) only a limited number of validated genetic markers have been available, and the cost of genotyping has been relatively high; (ii) the available genetic markers explain only a limited amount of genetic variation for the trait; (iii) marker or QTL effects were estimated on a within-family basis or in experimental crosses, which made it more difficult to incorporate them into breeding programmes; (iv) marker–trait associations did not replicate or were inconsistent across populations; and (v) QTLs were detected in crosses with wild boar or Chinese breeds and the favourable allele was already fixed in commercial breeds.

Table 16.6 provides a list of examples of genetic markers that have been available and/or used in pig breeding. It is interesting to note that most of these markers are LD markers that were identified using the candidate approach. In some cases, the genetic test is for the causative mutation. In addition to publicly available markers and genetic tests, several breeding organizations have had in-house programmes for the development and use of genetic markers for genetic improvement within their breeding programmes (Clutter, 2004; Clutter et al., 2004; McLaren, 2007).

The main use of genetic markers in commercial pig breeding has been to enhance genetic improvement within an established breed or line by using markers to reduce the incidence of deleterious effects (e.g. PSS) or to increase the frequency of alleles with favourable effects. Methods for incorporating genetic markers into breeding decisions are not clear, but probably vary from the preselection of candidates based on genetic markers to incorporating markers into routine genetic
evaluation procedures to enhance the accuracy of EBV (Rothschild, 2008). Other uses of genetic markers to enhance genetic improvement include:

- Introgression of favourable genes from one breed into another breed; an example of marker-assisted introgression is the introgression of the halothane-resistance allele into the Piétrain breed in Belgium (Hanset et al., 1995). By typing the closely linked GPI (glucose phosphate isomerase) locus, Hanset et al. (1995) were able to fix the Large White normal allele and to obtain a halothane-negative Piétrain strain after three backcrosses.
- Synthetic line development.
- Parental identification and verification, which can be used to create, and improves the accuracy of, pedigree information that is used for genetic evaluation in cases where the pedigree of phenotyped individuals is unknown or uncertain.

While most applications focus on additive genetic improvement, the availability of markers or genes with large non-additive effects also opens opportunities for the strategic use of specific genotypes in crossbreeding programmes. A prime example is the use of the paternally expressed IGF2 gene, which has been shown to have favourable effects on lean growth and reproductive performance, but only for the allele that is inherited from the sire (Van Laere et al., 2003; Jungerius et al., 2004; Liu et al., 2006; Mathur et al., 2007) owing to an imprinting effect. By producing sows from a cross between a boar that is homozygous for the wild (G) allele for IGF2, and mating this sow to a terminal sire that is homozygous for the mutant (A) allele for leanness, all market pigs will be lean because their sire allele is the lean allele, in this way also improving uniformity in their carcasses. However, by having inherited the wild allele (G) from their sire, the sows will have increased reproductive performance but will not pass the increased fatness trait on to their progeny (Buys et al., 2006, 2009; Mathur et al., 2007).

<table>
<thead>
<tr>
<th>Gene (action/target)</th>
<th>Traits affected</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYR1 (halothane sensitivity)</td>
<td>Lean growth, porcine stress syndrome, meat quality</td>
<td>Fuji et al., 1991</td>
</tr>
<tr>
<td>RN (Rendement Napole)</td>
<td>Meat quality</td>
<td>Milan et al., 2000</td>
</tr>
<tr>
<td>ESR (estrogen receptor)</td>
<td>Litter size</td>
<td>Rothschild et al., 1996; Chen et al., 2001</td>
</tr>
<tr>
<td>PRLR (prolactin receptor)</td>
<td>Nursing performance and litter size</td>
<td>Vincent et al., 1997; Farmer, 2000</td>
</tr>
<tr>
<td>RBP4 (retinol binding protein 4)</td>
<td>Litter size</td>
<td>Messer et al., 1996</td>
</tr>
<tr>
<td>MC4R (melanocortin-4 receptor)</td>
<td>Lean growth, fatness, feed intake</td>
<td>Kim et al., 2000</td>
</tr>
<tr>
<td>IGF2 (insulin-like growth factor 2)</td>
<td>Lean growth, litter size</td>
<td>Van Laere et al., 2003; Jungerius et al., 2004; Liu et al., 2006</td>
</tr>
<tr>
<td>HFABP/AFABP</td>
<td>Intramuscular fat</td>
<td>Gerbens et al., 1999</td>
</tr>
<tr>
<td>c-KIT receptor</td>
<td>Coat and skin colour</td>
<td>Andersson-Eklund et al., 1996</td>
</tr>
<tr>
<td>MC1R (melanocortin-1 receptor)</td>
<td>Red/black coat colour</td>
<td>Kijas et al., 1998</td>
</tr>
<tr>
<td>PRKAG3 (protein kinase AMP-activated gamma 3-regulatory subunit)</td>
<td>Meat quality</td>
<td>Ciobanu et al., 2001</td>
</tr>
<tr>
<td>HMGAl (high mobility group AT hook 1)</td>
<td>Backfat thickness</td>
<td>Kim et al., 2004</td>
</tr>
<tr>
<td>CCKAR (cholceystokinin type A receptor)</td>
<td>Feed intake and growth</td>
<td>Houston et al., 2006</td>
</tr>
<tr>
<td>CAST (calpain inhibitor)</td>
<td>Tenderness</td>
<td>Ciobanu et al., 2004</td>
</tr>
<tr>
<td>EPOR (erythropoietin)</td>
<td>Litter size</td>
<td>Vallet et al., 2005</td>
</tr>
<tr>
<td>F18</td>
<td>Escherichia coli diarrhoea</td>
<td>Vogeli et al., 1997</td>
</tr>
<tr>
<td>K88</td>
<td>E. coli diarrhoea</td>
<td>Meijerink et al., 1996, 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jørgensen et al., 2003</td>
</tr>
</tbody>
</table>

Table 16.6. Examples of genetic markers and genes used in swine.
In addition to the aforementioned uses of genetic markers to enhance genetic improvement, genetic markers also have several other uses that are relevant to commercial pig breeding programmes. Examples are:

- line verification and identification;
- measuring genetic distances between breeds (Foulley et al., 2006; SanCristobal et al., 2006); and
- traceability, using DNA barcodes to trace pork back to the farm of origin or to the breeding stock (Loftus, 2005; Webb, 2005).

**Genomic selection**

Recent developments in technology have removed some of the limitations of previous applications of QTL-mapping results for MAS, which, as already mentioned, have limited the use of markers in commercial breeding. These developments include genome sequencing, the identification of large numbers of genetic markers across the genome in the form of SNPs, and the cost-effective high-throughput genotyping of tens of thousands of such SNPs on individual animals. Combined with the further development of statistical methods for analysis of molecular data, this has led to a paradigm shift in the strategy of using genetic markers for the prediction of breeding values in the form of what has been termed ‘genomic selection’ (Meuwissen et al., 2001). Genomic selection (GS) is an enhanced version of MAS that involves the selection of animals for breeding on the basis of their genotype for tens of thousands of ‘random’ SNPs that cover the genome. In GS, the association of each SNP with phenotype is estimated using sophisticated statistical and quantitative genetics models without pre-screening markers based on significance. This is in contrast to ‘traditional’ MAS, which involves a two-step approach, with screening of markers based on significance of their association with phenotype as the first step, followed by the use of just those markers for selection (Lande and Thompson, 1990).

The main principle of GS is outlined in Fig. 16.3, which is based on methods outlined by Meuwissen et al. (2001). The first step is to collect phenotypes and DNA from a large group of individuals that have been phenotyped for the trait, or that have progeny with those phenotypes, and to genotype each animal using the SNPChip. The resulting ‘training data’ are used to ‘train’ a statistical model that estimates the effect of each of the SNPs on the SNPChip with the trait phenotype. In principle, the estimate for a given SNP is based on the comparison of the average phenotypes of individuals that have alternative genotypes at

---

**Fig. 16.3.** Steps in genomic selection. SNP, single nuclear polymorphism; BV, breeding variation.
that SNP, as described above, but with GS this is done simultaneously for all markers on the SNPChip. The resulting estimates can then be used to predict the ‘Genomic’ EBV (G-EBV) of new individuals based on their genotypes for the SNPChip.

In general, the prediction model for GS is developed by fitting the following linear model to phenotypes of individuals that make up the training population (Meuwissen et al., 2001):

\[ y_i = \mu + \sum X_{ij} b_j + e_i \]  

where \( y_i \) is the phenotype (or progeny mean phenotype) of individual \( i \) in the training data, \( \mu \) represents fixed effects, the summation \( \Sigma \) is over all genotyped SNPs, \( X_{ij} \) is the number (0, 1 or 2) of copies of allele ‘1’ (versus ‘0’) that individual \( i \) carries at SNP \( j \), \( b_j \) is the allele substitution effect for SNP \( j \), and \( e_i \) is a random residual. One statistical challenge to fitting this model is that the number of markers (>30,000 is common) typically is much greater than the number of animals with phenotypic records that are available to estimate their effects (typically less than 2000). The most common method to deal with this is to fit the effect of each SNP as a random effect using Bayesian methods (Meuwissen et al., 2001; Gianola et al., 2009). An important distinction exists between models that use prior distributions that assume genetic variance is equally distributed across all genotyped SNPs, the so-called genomic or G-BLUP method of Meuwissen et al. (2001), and methods that assume that a large proportion of SNPs have zero or very small effects, such as the Bayes-B method of Meuwissen et al. (2001), and other versions and implementations of these methods (Gianola et al., 2009; Calus, 2010).

Because the accuracy of the ‘genomic’ EBV (G-EBV) values that are derived using GS models are difficult to predict, a validation step is usually included (Fig. 16.3). This involves separating the data set into training and validation data sets, developing the GS prediction model on the training data set, applying the prediction model to compute G-EBV for individuals in the validation data set and correlating the resulting G-EBV with the phenotypes of these individuals. Typically, the training and validation data sets are separated by year of birth of individuals, with the older animals used for training and the younger animals for validation (Habier et al., 2007).

In principle, GS represents an extension of MAS, as described above, but with some important differences. The first is that MAS is based on the premise of using a limited number of markers that have been shown to have large effects on the trait(s) of interest in prior QTL-detection or marker-association analyses; in contrast, GS uses all SNPs for analysis and prediction, without the screening of associations for significance. The second is that genomic selection fits all SNPs in the model simultaneously, rather than one at a time or by genomic region, as is done in typical QTL-detection studies. Combined with fitting marker effects as random effects, which regresses estimates towards zero, depending on the amount of information available in the data to estimate the effect and the strength of prior information, marker-based EBV values obtained from GS methods have the potential to capture a much greater proportion of genetic variance than was possible with the limited number of markers included in previous methods for MAS. In addition, because estimates are based on LD between markers and QTLs that exist across the population, G-EBV can in principle be derived using estimates of SNP effects obtained from phenotypes on individuals that are not closely related to the selection candidates. Thus, with GS, in contrast to traditional selection, the phenotype that is collected on an individual can be used to estimate not only the breeding value of the animal itself or its relatives, but also that of unrelated individuals. Thus, the need to collect phenotypes on selection candidates or their close relatives is removed, and phenotypes collected can be leveraged for breeding value estimation across different families and generations, and even to different populations and breeds. The use of GS for training across breeds does require denser panels than those that may be needed within a breed (e.g. over 300,000 versus 50,000 SNPs, depending on the distance between the breeds involved), because only markers that are very close to the QTLs are expected to have associations that are consistent across breeds. In addition, this approach would fail if the effects of QTLs are not consistent across breeds because of non-additive
Recent applications in real data, in particular from dairy cattle (Hayes et al., 2009; Loberg and Dürr, 2009), and further theoretical work by Goddard (2009) and Meuwissen (2009) have shown that much larger training data sets are needed than indicated by the initial simulation results (several thousands of genotyped and phenotyped individuals). In addition, although the theory of GS, as proposed by Meuwissen et al. (2001), is predicated on capturing associations between markers and QTLs because of historic LD associations between markers and QTLs, it is now becoming clear that other factors can make major contributions to GS predictions from training data with dense and complex pedigree and family structures. These include genetic relationships (Habier et al., 2007) and recent LD and within-family effects (Goddard, 2008). Although these effects contribute to the accuracy of G-EBV for close relatives of individuals used for training (e.g. their progeny), they limit the accuracy of G-EBV for subsequent generations, for families that are not well represented in the training data and for other less related populations. The implication is that G-EBV will be most accurate for selection candidates that are closely related to individuals that are in the training data, and that accuracies may be much lower for individuals that are less related, in particular for individuals from other breeds. Thus, continuous phenotypic recording across families may be needed to maintain accuracies of G-EBV within a breed.

To fully capitalize on the benefits of genomic selection, pig breeding programmes may need to be changed substantially because genomic selection removes many of the limitations that exist in current phenotype-based programmes with regard to when and on which individuals phenotypes must be evaluated (see Fig. 16.4). Comprehensive strategies for selective recording of phenotypes with genomic selection must be evaluated. Redesign of breeding programmes with genomic selection can also capitalize on the greater emphasis that genomic selection provides for selection within families. This can be utilized in a number of ways to: (i) reduce rates of inbreeding; (ii) increase response to selection for a given rate of inbreeding by increasing selection intensities; or (iii) reduce the size of breeding populations and, thereby, costs, while maintaining the same rates of response and inbreeding.

**Conclusions**

The efficiency of breeding programmes can be assessed by measuring genetic changes occurring over time. The situation in pigs is well documented for the most important traits, based on the central-station testing and on-farm testing records available. From their review on genetic trends estimated in large national purebred populations, Sellier and Rothschild (1991) conclude that appreciable genetic gains have been obtained for growth and body composition traits, of the order of 0.5–1.5% of the mean annually over periods of time ranging from 5 to 10 years in most studies. Annual genetic gains of about 0.01–0.02 piglets born have been reported by Bidanel and Ducos (1994), over the period 1975–1991. Since 1995, breeding organizations have grown more and more commercial, as well as more and more international, and official publication of genetic trends has unfortunately disappeared. Phenotypic trends are clearly positive though, as can be observed from national production data. A phenotypic trend of around 0.5 piglets per sow per year appears to be feasible, with the larger part most probably being genetic.
It can be concluded that efficient breeding schemes can help the pork industry to continuously adapt to a changing world. Directions of these changes are adaptation to different environments, adaptation to a changing economy (reduction in labour), adaptation to changing society concerns (castration, tail docking, group housing), and the desire for stronger and more robust animals. These changes require appropriate data recording, preferably automated, and associated pedigrees, preferably on a crossbred level.

The most important driving factor for the pork industry is still cost price reduction: more piglets per sow per year with low feed costs and low labour input under a large range of environments.

The paradigm shift, which is currently under way, relates to the possibilities of genomics. The balance between phenotyping and genotyping is a delicate one. The aforementioned changes require that new phenotyping and appropriate phenotyping are prerequisites for: (i) QTL detection; and (ii) training for genomic selection. Costs involved in maintaining enough populations of adequate size, phenotyping, genotyping and genetic evaluation infrastructure are high and require more and more cooperation or consolidation.

Genetic variation in the pig, however, is very large, and populations can adapt quickly to changing demands of the farmers in terms of efficiency, of the consumer in terms of safety and of society in terms of adaptation to the environments offered. The new era of genomics offers tremendous opportunities for this adaptation.

Acknowledgement

Large sections of this chapter were based on the chapter by Louis Ollivier in the first edition of *The Genetics of the Pig*.

References


Introduction: Creating the Building Blocks – Genomics, Transgenesis and Cloning

Obtaining a complete draft of the pig genome sequence has been central to the development and broad acceptance of the pig as a biomedical model (Schook et al., 2005a,b). The pig genome sequence has recently been completed (http://www.ensembl.org/Sus_scrofa/Info/Index), and the key building blocks for full utilization of the pig as a biomedical model are now in place: completed genome sequence, ability to produce transgenic animals and the ability to replicate the model through somatic cell cloning (Schook et al., 2005b). The emergence of genetic information and the development of the necessary tools to target manipulations, in combination with the ability to clone pigs, provide a new and highly relevant animal model. These building blocks have stimulated the development of ‘genomic postulates’ (Table 17.1) for evaluating animal models and, relevant to this chapter, the significance of the pig. This chapter was developed to provide background on the need for relevant animal models and to address each of the aspects of the genomic postulates. Owing to the overwhelming physiological (Tumbleson and Schook, 1996) and genomic similarities between pigs and humans (Humphray et al., 2007), the pig provides a uniquely relevant animal model for human disease. In addition, a recent CRISP (Computer Retrieval of Information on Scientific Projects) search (1999–2003) indicated that the US National Institutes of Health (NIH, which has over 20 institutes and centres) sponsored research that supported 2400 separate grants that utilized the pig. Thus, a broad foundation supporting the pig as a model in biomedical research already exists from which to build future programmes. There is also growing...
interest within the biomedical community with respect to the utilization of pigs in bioengineering, imaging and behavioural studies.

The Animal Model Concept

The use of animals to study human physiology and anatomy can be traced back to the second century common era (CME) in which Galen, a Greek physician and philosopher, completed research studies on apes and pigs (Galen, 1586) (Fig. 17.1). Galen incorrectly assumed that all extracted information derived from his use of animals could be directly applied to humans. It was not, however, until the 16th century CME that his error was initially recognized (Nomura et al., 1987), when Bernard proposed the use of chemical and physical induction of disease in animals, thus becoming the first advocate for creating ‘induced animal models’ for biomedical research. At the turn of the 20th century came the development of infectious disease animal models and their use for evaluating antibacterial drugs, and the introduction of the ‘germ theory of disease’ (Koch, 1884; Fanning, 1908). The end of the 20th century and the beginning of the 21st century realized the ability to utilize naturally occurring models resulting from spontaneous mutations – severe combined immunodeficiency (SCID) or nude mice – and from genetically modified animal genomes through transgenesis or site-directed homologous recombination. Linkage with the ability to clone animals, through the utilization of either embryonic stem cells or somatic cell nuclear transfer, provided even further ability to use animals that have phenotypic characteristics close to humans as relevant animal models for dissecting human disease. Finally, the emergence of the whole genome sequencing of animals with many physiological similarities to the human, such as the pig, supports the ability to actually create a large animal model that is

Table 17.1. Genomic postulates, adapted from Koch’s postulates.

1. Isolate and propagate causal gene from animal
2. Characterize (manipulate) gene in vitro
3. Reintroduce putative gene (create transgenic animal) to test causality
4. Demonstration of causal relationship through induced phenotype

Fig. 17.1. Timeline of animal models. SCID, severe combined immunodeficiency. Sources: aGalen, 1586; bKoch, 1884; cFanning, 1908; dDunn, 1965; eMahley et al., 1975; fPantelouris, 1968; gHardy et al., 1981; hBrinster et al., 1984; iWaters et al., 1998; jCooper et al., 2002; kLaske et al., 2005; lAdam et al., 2007.
genetically and phenotypically similar to humans in terms of disease attributes.

Animal models represent important tools for investigating the pathogenesis of human disease and developing appropriate treatment strategies. The coupling of genomic information (genome sequence, gene expression profiling and proteomics) with enabling technologies (transgenesis and cloning) has revolutionized the development of human biomedical animal models. Traditionally, the mouse has been a powerful experimental system for understanding the complexity of cancer, diabetes and cardiovascular disease, among others. The dog is also considered a comparable model to human disease because of its similarities to human anatomy and physiology, particularly with respect to the cardiovascular, urogenital, nervous and musculoskeletal systems. As such, it has long been used as a model in drug discovery and development research. Human disease may best be recapitulated in a large mammal such as the pig. The pig is often the primary biomedical model for a number of diseases, for surgical research and for organ transplantation owing to the similarity in size, anatomy and physiology between pigs and humans (Swanson et al., 2004). Animal models, regardless of species, can be grouped into one of the following five categories:

(i) spontaneous models; (ii) genetically modified models; (iii) induced or experimental models; (iv) negative models; and (v) orphan models (Table 17.2).

One approach to studying human disease is to characterize a naturally occurring disease in an animal that corresponds to a human disease. The best-known spontaneous model is the athymic nude mouse, the use of which represented a turning point in the study of heterotransplanted tumours and enabled the first description of natural killer cells (Pantelouris, 1968). Genetically engineered models were created that harboured genetic changes commonly found in human disease. The first transgenic mouse tumour model was established by overexpression of viral and cellular oncogenes in specific tissues (Brinster et al., 1984; Stewart et al., 1984; Adams et al., 1985; Hanahan, 1989). Induced models involve healthy animals in which the condition to be studied is experimentally induced through surgical modifications, genetic modifications or chemical application – demonstrated in 1918 when Yamagiwa and Ichikawa showed that coal tar experimentally applied to rabbit ears caused skin carcinomas (Yamagiwa and Ichikawa, 1918). More recently, considerable insight has been gained into the strengths and weaknesses of toxicity and Table 17.2. Advantages and disadvantages of animal model types.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spontaneous</td>
<td>Similar disease phenotype to humans</td>
<td>Long latency</td>
<td>Nude/severe combined immunodeficiency (SCID) mice (Pantelouris, 1968)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Canine haemophilia (Giles et al., 1982); canine prostate cancer (Waters et al., 1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Porcine tumour model (Adam et al., 2007)</td>
</tr>
<tr>
<td>Genetically modified</td>
<td>Defined genetic background</td>
<td>Not genetically defined</td>
<td>Mouse tumour model (Brinster et al., 1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phenotypic expression of genes can differ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transgenesis and homologous recombination</td>
<td></td>
</tr>
<tr>
<td>Induced or experimental</td>
<td>Gene expression controlled through diet or inducers</td>
<td>Not predictive of therapeutic success</td>
<td>Atherosclerosis (Mahley et al., 1975; Bell and Gerrity, 1992; Dixon et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Rapid disease onset Free choice of species</td>
<td></td>
<td>Obesity (Spurlock and Gabler, 2008)</td>
</tr>
<tr>
<td>Orphan</td>
<td>Useful for evaluation of chemical/radiological treatments</td>
<td>Do not faithfully mimic human disease</td>
<td>Diabetes (Mordes and Rossini, 1981; Larsen et al., 2002; Larsen and Rolin, 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feline leukaemia (Hardy et al., 1981); bovine leucosis (Gillet et al., 2007)</td>
</tr>
</tbody>
</table>
carcinogenicity studies in laboratory rats and mice. Infectious disease models are often restricted to a limited number of susceptible species, and the remaining unresponsive species are considered negative models because they do not develop the disease when exposed to a particular stimulus (Hau, 2008). The main application of negative models is to gain insight into the physiological basis of disease resistance. There are functional disorders present in non-human species that have not yet been described in humans. Often, a similar disease will be identified in a human that was previously described in animals. These animals represent ‘orphan models’ for that particular disease as no human equivalent has been identified. Feline leukaemia (FeLV) represents a naturally occurring disease in domestic cats that is not transmissible to humans; like lymphoma in humans, lymphoma induced by FeLV in cats is characterized by immunosuppression.

The incidence of chronic disease due to complex genetic and environmental interactions, however, has continued to increase during the past century. Understanding human disease is difficult owing to the complexity of genetics and lifestyle interactions, and the high cost associated with developing therapeutics. As such, appropriate biomedical models are essential because most medical knowledge, treatment regimes and medical device developments are based on robust animal models. As genomic and bioinformatic technologies continue to advance, our knowledge of animal models will increase, thereby refining our choice of models and enabling the development of more applicable models. Animal models are essential tools for studying gene–gene interactions and gene–environment effects, and for preclinical testing of therapeutic interventions. Given that mice, the most common animal model, frequently do not faithfully recapitulate human disease, pigs will continue to serve as important biomedical models.

**Informing Human Physiology: Similarities between Pig and Human Phenotypes**

Animal physiology has significantly contributed to the basic understanding of human development and physiology related to disease (Table 17.3). For example, classical endocrinology studies in pigs have led to the current understanding of several reproductive and pituitary hormones, most notably the composition of insulin, which was first determined for porcine insulin and was used for several decades to treat human diabetes (Rohrer et al., 2003). The porcine biomedical model has provided a fundamental research platform for developing human reproductive techniques and for studying reproductive diseases. Ongoing research using the pig to study cancer and diabetes is contributing greatly to our understanding of these diseases and is further expanded upon in this chapter (Table 17.3). The pig has many similarities in structure and

**Utilizing the Pig to Improve Human Health**

During its multiple domestication events, the pig has undergone intense selection pressures for various phenotypes throughout the world (Chen et al., 2007). First domesticated in Asia from the wild boar, germplasm was quickly moved around the world by explorers and used for food and products. Intense selection and breeding have provided distinct phenotypes differing in metabolism, fecundity, disease resistance and meat products (Schook et al., 2005b; Schook, 2007). Such selective pressures have resulted in differentiated subpopulations and phenotypes extremely relevant to current and future human health research. The selection of ‘mini’ and ‘micro’ pigs for size, independently by investigators throughout the world, attests to the global relevance of this experimental animal in biomedical research. The porcine model is also relevant to human health research priorities such as obesity, female health, cardiovascular disease, nutritional studies (as the pig is an omnivore), and communicable diseases (reviewed in Tumbleson and Schook, 1996). The pig provides a valuable biological model in these priority areas because of the vast amount of research that has been conducted on the genetic and environmental interactions associated with complex, polygenic physiological traits.
Table 17.3. Validated swine biomedical models.

<table>
<thead>
<tr>
<th>Type of investigation</th>
<th>Model</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart physiology</td>
<td>Stent design, tissue engineering of blood vessels</td>
<td>Bedoya et al., 2006; Gyöngyösi et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Atherosclerosis</td>
<td>Turk and Laughlin, 2004; Turk et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Myocardial infarction</td>
<td>Ambrose, 2006; Boluyt et al., 2007</td>
</tr>
<tr>
<td></td>
<td>Ex vivo heart model</td>
<td>Laske et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Emergency procedures</td>
<td>Casas et al., 2005; Geddes et al., 2006</td>
</tr>
<tr>
<td>Reproductive function</td>
<td>Maternal–fetal interactions</td>
<td>Green et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Embryo development</td>
<td>Sun and Nagai, 2003; Rohrer et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Sperm</td>
<td>Strzezek et al., 2005; Lavitrano et al., 2006</td>
</tr>
<tr>
<td>Transplantation</td>
<td>Cell and organ transplants</td>
<td>Larsen and Rolin, 2004; Street et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Percutaneous permeation</td>
<td>Cooper et al., 2002; Ibrahim et al., 2006</td>
</tr>
<tr>
<td>Skin physiology</td>
<td>Contact dermatitis</td>
<td>Simon and Maibach, 2000; Dalton et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Skin culture model</td>
<td>Stuetz et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Melanoma</td>
<td>Huang et al., 2006</td>
</tr>
<tr>
<td>Brain</td>
<td>Stroke</td>
<td>Imai et al., 2006</td>
</tr>
<tr>
<td></td>
<td>AIDS, dementia</td>
<td>Tambuyzer and Nouwen, 2005</td>
</tr>
<tr>
<td></td>
<td>Drug-binding sites and interactions</td>
<td>Minuzzi et al., 2005</td>
</tr>
<tr>
<td>Gut physiology and nutrition</td>
<td>Gut structure and intestinal metabolism</td>
<td>Eubanks et al., 2006; Qiu et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Obesity</td>
<td>Brambilla and Cantafora, 2004</td>
</tr>
<tr>
<td></td>
<td>Probiotics and gut physiology</td>
<td>Reid et al., 2003; Domenechini et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Food allergies</td>
<td>Bailey et al., 2005; McClain and Bannon, 2006</td>
</tr>
<tr>
<td>Biochemical</td>
<td>Response to injury</td>
<td>Schmitt and Snedeker, 2006</td>
</tr>
<tr>
<td></td>
<td>Imaging techniques</td>
<td>Ellner et al., 2004; Goldberg et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Osteoporosis, bone density analysis</td>
<td>Teo et al., 2006</td>
</tr>
<tr>
<td>Tissue engineering</td>
<td>Cartilage repair</td>
<td>Chang et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Spinal fusion</td>
<td>Drespe et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Organ-specific gene delivery</td>
<td>Kawashita et al., 2005</td>
</tr>
<tr>
<td></td>
<td>Cataract repair</td>
<td>Lassota et al., 2006; van Kooten et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Polymer scaffolds</td>
<td>Brown et al., 2006; Moroni et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Tooth development</td>
<td>Hu et al., 2005</td>
</tr>
<tr>
<td>Respiratory function</td>
<td>Neonatal respiratory distress</td>
<td>Miller et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Asthma</td>
<td>Turner et al., 2002; Watremez et al., 2003</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>Therapeutics (vaccines, biotherapeutics, drug therapies)</td>
<td>González et al., 2004; Cheetham et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Developmental interactions</td>
<td>Hasslung et al., 2005; Butler et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Mucosal tissue responses</td>
<td>Dawson et al., 2005; Elahi et al., 2005; Pomeranz et al., 2005; Dvorak et al., 2006</td>
</tr>
<tr>
<td></td>
<td>Host response</td>
<td>Houdebine, 2005</td>
</tr>
</tbody>
</table>

function to humans, including size, feeding patterns, digestive physiology, dietary habits, kidney structure and function, pulmonary vascular bed structure, propensity to obesity, respiratory rates and social behaviours (Tumbleson and Schook, 1996). Because the pig is an omnivore, it provides an adaptable model to evaluate chronic and acute exposures to xenobiotics such as alcohol, tobacco, feed additives and environmental pollutants (Schook, 2007). Pigs have been used as models to evaluate alcoholism, total parenteral nutrition, organ
transplantation, atherosclerosis, exercise, hypertension, melanoma, nephropathy, dermal healing, shock and degenerative retinal diseases.

A severe shortage of organs and tissues for transplantation has also stimulated increased consideration of pigs as a potential solution, particularly with the recent ability to genetically modify pigs to overcome acute rejection (Lai et al., 2002). Targets for the genetic modification of pigs for xenotransplantation have thus far emphasized reducing the immunogenicity of porcine cells and tissues and preventing rejection after transplantation of porcine tissue. Acute rejection is mediated through preformed antibodies against galactosyl-α-1,3-galactose epitopes expressed on the surface of pig cells. Transgenic pigs have been developed that express regulators of the complement cascade, including CD55, CD59 and CD46, which suppress the attack on donor tissues (Bucher et al., 2005; Cox and Zhong, 2005; Houdebine, 2005). Another approach has focused on eliminating the galactosyl-α-1,3-galactose antigen from the surface of donor cells. Researchers have generated pigs without the gene encoding α-1,3-galactosyltransferase (Zhong, 2007). This was accomplished by the serial knockout of the gene in cultured pig fibroblasts, followed by somatic cell nuclear transfer to generate pigs. The convergence of transgenic and cloning techniques has enabled multilayered genetic modifications to be made in a single animal.

Breeding among multiple existing transgenic lines and introducing new genes by somatic cell nuclear transfer can be used in combination to overcome the various stages of xenograft rejection associated with xenotransplantation (Matsunari and Nagashima, 2009). The necessary genetic modifications are dependent on the specific transplant procedure. For example, the removal of the αGal epitope to prevent antibody reactivity and the insertion of complement regulators would increase the success of vascularized grafts, while pancreatic islet grafts would require the insertion of complement regulators, anticoagulants to prevent an inflammatory reaction and an anti-apoptotic gene to counteract ischaemia and reperfusion injuries (d’Apice and Cowan, 2009). Using these approaches, polytransgenic and α-1,3GalT-KO pigs have been produced, but further research is needed to create an efficient model (Rood et al., 2005; Tseng et al., 2005; Yamada et al., 2005; Cooper et al., 2007).

Phenotypic research utilizing unique pig breeds has identified genetically controlled differences in fat deposition (Rothschild and Ruvinson, 1998; Malek et al., 2001a,b). Such information provides the basis for developing an experimental model for understanding obesity and for the development of nutritional interventions from prenatal nutrition to aged cohorts. Porcine resource populations have been selected for phenotypic variation in bone density (osteoporosis), sex-expressed nutritional and reproductive characteristics, and growth and development (embryonic, prenatal and postnatal). Using comparative genomics, new models have been identified to study how metabolism is linked to obesity-induced diabetes (Milan et al., 2000). The porcine model will also be invaluable to study host–pathogen interactions for food safety (e.g. Salmonella), potential biological warfare agents (African swine fever; foot-and-mouth disease) and agents that affect food security and human health (e.g. porcine endogenous retroviruses and other zoonotic diseases).

Linking Genotypes and Phenotypes Relevant to Human Health

The discovery that mammalian genomes probably contain only 20,000–30,000 genes suggests that alternative transcripts and post-translational modifications must play a greater role in phenotypic expression than previously appreciated. It is also expected that single gene products affect different traits or disease states depending on the temporal and spatial presence of gene products. As an omnivore, the pig is prone to many of the same dietary health problems as humans. Depending on diet and genetics, pigs can suffer from hypertension, hypercholesterolaemia, dyslipidaemia, insulin resistance and atherosclerosis. The pig has mutations in similar genes affecting these metabolic disorders (e.g. ApoB and LDLR for hypercholesterolemia) (Ajiello et al., 1994; Hasler-Rapacz et al., 1998). Piglets are the preferred model organism to develop human
infant formula as their nutritional needs are comparable to those of human infants. Because of their similar digestive tracts, pigs are also susceptible to comparable enteric food-borne pathogens (e.g. *Salmonella*, enterohaemorrhagic *Escherichia coli*) and pig intestinal linings are used for *in vitro* studies of interactions with the intestine and these pathogens. Pigs are also susceptible to gastric ulcers that apparently are induced by diet and stress (Engstrand *et al*., 1990). Additional anatomical similarities with humans include renal morphology, eye structure, skin and tooth development. The pig is also one of few animals that will voluntarily eat to obesity, as well as being susceptible to alcoholism.

There are two reasons for research to investigate obesity-related genes in the pig. First, as already mentioned, the pig is a more realistic model organism for human obesity owing to its physiological similarities to humans (Tumbleson and Schook, 1996). As the pig is a true omnivore, the molecular basis and digestive tract anatomy of the pig are much closer to those of humans than any laboratory animal species, as identified by significant DNA polymorphisms of obesity-related genes in the pig genome that might provide useful targets for the genetic study of human obesity. The second reason is that the genetic components of human obesity can play important roles in pig performance traits such as fatness, growth rate and feed intake.

**Surrogate Systems for Human Experimentation**

The domesticated pig has provided numerous surrogate experimental models for biomedical research. There has been a long tradition of using abattoir tissues for the purification of enzymes and the elucidation of metabolic pathways. These tissues have also served as initial biologicals, with bovine and porcine insulin providing pre-recombinant DNA therapeutics and purified enzymes used to determine crystalline structure. Porcine gamete biology has played a critical role in our understanding of stem cells and *in vitro* fertilization (Wu *et al*., 2001; Yin *et al*., 2002). Because of the wealth of biological information derived from the porcine system, it has increasingly become important for studying epigenetic effects, as well as unravelling genomic imprinting. The demonstration that pigs can be cloned using *in vitro* cloning systems provides an invaluable technology platform for developing relevant clones of genetic models for biomedical research (Betthauser *et al*., 2000; see Chapter 11). In addition, a major obstacle for producing cloned genetically modified pigs has been overcome (Lai *et al*., 2002). Investigators have created a nuclear transfer technology using clonal fetal fibroblasts as nuclear donors for the production of gene-specific knockouts. This technology platform has significant applications beyond xenotransplantation, and the availability of genomic sequences will facilitate the broader utility of the pig as a surrogate system for human experimentation.

The phenotypic diversity of hundreds of porcine breeds distributed throughout the world provides a tremendous resource for ‘comparative phenomics’, the application of comparative genomic principles to the discovery of new genes underlying diverse phenotypes. In only a few thousand years, selective breeding has produced pig breeds that thrive in diverse environments (e.g. high altitude versus tropical), convert energy to muscle mass efficiently and rapidly, and tolerate specific pathogens. There can be little doubt that the understanding of what makes porcine breeds different with respect to reproductive efficiency, bone structure, growth rates, fat deposition, altitude or heat tolerance and resistance to specific pathogens will be important to understanding basic biological processes important to human health (see Chapter 18).

**Extrapolation from Animals to Humans**

The selection of an animal model depends on a number of factors relating to the hypothesis to be tested. Often a number of different models may advantageously be used to study a biological phenomenon associated with a human disease. For diseases such as cancer, there is a wide range of well-described models available,
both induced and spontaneous, in a variety of species. The key factor in using animal models for studying disease is that the results can be extrapolated to humans. Animal models of human disease are deemed relevant only if they are useful in recapitulating disease pathogenesis and assisting in the development of approaches to intervention or therapy (Hau, 2008). Thus, to ensure full utilization, a model needs to reliably mimic the normal anatomy and physiology of human organs and tissues of interest, as well as accurately reflect the morphological and biochemical aspects of disease pathogenesis.

The rationale behind extrapolating results from an animal model to humans is primarily based on the similarity between morphological structures and physiological processes. For example, an animal model of cancer should ideally undergo tumour development and progression in a similar fashion to humans. While many animals are more or less similar to humans in regard to biological characteristics, there are prominent differences in body size between species, which affects their appropriateness as a model for certain experiments. The validity of extrapolation may be further complicated by the prevalence of disease in humans, with certain sectors of the population having a higher incidence of one type of disease over another owing to genetic and environmental influences.

Traditionally, animal models were used to identify the genes responsible for a disease. Trends in the use of animal models are changing as new technologies are enabling researchers to use animal models to study the effects of changes in genetic pathways. Developments in the fields of genomics, proteomics, biotechnology and bioinformatics are changing the nature of biomedical research. The Human Genome Project is providing genetic information, not only from humans, but also from animals traditionally used as models. Increased insight into genetic pathways and gene–environment interactions that are involved in the aetiology of complex human disease is providing the knowledge required to select better animal models. This knowledge can be applied to produce specific transgenic animals or knockouts, which better mimic the physiological complexity of human disease than existing models. New, more precise models for the development of therapeutics can be created. Animal models are essential tools for studying gene–gene interactions and gene–environment effects, and for preclinical testing of therapeutic interventions.

An important theme in toxicology research is the search for and the assessment of animal models that are predictive for adverse effects of pharmaceuticals in humans. This process is based on the assumption that the current choice of animal models is truly predictive of a human response to a treatment. To validate this assumption, a large multinational pharmaceutical company survey analysed data compiled from 150 compounds to determine the concordance of the toxicity of pharmaceuticals observed in humans with that observed in experimental animal models (Olson et al., 2000). The concordance rate was found to be 71% for comparable target organs in rodent and non-rodent species, with non-rodents alone being predictive for 63% (primarily the dog) of human toxicity and rodents alone for 43% (primarily the rat). The highest incidence of overall concordance was seen in haematological, gastrointestinal and cardiovascular human toxicities, and the least was seen in cutaneous human toxicity. The results of this survey support the value of in vivo toxicology studies to predict for human toxicity associated with pharmaceuticals, and indicate that data collected from experiments in animals can be extrapolated to humans. It can also be concluded that the type of animal model chosen must be carefully evaluated. Traditionally, toxicology studies utilize rat and dog models, without considering whether there is an alternative species that might be more appropriate for testing a specific compound. While no animal model can completely recapitulate the effects of every drug administered to humans, previous research has shown that large animals are better preclinical models for drug toxicity than rodents (Olson et al., 2000).

**Modelling Human Disease in the Pig**

The pig has been used as an important large animal model for human disease for decades. The animal has a long lifespan of 10–15 years (Hau and Van Hoosier, 2003), so disease
progression is more similar to that seen in humans. Furthermore, as already discussed, the pig shares anatomical and physiological characteristics with humans that make it a unique and viable model for biomedical research (Tumbleson and Schook, 1996). Because of the similarity in body mass of pigs to humans, the pig has become a model of choice for tissue engineering and imaging studies (Lunney, 2007). Their large size also makes them ideal models for study in such medical fields as surgery, imaging, chemotherapy and radiation, which cannot be accurately tested in small animal models.

Their cardiovascular anatomy and physiology, in combination with the pig’s response to atherogenic diets, have made them a universally standard model for the study of atherosclerosis, myocardial infarction and general cardiovascular studies. Their gastrointestinal anatomy has some significant differences from that of humans; however, the physiology of their digestive processes has made them a valuable model for digestive diseases. The urinary system of swine is similar to humans in many ways, especially in the anatomy and function of the kidneys (Swindle and Smith, 2000). Swine are also a standard model for skin and reconstructive surgical procedures, and have been developed as models of transdermal toxicity. The anatomy and physiology of organs such as the liver, pancreas, kidney and heart have also made this species the primary species of interest as organ donors for xenograft procedures (Swindle and Smith, 2000).

In addition, the ability to use pigs from the same litter, and cloned or transgenic pigs, facilitates genetic mapping (Lunney, 2007) and minimizes immunological differences between animals in transplant studies. The availability of numerous well-defined cell lines from a broad range of tissues will assist in studies of gene expression and drug susceptibility testing. Sequencing of the swine genome (Schook et al., 2005a) has provided increasingly advanced genetic and proteomic tools for pigs. Many of these studies employ genomic approaches, as in heart, transplantation and melanoma models. The pig genome has a high sequence homology to humans, 60%, compared with a 40% sequence homology of rodents to humans (Thomas et al., 2003; Humphray et al., 2007), and the pig chromosomal structure has a higher similarity to humans than those of the mouse, rat, dog, cat, horse or cattle (Meyers et al., 2005; Murphy et al., 2005). Each model will be affected by the availability of the functional genomic tools and swine genome sequence and maps (Rothschild et al., 2007; Tuggle et al., 2007).

Creating a Porcine Cancer Model

The pig is an attractive model to study cancer biology and to help close the gap between basic science and patient benefit. Compared with rodents, the pig metabolizes drugs and undergoes tumorigenesis in a manner analogous to humans. Like humans, the incidence of cancer in pigs is rare, with a prevalence of childhood cancer – Wilm’s tumours in young pigs (Anderson and Jarrett, 1968) – and a broader spectrum of cancers in adults (Brown and Johnson, 1970). Furthermore, the pig provides an ideal system for preclinical studies of imaging, as well as of hyperthermia, radiation or photodynamic therapy of tumours. It is almost impossible to do intensity-modulated radiation therapy on mice owing to the small tumour size and the energy of the clinical accelerator. High-resolution intensity treatment in other rodents is hindered by the same problems, and devices used for hyperthermia treatment of tumours cannot be scaled down to be useful for studies in rodents.

Parallels in cancer biology between pigs and humans extend to the molecular level, as demonstrated by the reduced number of genes required to convert human and pig cells to a tumorigenic state compared with mouse cells (Kendall et al., 2005). Additionally, telomerase is suppressed in a number of tissues and reactivated during cancer in both humans and pigs (Pathak et al., 2000; Stewart and Weinberg, 2000), indicating that there are also similarities in the process of tumorigenesis between the species. The genomic sequence homology between pigs and humans is also very high (Swanson et al., 2004), and the porcine pregnane X receptor protein that regulates p450 cytochrome CYP3A, which metabolizes almost half of prescription drugs in humans, is more similar to that of humans than, for example, mice (Xie and Evans, 2002; Pollock et al., 2007).
It has been demonstrated that the enforced expression of transgenes that mimic genetic changes occurring in many types of human cancers can drive normal primary porcine cells to a tumorigenic state. Specifically, co-expression of human TERT (hTERT), p53<sup>DD</sup> (a dominant-negative truncation mutant of p53), cyclin D1, CDK4<sup>R24C</sup> (an activated version of a cyclin-dependent kinase 4 mutant), c-Myc<sup>T58A</sup> (a stabilized version of the oncogene c-Myc) and H-Ras<sup>G12V</sup> (a constitutively active form of Ras GTPase) have the ability to drive porcine fibroblasts to form tumors when explanted into immunocompromised pigs at different anatomical sites (Adam et al., 2007). These same genetic changes drive human kidney cells, mammary epithelial cells and myoblasts to a tumorigenic state (Kendall et al., 2005), indicating that tumorigenesis in pigs is similar to the process in humans. Genetically engineered porcine tumour cells provided the first method of inducing tumours in a large animal, and hence it is possible to tailor-make tumours of a defined background using the pig. Although this model is limited because the animals need to be immunosuppressed for tumours to grow (akin to xenograft mouse models), pigs nevertheless have a number of clear advantages that make them ideal for preclinical studies of human cancers. The resultant tumours in the pigs could be grown to very large sizes, ideal for a number of preclinical applications. This model can be exploited in different cell types to generate many different types of tumours potentially anywhere in the body (Table 17.4).

### Emerging Cancer Models Utilizing the Pig Phenotype

Basal cell carcinoma is the most prevalent human cancer, with over 750,000 cancers being diagnosed yearly in the USA alone, yet animal models remain limiting owing to molecular and skin type differences between humans and mice. While mouse skin and human skin share many similar features, there are also major differences, which may contribute to the differences in skin tumorigenesis with respect to tumour type and mechanism between the two species. In humans, the three main types of skin cancer are: basal cell carcinomas (BCC), squamous cell carcinomas (SCC) and cutaneous melanomas (CM), with BCC being the most common of the three, representing approximately 70% of all human skin cancers (de Gruijl et al., 2001). In contrast, mice do not develop BCC; the predominant malignant tumour type in mice is SCC (Peto et al., 1975; Bogovski, 1994). In addition, oncogenic Ras has an essential role in mouse skin tumorigenesis while it appears to have only a minor role in human skin cancer (Ananthaswamy and Pierceall, 1990; Pierceall et al., 1991a,b). Thus, mice are not always ideal in vivo models for the study of human skin cancer.

Among experimental animals, porcine skin is most similar to human skin and has been used extensively as a model of human wound healing (Lunney, 2007). More specifically, the porcine integument is morphologically (Montagna and Yun, 1964; Meyer et al., 1978; Monteiro-Riviere and Stromberg, 1985; Monteiro-Riviere, 1986), histochemically (Meyer et al., 1986; Rigal et al., 1991; Woolina et al., 1991), biochemically and biophysically similar to human skin. As such, the pig has been utilized as a model for drug toxicity and percutaneous absorption studies. Pig skin resembles human skin in having a sparse hair coat, a relatively thick epidermis, and similar epidermal turnover kinetics, lipid composition, carbohydrate biochemistry, lipid biophysical properties and arrangement of dermal collagen and elastic fibres (Weinstein, 1966; Forbes, 1967; Montagna, 1967; Meyer et al., 1981, 1982). Reported differences in pigs include a unique interfollicular muscle that spans the triad of the hair follicle (Stromberg

<table>
<thead>
<tr>
<th>Embryonic layer</th>
<th>Cell type transformed</th>
<th>Experimental model</th>
<th>Tumour type induced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoderm</td>
<td>Keratinocytes</td>
<td>In vitro cell transformation</td>
<td>N/A</td>
</tr>
<tr>
<td>Ectoderm</td>
<td>Fibroblasts; mammary, kidney and testes cells</td>
<td>In vitro cell transformation</td>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td>Mesoderm</td>
<td>T cells</td>
<td>Live virus injection</td>
<td>T cell lymphoma</td>
</tr>
</tbody>
</table>

Table 17.4. Porcine cell transformation.
et al., 1981), the presence of apocrine sweat glands only on the body surface (Montagna and Yun, 1964; Monteiro-Riviere and Stromberg, 1985) and a thicker stratum corneum (Montagna and Yun, 1964; Monteiro-Riviere and Stromberg, 1985) and a thicker stratum corneum (Montagna and Yun, 1964; Monteiro-Riviere and Stromberg, 1985). With regard to biochemical similarities between pigs and humans, for example, conservation of the matrix metalloproteinase genes MMP1 and MMP9 is greater between humans and pigs (89% and 85%, respectively) than between humans and mice (80% and 78%, respectively), based on the HomoloGene NCBI (US National Center for Biotechnology Information) database).

As discussed previously, it has been demonstrated that porcine fibroblasts can be transformed in vitro and explanted into the pig to form tumours. Fibroblasts, however, are the primarily transformed cell type in less than 1% of human malignancies (Khavari, 2006). BCC, the most common cancer in the USA, and SCC, the second most common cancer in the USA, arise from keratinocytes (Khavari, 2006). Isolated porcine keratinocytes, the target cell population, can be transformed following the same procedure. Specifically, the co-expression of hTERT, p53G30, cyclin D1, CDK4R24C, c-MycT58A and H-RasG12V is sufficient to drive porcine keratinocytes to form tumours when injected subcutaneously into immunocompromised mice. Further research has demonstrated that expression of only cyclin D1, CDK4R24C, H-RasG12V and c-MycT58A was sufficient to transform both porcine fibroblasts and keratinocytes to a tumorigenic state, indicating that fewer genes are required for successful porcine cell transformation and subsequent tumour formation (K.N. Kuzmuk, 2009, unpublished results).

The establishment of tumours using the pig as a model is possible, provided the animals remain on immunosuppressive therapy. When treatment with immunosuppressive drugs is halted, tumours, regardless of size, regress owing to an overwhelming host immune reaction to the tumour cells. Research using retroviruses as vectors is being conducted to determine whether this approach eliminates the need for immunosuppressed animals. It is theorized that the manipulation of cells in tissue culture during the transformation process makes the cells immunogenic. It has been demonstrated that the injection of a virus encoding mutated H-Ras directly into the mammary fat pads of wild-type rats is tumorigenic (McFarlin and Gould, 2003; McFarlin et al., 2003). For that reason, the direct in vivo injection of retroviruses containing the transgenes required for porcine cell transformation in vitro would be tumorigenic in immunocompetent pigs. To test this hypothesis, viruses expressing the transgenes used to transform both the porcine fibroblasts and keratinocytes (cyclin D1, CDK4R24C, H-RasG12V and c-MycT58A) were injected directly into the pig. Direct retroviral injection produced a low frequency of lymphoma of T cell origin (K.N. Kuzmuk, 2009, unpublished results).

Needs and Opportunities for Expanding the Use of Pig Biomedical Models

Novel approaches to harvesting genomic information to target genetic manipulations coupled with cloning have been identified as targets for further development (Schook et al., 2005b). Emerging technologies such as recombineering and gene trapping combined with relevant, standardized cell lines of targeted modifications could be used for cloning specific pigs for a given human disease. The National Swine Resource and Research Center (NSRRC) at the University of Missouri (http://www.nsrrc.missouri.edu) provides essential support for creating genetic pig models of human diseases. Specifically, NSRRC has established significant resources to assist researchers in creating transgenic pigs, as well as to support the distribution of created models to investigators, thus providing a mechanism for generating and distributing the ‘gold standard’ model for specific diseases or phenotypes.

Finally, the pig will continue to grow as the biomedical model of choice in bioengineering and experimental surgery, and in zoonosis research related to the emergence of new diseases such as swine influenza. With respect to bioengineering and experimental surgery, the growing popularity of the pig versus the dog has continued to rise, and the pig is now the most common large laboratory animal species. The number of pigs used in 2002 in registered research facilities as reported to
the US Department of Agriculture (USDA) was over 68,400, whereas the number of dogs declined from 201,000 in 1984 to 68,200 in 2002 (http://www.aphis.usda.gov/publications). Completion of the pig genome sequencing will only accelerate the popularity and value of swine in biomedical research. The pig is currently being developed as a model to understand the pathogenesis of and immunity to human viral pathogens such as rotavirus, calicivirus and coronavirus (CoV). Saif and co-workers (Costantini et al., 2004) have clearly demonstrated the utility of the pig as a model to understand the mechanisms for ‘super-spreaders’ and the atypical pneumonia and variable diarrhoea induced by the human CoV responsible for severe acute respiratory syndrome (SARS). The porcine model of SARS consists of utilizing the porcine respiratory CoV (PRCV), a spike deletion mutant of the enteric CoV transmissible gastroenteritis virus (TGEV), which shows striking pathogenetic similarities to the SARS CoV in its primary replication in the lung. Further research is justified to compare known immunological differences and similarities between mice, humans and pigs. Current work by Dawson et al. (2008) has revealed that pig immune responses are more similar to human responses than mouse responses for over 80% of the variables compared, and that the mouse immune responses were more similar to human than pig responses in less than 10% of comparisons (Dawson et al., 2008). Genomic tools will continue to push existing animal models to evolve and novel models to be developed (Table 17.5).

### Acknowledgements

The authors acknowledge the intellectual contributions of Chris Counter (Duke University) and Kelly Swanson (University of Illinois at Urbana-Champaign) for their thoughts on building animal models. This work was supported in part by funding from USDA grants AG2003-34480-13172 417, AG2004-34480-14417 and AG2005-24480-15939.

### Table 17.5. Evolution of animal models generated by genomic tools.

<table>
<thead>
<tr>
<th>Characteristic features</th>
<th>Traditional view</th>
<th>Current view</th>
<th>Future view</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to disease</td>
<td>Anatomy, physiology, pathology and responses to therapeutics</td>
<td>Disease characteristics and therapies or devices tested</td>
<td>Selected based on specific disease and therapeutic responses</td>
</tr>
<tr>
<td>Practical considerations</td>
<td>Dietary and housing requirements, husbandry, genetic uniformity and cost</td>
<td>Restricted to gene-rich species (worms, fruit fly, yeasts, rodents)</td>
<td>Emerging genomic profiles of animals with similar disease phenotypes to humans</td>
</tr>
<tr>
<td>Unique features</td>
<td>Clear laws, regulations and policies</td>
<td>Pain and stress protocol issues</td>
<td>Unknown issues in addition to use of new species for biomedical-regulated animal protocols</td>
</tr>
<tr>
<td>Ethical features</td>
<td>Practical and economical but relevance to human phenotype may be questioned</td>
<td>Genetically similar but is phenotype similar?</td>
<td>Ideal owing to recapitulating human condition</td>
</tr>
<tr>
<td>Overall characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Galen (1586) Galeni Librorum Quarta Classis. Apud Iuntas, Venetijs [Venice].


Domestication and Early Development of Breeds

Swine have been domesticated for at least 9000 years (Larson et al., 2007, Chapter 2). Domestication took place in at least two locations (Asia and Europe) with some intermingling of the strains as humans started to migrate across the Europe–Asia landmass. Like all domesticated animals, modern pigs were developed because the species (Sus scrofa) both received and provided a benefit in its relationship with humans (Homo sapiens). Unlike several other species, it was farmers who had established a location who assisted in this process rather than nomadic peoples. Nomadic peoples would be more likely to domesticate grazing animals which could be moved to locations with good forage while established farmers would use domesticated animals in a variety of ways.

It is likely that human attempts to ‘improve’ swine began soon after domestication. If an animal has a utility, breeding those that appear to perform that utility in a superior way is natural. Pigs have progressed through being a consumer of the ‘leftovers’ of human society, to pigs that provided a high amount of lard, to today’s pig, which consumes a carefully constructed diet of harvested feeds in order to produce animal protein as a source of human food as efficiently as possible.

Because local areas have different environments, different needs and different genetic stocks to initiate genetic improvement, it is not surprising that the pigs of the world are fairly diverse. Breeds are natural outgrowths of that diversity. Several aspects of domestication, as it applies to genetic diversity, are discussed in Chapter 2 of the previous edition of this book (Jones, 1998). It is generally agreed that pigs were domesticated to serve as a source of protein for human consumption and because of the unique taste of pork.

What Is a Breed?

The term ‘breed’ is a difficult one to define precisely because it means different things to different people. A breed might be defined as a group of animals with similar physical
characteristics (such as colour, body type, etc.). However, there are breeds that contain wide variation in such characteristics, while the members of other, different, breeds may be quite similar. There is general agreement that the concept of a breed denotes common ancestry, yet some organizations that protect the purity of a breed choose, periodically, to open their herd books to animals from exotic ancestry. Lush (1994), quoting from Lloyd-Jones (1915), makes the following observation:

A breed is a group of domestic animals, termed such by common consent of the breeders, a term which arose among breeders of livestock, created one might say for their own use, and no one is warranted in assigning to this word a scientific definition and in calling the breeders wrong when they deviated from the formulated definition. It is their word and the breeders’ common usage is what we must accept as the correct definition.

Wright (1977), in his description of breed formation, describes a breed as something that arises more rapidly than normal evolutionary processes would dictate, but more slowly than would be true in the laboratory. Breed development probably covers almost the entire range of rates in that spectrum. Some breeds arise almost entirely through natural forces (e.g. Chinese breeds of pigs, which developed in localized areas such as valleys where there were human communities). Other breeds and lines were developed by human managers in a highly directed fashion (e.g. breeds developed by the purposeful crossing of other breeds, research herds or company lines).

Breeds are easy to recognize in many developed countries because organizations have arisen to protect the breed purity and to pursue its improvement. These ‘breed societies’ originated in Great Britain during the early part of the 19th century (Willham, 1987), and spread to other countries, most notably the USA. Some breed societies are large businesses with many thousands of pigs registered, while others number their annual registrations with three digits and are organized by a single individual. The American Livestock Breeds Conservancy lists several pig breeds on their Threatened (<1000 annual registrations) or Critical (<200 annual registrations) lists, including Gloucestershire Old Spot, Large Black and Red Wattle.

Breed associations may periodically decide to open their herd books to allow registration of individuals from another breed in order to bring some desired characteristics into the breed. In the USA, this has happened with the Poland China, Spots and Chester White breeds in recent years. Additionally, pigs that represent the breed but are from other countries are, at times, registered. This provides the benefit of bringing in new genetic material and reducing the effects of inbreeding.

### Identifying Breeds of Swine for Inclusion

A difficulty associated with describing pig breeds is identifying which breeds to include in the discussion. Several hundred breeds have been identified (Mason, 1996), although many of these are a national derivative of a breed that is imported from its native country. It would be desirable to identify all of the ‘important’ breeds around the world. This task is rendered nearly impossible because of the difficulty in defining ‘important’. Breeds with high census numbers are likely to be considered important, but there may be breeds with low numbers which are important either historically or as a source of unique genetic material for some future use. This raises the issue of genetic conservation – or preservation of genetic material. Breeds may be conserved for economic, scientific or cultural reasons (Committee on Managing Global Genetic Resources, 1993; Chapter 13). There are organizations that are devoted to the identification and preservation of breeds that have declined in numbers (Rare Breeds Survival Trust, American Livestock Breeds Conservancy). These organizations identify breeds that are in danger of becoming extinct. There is value in preserving biodiversity through the preservation of breeds (Animal Genetic Resources Group, 2009). There is
also an effort to maintain a gene bank for European endangered breeds (Glodek, 2001; Labroue et al., 2001a,c; Ollivier et al., 2001b), and the USDA National Animal Germplasm Program was established to effectively conserve and facilitate uses of animal genetic resources.

Ideally, a description of breeds would include those breeds with well-understood origin and well-researched characteristics. However, many breeds are important, at least in some parts of the world, even though they fit neither of these characteristics. The origin of many breeds has been lost owing to inadequate historical records, or is irrelevant as a result of the large-scale introduction of individuals from outside the breed. There is considerable research information for some breeds but little, or none, for others.

We are left with only imperfect methods for identifying breeds to include in a publication such as this. We have chosen the following approach. Breeds are listed if they were included in Briggs and Briggs (1980), or are described at the Breeds of Livestock web site maintained by Oklahoma State University (http://www.ansi.okstate.edu/breeds/swine/).

**Previous Work on Breeds of Swine**

There have been many other summaries of the status and value of swine breeds in the world (Freedeen, 1957; Omtvedt, 1974; Sellier, 1976, 1988; Fahmy and Holtmann, 1977a; Briggs and Briggs, 1980; Johnson, 1981; Fahmy and Moride, 1983; Legault, 1985; Sutherland et al., 1985; Buchanan, 1987; McLaren, 1990; Sellier and Rothschild, 1991; Ollivier and Molénat, 1992; Jones, 1998; Ollivier et al., 2001a). All of these describe the considerable diversity that exists among existing breeds, and attempt to describe the value of breeds for commercial swine production. Additionally, there have been several published descriptions of the very interesting Chinese breeds of swine (Xue, 1991; Mercer and Hoste, 1994; Mao, 1995). These Chinese breeds are of special interest because many of them are very highly prolific (Sellier and Legault, 1986; Bidanel et al., 1989a,b,c).

**Describing Breeds of Swine**

The breeds are described here in a series of tables and colour plates (Plates 4–27). In Tables 18.1–18.5 breeds are included that have been developed in various parts of the world. Where information is available, breeds are described for size, lean-to-fat ratio and prolificacy. The descriptors for some of the breeds have been derived from a review of breed comparison research in North America (Johnson, 1981) and from a French evaluation of the Meishan breed (Bidanel et al., 1989a,b,c). Descriptors for the other breeds are highly subjective. They probably reflect performance that is dependent upon the environment in which the breeds are used, and may not indicate the performance levels that would be achieved if all of the breeds were managed in a uniform environment.

Tables 18.6–18.9 are included to provide references for research information concerning the breeds. Numerous research papers, published in refereed journals, are included with a list of the breeds evaluated in the project described in each paper. The four tables divide the experiments into those that: (i) evaluate breeds of American or British origin; (ii) include breeds of Asian origin; (iii) include at least one European local breed; and (iv) include the Piétrain breed.

There were intensive efforts at many state experiment stations in North America to compare breeds of swine during the 1960s, 1970s and 1980s. This is easy to observe in Table 18.6. This was followed by increased interest in the highly prolific breeds from China, with much research done by scientists in North America, Europe and Asia (Table 18.7). The Piétrain breed is interesting because of its very high ratio of lean to fat, and it has also been studied heavily (Table 18.9). The allele for porcine stress syndrome (PSS) is also present in high frequency in the breed.
Table 18.1. Breeds of pigs originating in North and South America.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Place of origin</th>
<th>Breed origin</th>
<th>Distribution</th>
<th>Colour</th>
<th>Ears</th>
<th>Other</th>
<th>Size and growtha</th>
<th>Lean-to-fat ratio</th>
<th>Prolificacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chester White&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td>Pennsylvania, USA</td>
<td>Yorkshire, Lincolnshire</td>
<td>North America</td>
<td>White</td>
<td>Droop</td>
<td></td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Duroc&lt;sup&gt;d&lt;/sup&gt;</td>
<td>New Jersey and New York, USA</td>
<td>Jersey Red, Berkshire, other red pigs</td>
<td>North America, Europe</td>
<td>Red</td>
<td>Droop</td>
<td></td>
<td>Moderate to high</td>
<td>Moderate to high</td>
<td>Moderate</td>
</tr>
<tr>
<td>Hampshire&lt;sup&gt;e,o&lt;/sup&gt;</td>
<td>Kentucky, USA</td>
<td>Old English</td>
<td>North America, Europe</td>
<td>Black, white belt</td>
<td>Erect</td>
<td></td>
<td>Moderate to high</td>
<td>Moderate to high</td>
<td>Moderate</td>
</tr>
<tr>
<td>Hereford&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Missouri, USA</td>
<td>Duroc, Chester White, Poland China</td>
<td>USA</td>
<td>Red, white face, belly, legs</td>
<td>Erect</td>
<td></td>
<td>Moderate to high</td>
<td>Moderate to high</td>
<td>Moderate</td>
</tr>
<tr>
<td>Lacombe&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Alberta, Canada</td>
<td>Landrace, Berkshire, Chester White</td>
<td>Canada, USA, Mexico, Europe</td>
<td>White</td>
<td>Droop</td>
<td>Composite breed developed by Canadian Dept Agriculture</td>
<td>Moderate</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Moura</td>
<td>Brazil</td>
<td>Duroc, Canastra, Canastra</td>
<td>Brazil</td>
<td>Blue roan</td>
<td>Droop</td>
<td></td>
<td>Rare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mulefoot</td>
<td>USA</td>
<td>Duroc</td>
<td>USA</td>
<td>Black</td>
<td>Droop</td>
<td>Rare, single toed</td>
<td>Rare</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Ossabaw Island&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Ossabaw Island (off coast of Georgia, USA)</td>
<td>Descendants of Spanish pigs brought in 1500s</td>
<td>Ossabaw Island</td>
<td>Varied</td>
<td>Erect</td>
<td>Used for diabetes research, rare</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Poland China&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Ohio, USA</td>
<td>Bedford, Berkshire, others</td>
<td>North America</td>
<td>Black, six white points</td>
<td>Droop</td>
<td></td>
<td>High</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Spotted&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Ohio, USA (from England)</td>
<td>Poland China Large White</td>
<td>North America, Europe</td>
<td>Black, white spots</td>
<td>Droop</td>
<td></td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Yorkshire&lt;sup&gt;j&lt;/sup&gt;</td>
<td>Ohio, USA</td>
<td>Large White</td>
<td>North America, Europe</td>
<td>White</td>
<td>Erect</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Pampa-Rocha</td>
<td>Uruguay</td>
<td>South America</td>
<td>South America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plau&lt;sup&gt;k&lt;/sup&gt;</td>
<td>Brazil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Terms such as high, moderate and low are necessarily relative. For breeds that have been well studied, the terms reflect information derived from objective breed comparisons. For other breeds, the terms are more subjective.

<sup>b</sup>Performance information inferred from Johnson (1981).

<sup>c</sup>Plate 4; <sup>d</sup>Plate 5; <sup>e</sup>Plate 6; <sup>f</sup>Plate 7; <sup>g</sup>Plate 8; <sup>h</sup>Plate 9; <sup>i</sup>Plate 10.

<sup>j</sup>The Yorkshire breed originated directly from the British Large White. The intermingling of the breeds continues as pigs from the Large White breed are sometimes registered as Yorkshires.

<sup>k</sup>Plate 11; <sup>l</sup>Plate 12.
Table 18.2. Breeds of pigs originating in the UK.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Place of origin</th>
<th>Breed origin</th>
<th>Distribution</th>
<th>Colour</th>
<th>Ears</th>
<th>Other</th>
<th>Size and growth(^a)</th>
<th>Lean-to-fat ratio</th>
<th>Prolificacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berkshire(^b)</td>
<td>England</td>
<td></td>
<td>North America, Europe</td>
<td>Black, six white points</td>
<td>Erect</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>British Lop</td>
<td>England</td>
<td>White pigs of Wales, Cumberland and Ulster</td>
<td>England</td>
<td>White</td>
<td>Droop</td>
<td>Moderate to high</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Gloucestershire Old Spot(^c)</td>
<td>England</td>
<td>Local breeds</td>
<td>England</td>
<td>White, black spots</td>
<td>Droop</td>
<td>Moderate to high</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Large Black</td>
<td>England</td>
<td>Devon, Cornwall and East Anglia</td>
<td>England, USA, South Africa, Australia</td>
<td>Black</td>
<td>Droop</td>
<td>Endangered</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Large White</td>
<td>England</td>
<td>Cumberland, Leicestershire, Middle and Small White</td>
<td>Europe, North America, Asia</td>
<td>White</td>
<td>Erect</td>
<td>Moderate</td>
<td>High</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Middle White(^d)</td>
<td>England</td>
<td>Berkshire, Tamworth, others</td>
<td>England</td>
<td>White</td>
<td>Erect</td>
<td>Rare</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Oxford Sandy and Black</td>
<td>England</td>
<td>Berkshire, Tamworth, others</td>
<td>England</td>
<td>Sandy with black blotches</td>
<td>Erect</td>
<td>Rare</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Saddleback</td>
<td>England</td>
<td>Essex, Wessex</td>
<td>England</td>
<td>Black, white belt</td>
<td>Erect</td>
<td>Moderate</td>
<td>Moderate to high</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Tamworth(^e)</td>
<td>England</td>
<td>England, North America</td>
<td></td>
<td>Red</td>
<td>Erect</td>
<td>Moderate to high</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

\(^a\)Terms such as high, moderate and low are necessarily relative. For breeds that have been well studied, the terms reflect information derived from objective breed comparisons. For other breeds, the terms are more subjective.

\(^b\)Plate 13; \(^c\)Plate 14; \(^d\)Plate 15; \(^e\)Plate 16.
Table 18.3. Breeds of pigs originating in Europe.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Place of origin</th>
<th>Breed origin</th>
<th>Distribution</th>
<th>Colour</th>
<th>Ears</th>
<th>Other</th>
<th>Size and growth</th>
<th>Lean-to-fat ratio</th>
<th>Prolificacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angeln Saddleback</td>
<td>Germany</td>
<td>Landrace, Wessex-Saddleback</td>
<td>Europe</td>
<td>Black, white belt</td>
<td>Droop</td>
<td>Nearly extinct</td>
<td>Moderate</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Bazna</td>
<td>Romania</td>
<td>Berkshire, Mangalitsa</td>
<td>Europe</td>
<td>Black, white belt</td>
<td>Droop</td>
<td>Moderate</td>
<td>Low</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Belarus Black Pied</td>
<td>Belarus</td>
<td>Large White, Large Black, Berkshire, Middle White, local breeds</td>
<td>Belarus</td>
<td>Usually black</td>
<td>Droop</td>
<td>Moderate</td>
<td>Low</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Bentheim Black Pied</td>
<td>Germany</td>
<td>Berkshire, Cornwall, local breeds</td>
<td>Germany</td>
<td>White, black spots</td>
<td>Droop</td>
<td>Nearly extinct</td>
<td>Moderate</td>
<td>Low</td>
<td>Moderate to high</td>
</tr>
<tr>
<td>Black Slavonian White</td>
<td>Croatia</td>
<td>Berkshire, Poland China, Black Mangalitsa</td>
<td>Croatia</td>
<td>Black</td>
<td>Droop</td>
<td>Nearly extinct</td>
<td>Moderate</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Bulgarian White</td>
<td>Bulgaria</td>
<td>Large White, Edelschwein, native breeds</td>
<td>Bulgaria</td>
<td>White</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Improved White</td>
<td>Czech Republic</td>
<td>Large White, Edelschwein, Landrace</td>
<td>Czech Republic</td>
<td>White</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dermantsi Pied Iberian</td>
<td>Bulgaria</td>
<td>Berkshire, Mangalitsa, local breeds</td>
<td>Bulgaria</td>
<td>Black and white</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iberian</td>
<td>Spain</td>
<td>Extremadura Red, Jabugo Spotted, Black Iberian</td>
<td>Spain</td>
<td>Varies</td>
<td>Droop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krskopolje</td>
<td>Slovenia</td>
<td>Suffolk, Carniola, Berkshire, Yorkshire</td>
<td>Slovenia</td>
<td>Black, white belt</td>
<td>Droop</td>
<td>Rare</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Laconie</td>
<td>Haiti</td>
<td>Hampshire, Piétrain, Large White</td>
<td>Many strains in different countries</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landrace²</td>
<td>Denmark</td>
<td>Large White, native breeds</td>
<td>North America, White, Europe</td>
<td>White</td>
<td>Droop</td>
<td></td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Breed</td>
<td>Origin</td>
<td>Type</td>
<td>Description</td>
<td>Ear Posture</td>
<td>Hair</td>
<td>Meat Production</td>
<td>infertility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td>------</td>
<td>------------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuanian Native</td>
<td>Lithuania</td>
<td>Local breeds</td>
<td>Lithuania</td>
<td>Spotted, varied colours</td>
<td>Erect</td>
<td>Rare</td>
<td>Moderate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mangalitsa</td>
<td>Austria-Hungary</td>
<td>Local breeds</td>
<td>Europe</td>
<td>Varied</td>
<td>Erect</td>
<td></td>
<td>Low to moderate</td>
<td>Moderate to high</td>
<td></td>
</tr>
<tr>
<td>Mora</td>
<td>Ravenna province, Italy</td>
<td>Indigenous breed of region</td>
<td>Italy</td>
<td>Dark brown</td>
<td>Droop</td>
<td></td>
<td>Moderate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romagnola</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piétrain</td>
<td>Belgium</td>
<td></td>
<td>Europe</td>
<td>White, black spots</td>
<td>Erect</td>
<td></td>
<td>Moderate</td>
<td>Very high</td>
<td>Moderate</td>
</tr>
<tr>
<td>Swabian-Hall Swine</td>
<td>Germany</td>
<td></td>
<td>Germany</td>
<td>White, black head and rump</td>
<td>Droop</td>
<td></td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Turopolje</td>
<td>Croatia</td>
<td>Local breeds, Siska, Krskopoljski</td>
<td>Croatia</td>
<td>White, black spots</td>
<td>Droop</td>
<td>Nearing extinction</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Terms such as high, moderate and low are necessarily relative. For breeds that have been well studied, the terms reflect information derived from objective breed comparisons. For other breeds, the terms are more subjective.

| Plate 17; cPlate 18; dPlate 19; ePlate 20; fPlate 21. |
Table 18.4. Breeds of pigs originating in Asia.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Place of origin</th>
<th>Breed origin</th>
<th>Distribution</th>
<th>Colour</th>
<th>Ears</th>
<th>Other</th>
<th>Size and growth $^a$</th>
<th>Lean-to-fat ratio</th>
<th>Prolificacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ba Xuyen</td>
<td>Vietnam</td>
<td>Berkshire, Chinese breeds</td>
<td>Asia</td>
<td>White, black spots</td>
<td>Erect</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Beijing Black</td>
<td>China</td>
<td>Berkshire, Large White, local breeds</td>
<td>China</td>
<td>Usually black</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>Black and white</td>
<td>Droop</td>
<td>Imported to USA in 1989</td>
<td>Low</td>
<td>Very low</td>
<td>Very high</td>
</tr>
<tr>
<td>Da Min$^e$</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>Black</td>
<td>Droop</td>
<td>Imported to USA in 1989</td>
<td>Low</td>
<td>Very low</td>
<td>Very high</td>
</tr>
<tr>
<td>Fengjing$^c$</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black head and rump</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Hezuo</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Very low</td>
<td>Low</td>
</tr>
<tr>
<td>Jinhua</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Kele</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Very low</td>
<td>Low</td>
</tr>
<tr>
<td>Luchuan$^d$</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Very low</td>
<td>Low</td>
</tr>
<tr>
<td>Large Black-White</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Very high</td>
</tr>
<tr>
<td>Meishan$^{e,f}$</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China, North America</td>
<td>Black, white points</td>
<td>Droop</td>
<td>Imported to USA in 1989</td>
<td>Low</td>
<td>Very low</td>
<td>Very high</td>
</tr>
<tr>
<td>Minzhu</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China, North America</td>
<td>Black</td>
<td>Droop</td>
<td>Imported to USA in 1989</td>
<td>Low</td>
<td>Very low</td>
<td>Very high</td>
</tr>
<tr>
<td>Mong Caï$^g$</td>
<td>Vietnam</td>
<td>Berkshire, Chinese breeds</td>
<td>Vietnam</td>
<td>Black and white</td>
<td>Erect</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Neijiang</td>
<td>China</td>
<td>Berkshire, Chinese breeds</td>
<td>China</td>
<td>White, black spots</td>
<td>Droop</td>
<td>Very early sexual maturity</td>
<td>Low</td>
<td>Very low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Breeds</td>
<td>Country</td>
<td>Origin</td>
<td>Coat Color</td>
<td>Ears</td>
<td>Droop</td>
<td>Adaptation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------------------------------</td>
<td>------------</td>
<td>--------</td>
<td>---------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ningxiang</td>
<td>China</td>
<td>Local breeds of Hunan Province</td>
<td>Black and white</td>
<td>Droop</td>
<td>Low</td>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thuoc Nhieu</td>
<td>Vietnam</td>
<td>Bo Xu, Yorkshire</td>
<td>White</td>
<td>Erect</td>
<td>Very low</td>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibetan</td>
<td>Tibet</td>
<td>China</td>
<td>Black</td>
<td>Erect</td>
<td>Low</td>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tongcheng Pot Belly</td>
<td>Vietnam</td>
<td>Breed of Vietnam</td>
<td>Black and white</td>
<td>Erect</td>
<td>Very low</td>
<td>Imported in 1986 as pets</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Terms such as high, moderate and low are necessarily relative. For breeds that have been well studied, the terms reflect information derived from objective breed comparisons. For other breeds, the terms are more subjective.

Plate 22; Plate 23; Plate 24; Plate 25.
Information for Meishan derived from Bidanel et al., 1989a,b,c.
Plate 26; Plate 27.
<table>
<thead>
<tr>
<th>Breed</th>
<th>Place of origin</th>
<th>Breed origin</th>
<th>Distribution</th>
<th>Colour</th>
<th>Ears</th>
<th>Other</th>
<th>Size and growth(^a)</th>
<th>Lean-to-fat ratio</th>
<th>Prolificacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arapawa Island</td>
<td>New Zealand</td>
<td>New Zealand</td>
<td>Varied</td>
<td>Feral</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bantu</td>
<td>South Africa</td>
<td>Guinea coast of Africa</td>
<td>Usually brown</td>
<td>Black</td>
<td>Rare</td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinea</td>
<td>New Zealand</td>
<td>New Zealand</td>
<td>Varied</td>
<td>Erect</td>
<td>Tassels on lower jaw</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Kunekune</td>
<td>New Zealand</td>
<td>From introduced pigs in early 1800s</td>
<td>Varied</td>
<td>Erect</td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Mukota</td>
<td>Rhodesia, Zimbabwe</td>
<td>Africa</td>
<td>Black</td>
<td></td>
<td>Adapted to harsh tropical environment</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Red Wattle</td>
<td>New Caledonia</td>
<td>North America</td>
<td>Red</td>
<td>Erect</td>
<td>Wattles on jaw</td>
<td>Low</td>
<td>High</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Terms such as high, moderate and low are necessarily relative. For breeds that have been well studied, the terms reflect information derived from objective breed comparisons. For other breeds, the terms are more subjective.
Table 18.6. References for research information comparing American and British breeds.

<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dufour and Fahmy, 1975</td>
<td>Agriculture Canada</td>
<td>Hampshire, Lacombe, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Fahmy, 1972; Fahmy et al., 1975, 1976, 1978</td>
<td>Agriculture Canada</td>
<td>Berkshire, Duroc, Hampshire, Lacombe, Landrace, Large Black, Yorkshire</td>
</tr>
<tr>
<td>Jeremiah et al., 1999</td>
<td>Agriculture Canada</td>
<td>Duroc, Hampshire, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Luiting et al., 1995</td>
<td>Agricultural University of Norway</td>
<td>Duroc, Landrace</td>
</tr>
<tr>
<td>Jones et al., 1980; Richmond et al., 1979</td>
<td>Alberta, Canada</td>
<td>Duroc, Hampshire, Lacombe, Yorkshire</td>
</tr>
<tr>
<td>Bereskin and Davey, 1976; Davey and Bereskin, 1977; Bereskin and Steele, 1986</td>
<td>BARC, USA</td>
<td>Duroc, Yorkshire</td>
</tr>
<tr>
<td>Park and Yi, 2002</td>
<td>Daejeon, Korea</td>
<td>Duroc, Yorkshire</td>
</tr>
<tr>
<td>Cameron, 1990; Cameron et al., 1990</td>
<td>Edinburgh, UK</td>
<td>Duroc, Landrace</td>
</tr>
<tr>
<td>Kennedy and Conlon, 1978</td>
<td>Guelph, Canada</td>
<td>Duroc, Hampshire, Landrace</td>
</tr>
<tr>
<td>Ruusunen and Puolanne, 1997</td>
<td>Helsinki, Finland</td>
<td>Hampshire, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Zheng et al., 2002</td>
<td>Hunan, China</td>
<td>Duroc, Landrace, Large White</td>
</tr>
<tr>
<td>Baas et al., 1992a,b</td>
<td>Iowa, USA</td>
<td>Hampshire, Landrace</td>
</tr>
<tr>
<td>Meeker et al., 1985, 1987</td>
<td>Iowa, USA</td>
<td>Duroc, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Schneider et al., 1982a,b</td>
<td>Iowa, USA</td>
<td>Chester White, Duroc, Hampshire, Yorkshire</td>
</tr>
<tr>
<td>Zhang et al., 2007</td>
<td>Iowa, USA</td>
<td>Berkshire, Chester White, Duroc, Hampshire, Landrace, Poland China, Spot</td>
</tr>
<tr>
<td>Shu, 1996</td>
<td>Jiangxi, China</td>
<td>Duroc, Landrace, Large White</td>
</tr>
<tr>
<td>Fahmy and Holtmann, 1977a,b; Langlois and Minvielle, 1989a,b</td>
<td>Laval University, Canada</td>
<td>Berkshire, Duroc, Hampshire, Lacombe, Landrace, Large Black, Tamworth</td>
</tr>
<tr>
<td>Holtmann et al., 1975</td>
<td>Laval University, Canada</td>
<td>Duroc, Landrace, Large White</td>
</tr>
<tr>
<td>Purchas et al., 1990; Smith et al., 1988, 1990</td>
<td>Massey University, New Zealand</td>
<td>Duroc, Hampshire, Landrace, Large White</td>
</tr>
<tr>
<td>Suzuki and Watanabe, 1996; Suzuki et al., 2001</td>
<td>Miyagi, Japan</td>
<td>Duroc, Landrace, Large White</td>
</tr>
<tr>
<td>Suzuki et al., 2003</td>
<td>Miyagi, Japan</td>
<td>Berkshire, Duroc</td>
</tr>
<tr>
<td>Lishman et al., 1975</td>
<td>Newcastle, UK</td>
<td>Hampshire, Large White, Landrace</td>
</tr>
<tr>
<td>Neely et al., 1980; Neely and Robison, 1983; Toelle and Robison, 1983</td>
<td>North Carolina, USA</td>
<td>Duroc, Yorkshire</td>
</tr>
<tr>
<td>Nelson and Robison, 1976</td>
<td>North Carolina, USA</td>
<td>Duroc, Hampshire, Yorkshire</td>
</tr>
<tr>
<td>Neal and Irvin, 1994</td>
<td>Ohio, USA</td>
<td>Large White, Yorkshire</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shurson and Irvin, 1992</td>
<td>Ohio, USA</td>
<td>Duroc, Landrace</td>
</tr>
<tr>
<td>Yen et al., 1987</td>
<td>Ohio, USA</td>
<td>Chester White, Duroc, Hampshire, Landrace, Spot, Yorkshire</td>
</tr>
<tr>
<td>Hutchens et al., 1982; Fent et al., 1983; Buchanan and Johnson, 1984; Gaugler et al., 1984; McLaren et al., 1987a,b,c,d</td>
<td>Oklahoma, USA</td>
<td>Duroc, Landrace, Spot, Yorkshire</td>
</tr>
<tr>
<td>Johnson and Omtvedt, 1973, 1975; Johnson et al., 1973, 1978; Young et al. 1976a,b; Wilson et al., 1977; Wilson and Johnson, 1981a,b</td>
<td>Oklahoma, USA</td>
<td>Duroc, Hampshire, Yorkshire</td>
</tr>
<tr>
<td>Goenaga and Carden, 1979</td>
<td>Pergamino, Argentina</td>
<td>Duroc, Hampshire, Landrace</td>
</tr>
<tr>
<td>Kim et al., 1983</td>
<td>Seoul National University, Korea</td>
<td>Duroc, Hampshire, Landrace, Spot, Yorkshire</td>
</tr>
<tr>
<td>Kim et al., 2007</td>
<td>Seoul National University, Korea</td>
<td>Berkshire, Duroc, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Tan and Chen, 1983</td>
<td>Singapore</td>
<td>Duroc, Hampshire, Landrace, Poland China, Yorkshire</td>
</tr>
<tr>
<td>Lee et al., 1986</td>
<td>Suweon, Korea</td>
<td>Lacombe, Large White</td>
</tr>
<tr>
<td>Edwards et al., 1992</td>
<td>Terrington, UK</td>
<td>Duroc, Large White</td>
</tr>
<tr>
<td>Chang et al., 1998</td>
<td>TLRI,^b Taiwan</td>
<td>Berkshire, Duroc, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Martinez Gamba et al., 2006</td>
<td>Universidad Nacional Autónoma de México</td>
<td>Duroc, Landrace, Yorkshire</td>
</tr>
<tr>
<td>Blasco et al., 1994</td>
<td>Universidad Politécnica de Valencia, Duroc, Landrace, Large White Spain</td>
<td></td>
</tr>
</tbody>
</table>

^aHenry A. Wallace Beltsville Agricultural Research Center, Maryland (US Department of Agriculture Agricultural Research Service).

^bTaiwan Livestock Research Institute.
Table 18.7. References for research information that includes a Chinese breed.

<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xu et al., 1998</td>
<td>Alberta, Canada</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Kouba et al., 1997</td>
<td>ENSAR, France</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Bazer et al., 1988</td>
<td>Florida</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Xu et al., 2009</td>
<td>Huazhong Agricultural University, China</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Gerfen et al., 1994</td>
<td>Illinois, USA</td>
<td>Fengjing, Meishan, Yorkshire</td>
</tr>
<tr>
<td>Wolter et al., 2000</td>
<td>Illinois, USA</td>
<td>Duroc, Landrace, Meishan, Yorkshire</td>
</tr>
<tr>
<td>Zou et al., 1992</td>
<td>Illinois, USA</td>
<td>Hampshire, Landrace, Meishan, Yorkshire</td>
</tr>
<tr>
<td>Bidanel, 1990; Bidanel et al., 1990a,b, 1991</td>
<td>INRA, France</td>
<td>Jiaxing, Jinhua, Meishan, European breeds</td>
</tr>
<tr>
<td>Canario et al., 2006, 2009; Laloé et al., 2006</td>
<td>INRA, France</td>
<td>Duroc, Laconie, Large White, Meishan</td>
</tr>
<tr>
<td>Geverink et al., 2006; Foury et al., 2007</td>
<td>INRA, France</td>
<td>Duroc, Landrace, Large White, Meishan, Piétrain</td>
</tr>
<tr>
<td>Legault and Caritez, 1982; Legault et al., 1982, 1984; Rombauts et al., 1982</td>
<td>INRA, France</td>
<td>Jiaxing, Jinhua, Large White, Landrace, Meishan</td>
</tr>
<tr>
<td>Legault et al., 1985</td>
<td>INRA, France</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Bidanel, 1989, 1993; Bidanel et al., 1989a,b,c, 1993; Després et al., 1992; Mandonnet et al., 1992; Terqui et al., 1992; Camara et al., 1994; Downey and Driancourt, 1994; Reviere et al., 1997; Driancourt et al., 1998; Mourot and Kouba, 1999; Lefaucheur and Ecolan, 2005</td>
<td>INRA, France</td>
<td>Jia Xiu, Large White, Meishan, Piétrain</td>
</tr>
<tr>
<td>Touraille et al., 1989</td>
<td>INRA, France</td>
<td>Meishan, Jiaxing, Jinhua, European breeds</td>
</tr>
<tr>
<td>Vrillon and Caritez, 2000</td>
<td>INRA, France</td>
<td>Meishan, Yorkshire</td>
</tr>
<tr>
<td>Halbur et al., 1998</td>
<td>Iowa, USA</td>
<td>Duroc, Hampshire, Meishan</td>
</tr>
<tr>
<td>Youngs et al., 1993; Biensen et al., 1998; Vonnahme et al., 2002</td>
<td>Iowa, USA</td>
<td>Meishan, Yorkshire</td>
</tr>
<tr>
<td>Gispert et al., 2007</td>
<td>IRTA, Spain</td>
<td>Duroc, Large White, Landrace, Meishan, Piétrain</td>
</tr>
<tr>
<td>Franck et al., 1998</td>
<td>Lyon, France</td>
<td>Duroc, Large White, Meishan, Piétrain</td>
</tr>
<tr>
<td>Biggs et al., 1993; Hunter et al., 1993; Hunter, 1994</td>
<td>Nottingham, UK</td>
<td>Landrace, Large White, Meishan</td>
</tr>
<tr>
<td>Lee et al., 1994</td>
<td>Roslin, UK</td>
<td>Large White, Meishan</td>
</tr>
<tr>
<td>Wilmut et al., 1992; Ashworth et al., 1994</td>
<td>Rowett Institute, UK</td>
<td>Landrace, Large White, Meishan</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al., 2001</td>
<td>Seoul National University, Korea</td>
<td>Landrace, Large White, Meishan</td>
</tr>
<tr>
<td>Yen et al., 2001</td>
<td>Taichung, Taiwan</td>
<td>Duroc, Taoyuan</td>
</tr>
<tr>
<td>Tzeng, 1991</td>
<td>Taiwan Livestock Research Institute</td>
<td>Landrace, Taoyuan</td>
</tr>
<tr>
<td>Young, 1994</td>
<td>USMARC, USA</td>
<td>Duroc, Fengjing, Meishan, Minzhu</td>
</tr>
</tbody>
</table>

*aEcole Nationale Supérieure Agronomique de Rennes.
*bInstitut National de la Recherche Agronomique.
*cInstitut de Recerca i Tecnologia Agroalimentàries.
*dUS Meat Animal Research Center.
Table 18.8. References for research information that includes a local breed.

<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Franci et al., 2001</td>
<td>Dipartimento di Scienze Zootecniche, Italy</td>
<td>Italian local breeds, Large White</td>
</tr>
<tr>
<td>Glodek et al., 2001</td>
<td>Göttingen, Germany</td>
<td>German local breeds, Piétrain</td>
</tr>
<tr>
<td>Labroue et al., 2000, 2001b</td>
<td>ITP, France</td>
<td>French local breeds, Large White</td>
</tr>
<tr>
<td>Gourdine et al., 2006</td>
<td>INRA, France</td>
<td>Creole, Large White</td>
</tr>
<tr>
<td>Barba et al., 2001</td>
<td>Universidad de Córdoba, Córdoba, Spain</td>
<td>Spanish and Iberian local breeds</td>
</tr>
</tbody>
</table>

*Institut Technique du Porc.
*Institut National de la Recherche Agronomique.

Table 18.9. References for research information that includes Piétrain.

<table>
<thead>
<tr>
<th>Reference(s)</th>
<th>Location</th>
<th>Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castaing et al., 2002</td>
<td>ADAESQ, France</td>
<td>Duroc, Large White, Piétrain</td>
</tr>
<tr>
<td>Mathies et al., 2008</td>
<td>Berlin, Germany</td>
<td>Duroc, Piétrain</td>
</tr>
<tr>
<td>Wood and Lister, 1973; Moody et al., 1978</td>
<td>Bristol, UK</td>
<td>Large White, Piétrain</td>
</tr>
<tr>
<td>Davies, 1974a,b,c; King et al., 1975</td>
<td>Edinburgh, UK</td>
<td>Landrace, Piétrain, Yorkshire</td>
</tr>
<tr>
<td>Glodek et al., 2001</td>
<td>Göttingen, Germany</td>
<td>German local breeds, Piétrain</td>
</tr>
<tr>
<td>Bout et al., 1988; Sellier et al., 1988; Caugant et al., 1989; Guebléz et al., 1993a,b; Aubry et al., 2001</td>
<td>INRA, France</td>
<td>Landrace, Large White, Piétrain</td>
</tr>
<tr>
<td>Legault et al., 1987; Monin et al., 2003; Tiran et al., 2003</td>
<td>INRA, France</td>
<td>Large White, Piétrain</td>
</tr>
<tr>
<td>Touraille et al., 1989</td>
<td>INRA, France</td>
<td>Jia Xiu, Large White, Meishan, Piétrain</td>
</tr>
<tr>
<td>Gispert et al., 2007</td>
<td>IRTA, Spain</td>
<td>Duroc, Large White, Landrace, Meishan, Piétrain</td>
</tr>
<tr>
<td>Franck et al., 1998</td>
<td>Lyon, France</td>
<td>Duroc, Large White, Meishan, Piétrain</td>
</tr>
<tr>
<td>Wolf et al., 2006</td>
<td>Prague, Czech Republic</td>
<td>Duroc, Hampshire, Large White, Landrace, Piétrain, Yorkshire</td>
</tr>
<tr>
<td>Maurer et al., 1985; Young et al., 1989</td>
<td>USMARC, USA</td>
<td>Chester White, Duroc, Hampshire, Landrace, Large White, Piétrain, Spot, Yorkshire</td>
</tr>
<tr>
<td>Curran et al., 1972; Lean et al., 1972</td>
<td>Wye College, UK</td>
<td>Hampshire, Large White, Landrace, Piétrain</td>
</tr>
</tbody>
</table>

*Association pour le Développement Agro-Environnemental du Sud-Ouest.
*Institut National de la Recherche Agronomique.
*Institut de Recerca i Tecnologia Agroalimentàries.
*US Meat Animal Research Center.

**Breeding Companies**

The worldwide pork industry obtains much of its genetic stock from companies that provide seedstock and develop lines that fit into specified breeding programmes (e.g. Babcock Genetics, Danbred, Fast Genetics, Genesis, Genetiporc, Hypor, Hermitage, JSR, Newsham Choice Genetics, PIC, TOPIGS). These companies develop their own lines, some of which are based on conventional purebreds, while others represent crosses of various breeds but have been kept closed for a sufficient period to be considered as independent from their foundation breeds. The companies generally sell a breeding programme with specialized sire and dam lines. The PIC Camborough gilt, which is a hybrid of Landrace and Large White, is very commonly used in commercial swine production throughout the
world. The genetic selection procedures and performance characteristics are generally kept private by the company, but there have been attempts by the industry to characterize the performance of some of the lines (Cassady et al., 2002; Moeller et al., 2004; Serenius et al., 2007). These company lines present a departure from conventional thought about what constitutes a breed. However, the procedures used by the companies to preserve and improve their lines are very much in keeping with a simple modernization of the techniques used by purebred swine producers for many decades. Breeding companies continue to sample pigs from conventional breeds to identify sources of new genetic material for their lines.

**Molecular Biology and Breeds of Pigs**

Modern molecular biology tools are now available to the swine industry. Genome mapping efforts have made great progress in the USA (NSRP8, 2010; Rothschild, 2010; University of Illinois, 2010), the UK (Roslin Institute, 2010; Sanger Institute, 2010), France (INRA, 2010) and elsewhere. Other chapters in this book will address these efforts. In brief, progress is being made in identifying genes to assist in genetic improvement and aid in maintaining breed purity (Milan et al., 1998; Bidanel et al., 2000, 2002, 2008; Peischl et al., 2005).

Molecular biology tools have led to greater understanding of domestication (Larson et al., 2007) and breed diversity (Franceschi and Ollivier, 1981; Laval et al., 2000; Okumura et al., 2001; Kim et al., 2002; Alves et al., 2003; Ollivier et al., 2003, 2005; Yang et al., 2003; Gongora et al., 2004; Fang and Andersson, 2006; Fang et al., 2006; SanCristobal et al., 2006; Thuy et al., 2006; Wu et al., 2007; Jiang et al., 2008; Chang et al., 2009; Ollivier, 2009). Data from the Roslin Institute Pig Biodiversity Project were used to construct a phylogenetic tree (Fig. 18.1). The tree includes one Asian breed (Meishan), and it clearly shows the considerable genetic distance between it and the breeds with origins in Europe. The presence of the Meishan in the middle of the tree is surprising in view of the separate domestication of pigs in Asia and Europe. However, it is also documented (Larson et al., 2007) that Asian pigs were brought to Europe, and that genes from those pigs were incorporated into pigs which contributed to breeds that developed in Europe. For example, the Berkshire breed has been shown to have some genetic elements that are similar to those in several Chinese breeds (Wu et al., 2007). As this phylogenetic tree is examined, it is important to remember that when the entities in the tree are breeds, interpretation is different from when the entities are species. Speciation frequently occurs by a single species subdividing into two or more species. A phylogenetic tree illustrates those speciation events. Breed development occurs through a process of crossing representatives of other genetic stocks and the periodic insertion of genes from other breeds. As a result, a phylogenetic tree of breeds represents degrees of genetic similarity for the markers used in the research to evaluate genetic distance. However, as breeds draw genetic input from a variety of sources, a phylogenetic tree that truly represents breed development would have multiple connections instead of single branch points.

**Summary**

The concept of a breed is likely to remain rather fluid. It is tempting to assume that the important breeds of today will continue to be important in the future. One has only to examine the history of breeds during the 20th century, in swine and in other species, to see the fallacy of this assumption. The popularity of the Berkshire, followed by a decline in numbers but a recent resurgence, is an example. Improved techniques for the identification of superior genetic material, including techniques from molecular biology, will be likely to speed the evolutionary pace in swine. This will mean even more rapid assembly and recombination of genetic stocks.
Fig. 18.1. Phylogenetic tree of breeds. Data adapted from the pig biodiversity database from the Roslin Institute (http://projects.roslin.ac.uk/pigbiodiv/). The phylogenetic tree was built using the program PHYLIP (Felsenstein, 2004).
References


Introduction

Genetics includes the study of genotypes, phenotypes and the mechanisms of genetic control between them. Genetic terms describe the processes, genes and traits with which genetic phenomena are described and examined. The genetic process terminologies are thoroughly discussed in the previous chapters. Therefore, in this chapter, we will only list the terms for genetic processes and concepts in Appendix I (a general genetic glossary), and concentrate the discussion on pig gene and trait terminologies (Appendix II); a glossary for pig diseases and defects is also included (Appendix III).

A standardized genetic nomenclature is vital for unambiguous concept description, efficient genetic data management and effective communications among not only scientists, but also those who are involved in the pig production and genetic improvement industries. This issue becomes even more evident in the post-genomics era, owing to the rapid accumulation of large quantities of genetic and phenotypic data, and the use of computer software to manage such data. This imposes a challenge for the precise definition and interpretation of gene and trait terms.

For example, Myostatin (gene MSTN) is also known as Growth and Differentiation Factor 8 (gene GDF8) (one can also find inappropriate abbreviations such as GDF-8 in the literature) and is referred to as the ‘muscle hypertrophy’ locus in cattle. While all these names are interchangeably used in the literature, it gets more complicated when one considers paralogous gene duplications across species, which led Rodgers et al. (2007) to propose MSTN-1 and MSTN-2 as names. Unfortunately, this naming scheme does not
follow Human Gene Nomenclature Committee (HGNC) guidelines, which would indicate that they should be named MSTN1 and MSTN2.

In terms of traits, an example that would benefit from consistent nomenclature is the longissimus dorsi muscle area, which is also referred to as the loin eye area (LEA), loin muscle area (LMA), meat area (MLD), ribeye area (REA), etc. Each of these is known by different individuals as the default name for the trait. Complexity is further increased by variation in anatomical locations, physiological stages and methods used to measure a given trait. This may seem manageable at first, but once one starts to compare data across different labs, publications or species, it quickly becomes very confusing.

While we may not want to dictate to the community how genetic terms are defined, there are good reasons why all researchers need to adopt standardized genetic nomenclatures. The emergence of the use of ontologies in biological research has contributed a new way to effectively use, standardize and manage genetics terms. The Gene Ontology (GO) consortium has provided a good example (The Gene Ontology Consortium, 2000). When genomics information must be transferred across species to perpetuate genetic discoveries, the role of a standardized genetic nomenclature becomes even more important.

The goal of this chapter is to help establish guidelines for nomenclature, with the hope that it will facilitate the comparison of results between experiments and, most importantly, prevent confusion.

**Locus and Gene Names and Symbols**

**Locus name and symbol**

These guidelines for gene nomenclature are adapted and abbreviated from the HGNC Guidelines [http://www.genenames.org/guidelines.html](http://www.genenames.org/guidelines.html).

A ‘gene’ is a functional hereditary unit that occupies a fixed location on a chromosome, has a specific influence on phenotype, and is capable of mutation to various allelic forms. In the absence of demonstrated function a gene may be characterized by sequence, transcription or homology. A ‘locus’, which is not synonymous with a gene, refers to a position in the genome that can be identified by a marker. A ‘chromosome region’ is defined as a genomic region that has been associated with a particular syndrome or phenotype.

Gene names and symbols will follow the human gene when 1:1 orthology is known. Gene names should be short and specific and convey the character or function of the gene. Gene names will be written using American spelling and contain only Latin letters or a combination of Latin letters and Arabic numerals. The first letter of a gene symbol should be the same as for the gene name. The symbol will consist of upper-case Latin letters and possibly Arabic numerals. Gene symbols must be unique.

The locus name should be in capitalized Latin letters or a combination of Latin letters and Arabic numerals. If the locus name is two or more words, each word after the first word should be in capital Latin characters. The locus symbol should consist of as few Latin letters as possible or a combination of Latin letters and Arabic numerals. The characters of a symbol should always be capital Latin characters, and should begin with the initial letter of the name of the locus. If the locus name is two or more words, then the initial letters should be used in the locus symbol. The locus name and symbol should be printed in italics wherever possible; otherwise they should be underlined.

When assigning gene nomenclature, the gene name and symbol should be assigned based on existing HGNC nomenclature where possible (e.g. 1:1 for swine:human orthologues). Ensembl (a joint project between EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) and the Wellcome Trust Sanger Institute to develop a software system that produces and maintains automatic annotation on selected eukaryotic genomes) has used the new EPO (Enredo, Pecan, Ortheus) pipeline (Paten et al., 2008) for whole genome alignment of the swine genome. Initial efforts to provide information about genes predicted during the swine genome sequencing effort assigned standardized nomenclature based on human gene nomenclature for 2798 swine genes [http://www.ensembl.org/Sus_scrofa/Info/StatsTable](http://www.ensembl.org/Sus_scrofa/Info/StatsTable).

There are two categories of novel swine genes: (i) novel genes predicted by bioinformatic
gene prediction programs; and (ii) novel swine genes that were studied before the completion of the swine genome. In cases where no strict 1:1 human orthologue exists that has been assigned nomenclature, the LOC# or Ensembl ID should be used as a temporary gene symbol. In order to assign a name to a novel gene, it will need to be manually curated and assigned a unique name, following HGNC guidelines.

Allele name and symbol

These guidelines for allele nomenclature are adapted from Young (1998) and mouse genome nomenclature guidelines (http://www.informatics.jax.org/mgihome/nomen/gene.shtml), in accordance with HGNC guidelines.

The allele names should be as brief as possible, yet still convey the variation associated with the allele. Alleles do not have to be named, but should be given symbols. If a new allele is similar to one that is already named, it should be named according to the breed, geographical location or population of origin. If new alleles are to be named for a recognized locus, they should conform to nomenclature established for that locus. The first letter of the allele name should be lower case. However, this does not apply when the allele is only a symbol.

An allele symbol should be as brief as possible and consist of Latin letters or a combination of Latin letters and Arabic numerals. Like a gene symbol, an allele symbol should be an abbreviation of the allele name, and should start with the same letter. The allele name and symbol may be identical for a locus detected by biochemical, serological or nucleotide methods. The wild-type allele can be denoted as a + (e.g. MSTN+). Neither + nor – symbols should be used in alleles detected by biochemical, serological or nucleotide methods. Null alleles should be designated by the number zero. The initial letter of the symbol of the top dominant allele should be capitalized. All alleles that are codominant should have an initial capital letter. The initial letter of all other alleles should be lower case. A single nucleotide polymorphism (SNP) allele should be designated based on its dbSNP_id (http://www.ncbi.nlm.nih.gov/projects/SNP/), followed by a hyphen and the specific nucleotide (e.g. MSTNrs1234567T). If the SNP occurs outside an identified gene, the SNP locus can be designated using the dbSNP_id as the locus symbol, and the nucleotide allelic variants are then superscripted as alleles (e.g. rs1234567T).

The allele symbol should always be written with the locus symbol. Specifically, the allele symbol is written as a superscript following the locus symbol. For example, an SNP allele can be designated based on its dbSNP_id, followed by a hyphen and the specific nucleotide, as in MSTNrs1234567T. The allele symbol should be printed immediately adjacent to the locus symbol, with no gaps. The allele name and symbol should be printed in italics whenever possible, or otherwise be underlined.

Genotype terminology

The genotype of an individual should be shown by printing the relevant locus and allele symbols for the two homologous chromosomes concerned, separated by a slash, e.g. MSTNrs1234567T/r1234567C. Unlinked loci should be separated by semicolons, e.g. CD11^Rod2400;2000, ESR^Pod5700;4200. Linked or syntenic loci should be separated by a space or dash and listed in linkage order (e.g. POU1F1AG-STCHF^G-PRSS7^T), or in alphabetical order if the linkage order is not known. For X-linked loci, the hemizygous case should have a /Y following the locus and allele symbol, e.g. APR^B8571094/Y. Likewise, Y-linked loci should be designated by /X following the locus and allele symbol.

Trait and Phenotype Terminology

Pig traits and trait ontology

Pig traits are conventionally named based on their performance (e.g. average daily gain), physiological parameters (e.g. ovulation rate), anatomical locations/dissections (e.g. loin muscle area), physicochemical properties (e.g. muscle protein percentage, pH), livelihood soundness (e.g. immune capacity) and exterior appearance (e.g. coat color), etc. As such, there is a good chance that a trait will be named differently by different people, even within a
species community. Furthermore, traits have been studied across many species, which adds additional complexity to their naming. The study of the traits may also involve the study of underlying genes and markers, environments and management protocols that contribute to the manifestation of a trait. Therefore, it is obvious that factors that contribute to the naming of a trait are multidimensional. As the amount of trait information that is associated with a gene or chromosomal region is growing exponentially, we cannot overemphasize the need for a standard nomenclature to be used by researchers to communicate as consistently and unambiguously as possible, with the aid of bioinformatics tools.

Traits

Pig trait terms can be found ubiquitously throughout journal articles, farm reports and daily communications among scientists and pig industry personnel. A trait term can be created by anyone, and each person may have a slightly different definition for said term, so hundreds of thousands of terms can be found in the literature with various naming conventions used. Previously, there was no central repository where the uniqueness of a trait term could be maintained and checked, but two relatively recent database development efforts have now emerged: the Online Mendelian Inheritance in Animals (OMIA) database and the Animal Trait Ontology (ATO) database.

OMIA was initiated in 1978. To date, it contains 214 pig trait variations and/or abnormalities from pig genetic research publications (OMIA, 2009). ATO has a collection of over 500 pig traits, including measurement method variations (ATO, 2009), of which 390 traits are associated with at least one quantitative trait locus (QTL). Appendices II and III contain a list of pig traits and a list of pig diseases and defects, respectively, which were mainly abridged from ATO and OMIA. In both cases, efforts have been made to make each database entry unique in terms of the names and their representations. ATO has been expanded to include trait data from cattle, pig and chicken, but was originally developed as part of the PigQTLdb (Pig QTL database, http://www.animalgenome.org/QTLdb/pig.html, later known as Animal QTLdb when QTL data from multiple animal species were included), and the number of pig traits collected is more complete than that of cattle or the chicken. The initial purpose of ATO was to help with organization and management of trait information through the use of a controlled vocabulary to facilitate comparison of QTL experiment results and standardized trait data annotation and retrieval (Hu et al., 2005, 2007), but it was later introduced to the community (Hughes et al., 2008).

‘Super-traits’

Compared with standard gene nomenclature, trait name standardization is far more complex, not only because the same trait can be named differently (e.g. ‘loin eye area’ versus ‘ribeye area’), but also because many factors contribute to how a trait is defined under various circumstances. For example, Table 19.1 gives a list of 16 ‘backfat’ variations, each defined by their different measurement methods, measuring time and specific anatomical locations, which may contribute to trait comparison difficulties and, sometimes, confusion.

One attempt to simplify the comparisons is by introducing the concept of ‘trait types’ or ‘super-traits’ (Hu et al., 2005). Hu et al. described trait type as a general physical or chemical property of, or the processes that lead to, or types of measurements that result in an observation (phenotype). The ‘trait types’ or ‘super-traits’ were initially used to serve as a general concept for a trait, regardless of possible variations in the trait names by measurement times, locations or methods. As the ATO project progressed, the factors in the methods of trait measurements, such as point in time or timespan, anatomical locations, instruments, etc., were classified as ‘trait modifiers’, because they do not constitute a component of a trait, but only affect the way a trait is described (authors’ unpublished data). Therefore, the ‘super-trait’ may only be employed to categorize variations in how a trait is defined or named. For example, ‘ribeye area’, ‘rib muscle area’, ‘longissimus dorsi muscle area’, ‘longissimus muscle area’, ‘loin eye area’, ‘loin muscle area’, etc. can be unified as ‘longissimus dorsi muscle area (LMA)’; and ‘backfat’, ‘backfat depth’, ‘backfat thickness’, ‘backfat above muscle dorsi’, ‘backfat
'Intercept', 'backfat linear', etc. can all simply be 'subcutaneous fat thickness'.

**Trait hierarchy and ontology**

In order to compare QTLs across experiments, the PigQTLdb uses a trait hierarchy (Table 19.2) to provide a framework for organizing the traits and easily locating them (Hu et al., 2005). This approach simplifies the procedures by which traits are defined, linked and compared. Subsequently, a computer program could be implemented to automatically process the database searches so that, when a user queries for a trait by keywords, the database can gather and retrieve related trait names and their associated QTLs, put them together and present them to the user in real time.

However, people of different disciplines may see the need for a different trait hierarchy, which may better capture the subtleties required in their field. For example, for the body weight gained over a period of time (e.g. average daily gain, or ADG, a measure of body weight gain), a farmer considers it a production trait, a nutritionist may see it as an indicator for feed conversion efficiency and a veterinarian may find it a health status parameter. Similarly, blood cholesterol levels may be used to predict pork quality by pork producers, and may also be a parameter to predict coronary heart disease by those who use the pig as an animal model for human heart disease research. Therefore, a simple hierarchy may be helpful to solve problems in some cases, but may be inadequate in...
other cases. In addition, owing to the existence of multiple overlapping hierarchies for pig traits, the management of such data may be difficult. Thankfully, progress is being made to resolve these issues.

Ontologies use controlled vocabularies to describe objects and the relationships between them in a formal manner. In an ontology, the Directed Acyclic Graph (DAG), a mathematical graphic modelling method, is used to solve data management problems with complex hierarchical structure. In the example shown in Fig. 19.1, the trait ‘marbling’ is shown to belong to the ‘meat quality’, ‘adipose trait’ and ‘muscular system physiology’ hierarchies. Figure 19.2 shows how these traits are linked by their respective hierarchy paths. Computer tools have been developed and are freely available to manage ontology data with DAG structures. The two most popular tools, which are likely to be useful to the pig genetics community, are AmiGO and OBO-Edit (Gene Ontology Tools, 2009). AmiGO is an ontology browser adapted to the ATO database, which allows users to share and view trait data stored in ATO with any web browser on the Internet (ATO, 2009). OBO-Edit is a Java-based ontology data editor, which can be used by anyone to edit ontology term definitions and relationships.

ATO has been a successful project since its development from the QTLdb several years ago. Recently, the developers of ATO have begun working with Mouse Genome Informatics, the Rat Genome Database, the European Animal Disease Genomics Network of Excellence (EADGENE) and the French National Institute for Agricultural Research (INRA) to incorporate Mammalian Phenotype (MP) ontology and ATO into a unified Vertebrate Trait (VT) Ontology (2009). This project is aimed at enhancing the ability to standardize trait nomenclature within and across species. For example, a disease such as Atrophic rhinitis may have been considered a ‘trait’ in classical animal genetic studies. In fact, in terms of concept specifications, it is not a characteristic pig trait observable in the general population, but rather a morphological abnormality caused by a disease only in some pigs. In addition, a trait name may have variations because it is ‘modified’ by measurement time or method (Table 19.1), but these are actually the same trait. The separation of diseases from traits...
technical issue, but a community issue, in the sense that it has to be commonly recognized, mutually agreed upon and widely shared.

**Trait nomenclature recommendations**

Currently, pig traits are named in many different ways, ranging from the very abbreviated to other names that are verbally lengthy and self-descriptive. We suggest a trait name be chosen to represent that which is measured on an individual animal. We have compiled and edited known pig traits into a list in Appendix II. We strongly recommend users to refer to the specialized online databases for the most up-to-date reference, as far as the use of trait terms is concerned. While it might not be realistic to enforce a naming guideline without a community consensus, we recommend the following steps to take or factors to consider when naming a trait:

1. Check against existing databases and literature, to make sure a trait term does not previously exist. Good places to start are online databases such as ATO, OMIA, AGRICOLA and PubMed. Checking journals in the field (e.g. *Journal of Animal Science*), and in closely related fields (e.g. *Meat Science*), is essential too. Sometimes, reviewing similar names also helps to improve ideas.

2. Try to be informative. Although more details can be given in the ‘description’ of the trait, the name itself should already carry most of the conceptual information of that trait.

3. Try to be brief. One or a few key nouns, or a short noun phrase, may be good candidates. Sentences should be avoided (i.e. no verb). Generally, the name should be short, but do not abbreviate so much as to lose necessary information.

4. The name should be neutral. It should represent a concept, not a type or a part of the concept.

5. Avoid adjectives, adverbs and excessive symbols in the name.

6. If similar trait terms exist in another species, consider adapting them for use in pigs. This will help with future cross-species data comparisons and in bringing comparative information from other species.

7. Consider a good abbreviation for the name. Making it phonically easy to pronounce and easy to remember will make it more likely to be well used.

8. The name should be in English, using the American spelling. The abbreviations should be in capital Latin letters. If an original trait term used in a study is not English, its English equivalent is needed for publications.

9. Lastly, check for possible name conflicts to avoid confusion.

---

**Fig. 19.2.** An OBO-Edit graphic view showing that multiple relationship paths may exist between a child term and its ancestral term. The symbol ‘I’ means that the relationship type is ‘*is_a*’; The symbol ‘P’ means that the relationship type is ‘*part_of*’.
Phenotype nomenclature recommendations

A phenotype is the actual manifestation of observable traits. It is a trait observed in an individual. It usually consists of a trait with characteristic features (e.g. a litter size of nine piglets), variations that can be described (e.g. black spots on the body) or qualities that can be measured (e.g. birth weight of 1.5 lb). Because there are so many variations in how a phenotype can be ‘observed’ (often such observation is made indirectly with instruments or through tests) and obtained, a technical guide for recording each trait might be ideal. None the less, we recommend that the following components are included for a complete phenotype record:

1. What is the observation for? This is normally the trait type or super-trait name, such as ADG or carcass weight.
2. How is the observation made? This refers to the methods used to measure or make the observation (e.g. balance, ruler, ultrasound machine, scores, etc.).
3. How are the observational data processed? This is the calculation necessary to derive the results for recording (e.g. calculation of ADG, conversion of ultrasound value to length).
4. What is the unit of the observed data (e.g. inch or centimetre for length, pound or gram for weight, percentage or ratio for fractions, etc.)?

In addition, a description of comments for a phenotype record may be necessary for the data users to correctly understand and use the data. For example, when blood samples are taken, the number of hours the pig is fasted might be an important cofactor for the blood cholesterol concentration measured.

When a phenotype is a reflection of certain genotypes, the phenotype symbol should be in the same characters as the genotype and allele symbols. The difference is that the characters should not be underlined or in italics, and they should be written with a space between locus characters and allele characters, instead of an asterisk. Square brackets, [], may also be used.

In classical genetics, phenotypes are often used to denote Mendelian genotypes. This is done using an abbreviation of the trait, followed by a superscripted plus (+) or minus (−) sign to represent ‘presence’ or ‘absence’ of certain trait features. For example, halothane-negative is denoted as ‘Hal−’ and halothane-positive as ‘Hal+’. A phenotype denotation can also be used to represent genetic haplotypes, such that ‘K88ab+, ac−, ad−’ are written together as an entire denotation. Likewise, numbers or letters may be used to denote alleles when polymorphisms are observed, for example, ApoB1/2, ApoB2/3, etc. (note the difference from recording genotypes, where italics or asterisks are required).

Future Prospects

The Gene Ontology and Mammalian Phenotype Ontology are already playing a role in robust annotation of mammalian genes and phenotypes in the context of mutations, quantitative trait loci, etc. (Smith et al., 2005). Undoubtedly, a standardized pig genetic nomenclature will more effectively facilitate efficient pig genome annotation and transfer of knowledge from information-rich species such as humans and the mouse, and make it possible for new bioinformatics tools to easily streamline data management and genetic analysis.

Several genome databases, such as ArkDB (http://www.thearkdb.org/), Animal QTLdb (http://www.animalgenome.org/QTLdb), Ensembl (http://www.ensembl.org/) and NCBI GeneDB (http://www.ncbi.nlm.nih.gov/gene), have played a role in the usage of commonly accepted gene/trait notations. Undoubtedly, existing and new genome databases and tools will further develop and evolve. As such, a standardized genetic nomenclature in pigs will definitely become crucial for information sharing and comparisons between different research groups, across experiments and even across species.

Acknowledgements

The authors wish to thank Drs Max Rothschild and John Bastaanssen for their useful discussions in the early stages of PigQTLdb development.
References


Appendix I  Genetic Glossary

**Bold words** are glossary entries. **Italicized words** are concepts that may be independent glossary entries as well.

**Adaptation traits** – Adaptation traits contribute to individual fitness and to the evolution of animal genetic resources. By definition, these traits are also important to the ability of the animal genetic resource to be sustained in the production environment.

**Additive genetic effects** – The effect of an allele on animal performance, independent of the effect of the other allele at a locus. These effects of the two alleles at a locus add up (thus ‘additive’). Alleles at a locus may have other effects (dominance, epistasis), so that there are not genes that have just ‘additive’ effects and other genes with only ‘dominance’ effects. Additive genetic effects can be inherited; other genetic effects such as dominance and epistasis are the result of allele combinations that are lost between generations. The additive genetic effect that an animal has for a trait is equal to its breeding value.

**Allele** – One of a pair, or series of alternative forms, of a gene that can occur at a given locus on homologous chromosomes.

**Amino acids** – Any one of a class of organic compounds containing the amino (NH₂) group and the carboxyl (COOH) group. Amino acids are combined to form proteins.

**Ancestor** – Any individual from which an animal is descended.

**Animal model** – A system for genetic evaluations that estimates breeding values of individual animals (males, females) at the same time. The system uses production data on all known relatives in calculating a genetic evaluation.

**Assortative mating** – Assigning animals as mates based on phenotypic or genetic likeness. **Positive assortative mating** is mating animals that are more similar than average. **Negative assortative mating** is mating animals that are less similar than average.

**Autosome** – Any chromosome that is not a sex chromosome.

**Backcross** – The cross produced by mating a first-cross animal back to one of its parent lines or breeds.

**Best linear unbiased prediction (BLUP)** – A method of genetic prediction (with the properties of smallest variance, linear and unbiasedness) that is particularly appropriate when performance data come from genetically diverse contemporary groups.

**Breed** – Either a subspecific group of domestic livestock with definable and identifiable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species, or a group for which geographical and/or cultural separation from phenotypically similar groups has led to acceptance of its separate identity.

**Breeding value** – The mean genetic value of an individual as a parent. It can be estimated as the average superiority of an individual’s progeny relative to all other progeny under conditions of random mating.

**Categorical trait** – Scores are given usually in a few categories up to several categories (e.g. scores 1–5 for leg movement).

**Centromere** – Spindle-fibre attachment region of a chromosome.

**Chromosome** – Microscopically observable linear arrangement of DNA in the nucleus of a cell. Chromosomes carry the genes responsible for the determination and transmission of hereditary characteristics.

**Codominant alleles** – Alleles, each of which produces an independent effect in heterozygotes.

**Combining ability** – The mean performance of a line when involved in a crossbreeding system. General combining ability is the average performance when a breed or line is crossed with two or more other breeds or lines. Specific combining ability is the degree to which the performance of a specific cross deviates from the average general combining ability of two lines.

**Composite (synthetic) breed** – A hybrid with at least two, and typically more, breeds in its background. Composites are expected to be bred to their own kind, retaining a level of hybrid vigour normally associated with traditional crossbreeding systems.

**Control line** – A line that is randomly selected and randomly mated. Usually used in selection experiments to monitor environment effects in order to estimate genetic change in a selected line.

**Correlation coefficient** – A measure of the interdependence of two random variables that ranges in value from −1 to +1, indicating perfect negative correlation at −1, absence of correlation at zero, and perfect positive correlation at +1. It determines the degree to which the movements of two variables are associated. No cause and effect is implied.
**Covariance** – The degree to which two measurements vary together. A positive covariance is when two measurements tend to increase together. A negative covariance is when one measurement increases and the other measurement tends to decrease.

**Crossbreeding** – Matings between animals of different breeds or lines.

**Crossover** – The process during meiosis when chromosomal segments from different members of a *homo-logous* pair of chromosomes break, and part of one will join a part of the other, so that two gametes form possessing new combinations of *genes*. The frequency of crossover between two *loci* is proportional to the physical distance between them.

**Crossover unit** – Each unit is equal to a 1% frequency of crossover *gametes*.

**Cytoplasm** – The protoplasm outside a cell *nucleus*.

**Descendant** – An individual descended from other individuals.

**Diallel cross** – When both males and females from each breed (or line) in a set of breeds (or lines) are mated to males and females of each breed (or line) in the set including their own breed (or line).

**DNA** – Deoxyribonucleic acid, the chemical material that carries information to code for a gene.

**Dominant** – Applied to one member of an allelic pair of genes, which has the ability to express itself wholly or largely at the exclusion of the expression of the other allele.

**Dominance genetic effects** – The effect that an allele has on animal performance, which depends upon the genotype at the locus. For example, the ‘a’ allele may have a different effect on animal performance in ‘aa’ animals than in ‘Aa’ animals. See *additive genetic effects*.

**Economic trait loci** – Loci that have effects on traits of economic importance.

**Economic value** – A measure of the contribution an individual trait makes to the overall economic value of an animal.

**Environment** – The aggregate of all the external conditions and influences affecting the life and development of the organism.

**Environmental correlation** – When two traits tend to change in association with each other as a result of environmental effects.

**Environmental variance** – Variation in phenotype that results from variation in environmental effects.

**Epistasis** – When the gene at one locus affects the expression of the gene at another locus.

**Estimated breeding value** – A prediction of a breeding value. See *breeding value*.

**F1** – Animals resulting from crossing parents from different lines or breeds.

**F2** – Animals resulting from matings among *F1* parents.

**F3** – Animals resulting from matings among *F2* parents.

**Family size** – The mean number of offspring per parent that successfully reproduce.

**Full sibs** – Individuals having the same male and female parents.

**Gamete** – A sperm or egg cell containing the haploid (1n) number of chromosomes.

**Gene** – A functional hereditary unit that occupies a fixed location on a chromosome, has a specific influence on phenotype and is capable of mutation to various allelic forms.

**Generation interval** – The average age of the parents when the progeny that will replace them are born.

**Genetic correlation** – When two traits tend to change in the same or opposite directions as a result of genetic effects.

**Genetic distances** – A measure of gene differences between populations (hence genetic relationships among them) described by some numerical quantity; the gene differences are usually referred to as measured by a function of gene frequencies.

**Genetic drift** – Changes in gene frequency in small breeding populations due to chance fluctuations.

**Genetic gain** – The amount of increase in performance that is achieved through genetic selection after one generation of selection.

**Genetic maps** – See *linkage map*.

**Genetic marker** – A gene or DNA sequence having a known location on a chromosome and associated with a particular gene or trait; a gene phenotypically associated with a particular, easily identified trait and used to identify an individual or cell carrying that gene.

**Genetic variance** – Variation in phenotype that results from variations in genetic composition among individuals.

**Genome** – The complete set of *genes* and *non-coding* sequences present in each cell of an organism, or the genes in a complete haploid set of *chromosomes* of a particular organism.

**Genotype** – The genetic constitution of one or a few *gene(s)* or *locus (loci)*, or total genetic make-up (genes) of an individual organism.
Genotype–environment interaction – When the difference in performance between two genotypes differs, depending upon the environment in which performance is measured. This may be a change in the magnitude of the difference or a change in rank of the genotypes.

Germplasm – The germinal material or physical basis of heredity; the sum total of the genes.

Grade-up – The process of repeated backcrossing to one parental line to produce a population that is nearly purebred.

Half sibs – Individuals that share only one common parent.

Haplotype – A set of alleles at a closely linked group of loci, so closely linked that the allelic set behaves almost as one allele in terms of inheritance.

Hardy–Weinberg law – A population is in genotypic equilibrium if \( p \) and \( q \) are the frequencies of alleles \( A \) and \( a \), respectively, and \( p^2, 2pq \) and \( q^2 \) are the genotypic frequencies of \( AA \), \( Aa \) and \( aa \) under the condition of random mating.

Heritability – Degree to which a given trait is controlled by inheritance; proportion of total phenotypic variation that is attributable to genetic variation (in contrast to environment-caused variation).

Heterosis – The degree to which the performance of a crossbred animal is better or worse than the average performance of the parents.

Heterozygote, adj. heterozygous – An organism with unlike members of any given pair or series of alleles, which consequently produces unlike gametes.

Homologous chromosomes – Chromosomes that occur in pairs and are similar in size and shape, one having come from the male and one from the female parent.

Homozygote, adj. homozygous – An organism whose chromosomes carry identical members of a given pair of genes. The gametes are therefore all alike with respect to this locus.

Inbreeding – Matings among related individuals, which result in progeny that have less heterozygosity and hence more homozygous gene pairs than the average of the population.

Inbreeding coefficient – A measurement of the increase in homozygosity; each unit is equal to a 1% increase in homozygosity relative to the average homozygosity in the base population.

Inbreeding depression – The decreased performance normally associated with accumulation of inbreeding. Many recessive genes result in undesired traits or decreased performance when they are expressed. Inbred animals have more recessive genes in the homozygous condition that are expressed and result in reduced performance or undesired traits.

Independent culling – When animals are culled if they do not meet all of the minimum levels of performance for a set of traits.

Introgression – A breeding strategy for transferring specific favourable alleles from a donor population to a recipient population. This would, for example, be of great interest for genes responsible for disease resistance, which could be introgressed into a susceptible, but otherwise economically superior, breed.

Karyotype – The appearance of the metaphase chromosomes of an individual or species, which shows the comparative size, shape and morphology of the different chromosomes.

Lethal gene – A gene that results in the death of the animal.

Liability – Both internal (e.g. genetic merit) and external (e.g. nutrition, disease, exposure) forces that influence the expression of a threshold character (e.g. disease, conception, abnormalities, etc.).

Line-breeding – Mating of selected individuals from successive generations to produce animals with a high relationship to one or more selected ancestors. It is a mild form of inbreeding.

Linkage – Association of genes physically located on the same chromosome. A group of linked genes is called a linkage group.

Linkage map – A linear map of an experimental population that shows the position of its known genes and/or genetic markers relative to each other in terms of recombination frequency.

Locus, pl. loci – A fixed position on a chromosome occupied by a given gene or one of its alleles.

Major gene – A gene that has an easily recognizable and measurable effect on a characteristic.

Marker – Specific and identifiable sequences of the DNA molecule. These markers may or may not be functional genes.

Marker-assisted selection (MAS) – Selection for specific alleles using genetic markers.

Maternal heterosis – The advantage of the crossbred mother over the average of purebred mothers.

Mating systems – The rules that describe how selected breeds and/or individuals will be paired at mating.

Meiosis – The process by which the chromosome number of a reproductive cell becomes reduced to half the diploid \((2n)\) or somatic number and results in the formation of eggs or sperm.
Migration – Movement of animals, and consequently genes, from one population to another.

Mitochondria – Small bodies in the cytoplasm of most plant and animal cells that are responsible for energy production.

Mitosis – Cell division process in which there is first a duplication of chromosomes, followed by migration of chromosomes to the ends of the spindle and a dividing of the cytoplasm, resulting in the formation of two cells with a diploid (2n) number of chromosomes.

Molecular genetics – The branch of genetic studies that deals with hereditary transmission and variation on the molecular level. It deals with the expression of genes by studying the DNA sequences of chromosomes.

Multiple alleles – Three or more alternative forms of a gene representing the same locus in a given pair of chromosomes.

Mutation – A sudden change in the genotype of an organism. The term is most often used in reference to point mutations (changes in base sequence within a gene), but can refer to chromosomal changes.

Natural selection – Natural processes favouring reproduction by individuals that are better adapted, and tending to eliminate those less adapted to their environment.

Nucleus – Part of a cell containing chromosomes and surrounded by cytoplasm.

Outcrossing – Mating of individuals that are less closely related than the average of the population.

Overdominance – A form of dominance where the performance of the heterozygote exceeds that of the best homozygote.

Partial dominance – A form of dominance where the performance of the heterozygote is intermediate between the two homozygotes, but more closely resembles the performance of the homozygous dominant type.

Pedigree – Usually refers to a pedigree chart or what a pedigree chart represents in genetics. It is a document to record the ancestry of an individual. A pedigree can also be used to illustrate the family structure or breeding scheme.

Penetrance – The proportion of the individuals with a particular gene combination that express the corresponding trait.

Permanent environmental effects – Environmental effects that result in permanent effects on the phenotypic expression of a trait. For example, severe mastitis during lactation may have a permanent effect on milk production and litter weaning weight for an animal in subsequent litters.

Phenotype – Actual exhibit of observable traits. Normally, it refers to characteristics of an individual such as size, shape, colour or performance.

Phenotypic correlation – When two traits tend to change in the same or different direction as a net result of genetic and environmental effects.

Phenotypic value – A performance record; a measure of an animal's performance for a trait.

Phenotypic variation – Variation in phenotype that results from variation in genetic and environmental effects on the individuals.

Pleiotropy – The property of a gene whereby it affects two or more characters, so that, if the gene is segregating, it causes simultaneous variation in the characters it affects.

Polymorphism – Where DNA or genes have more than two forms or alleles in the population.

Population – Entire group of organisms of a kind that interbreed.

Population genetics – The branch of genetics that deals with frequencies of alleles in groups of individuals.

Progeny – Offspring or individuals resulting from specific matings.

Progeny test – A test used to help predict an individual's breeding values, involving multiple matings of that individual and evaluation of its offspring.

Protein – Any of a group of complex nitrogenous organic compounds that contain amino acids as their basic structural units, occur in all living matter and are essential for the growth and repair of animal tissue.

Qualitative trait – A trait that can generally be classified into a limited number of categories, and the animal can be said to 'possess' the quality or not. Examples include hair colour, skin colour and ear stature.

Quantitative trait – A trait that is represented by an almost continuous distribution of measurements. Examples include average daily gain, backfat thickness and height.

Quantitative trait locus (QTL) – A locus that affects a quantitative trait.

Random mating – A mating system in which animals are assigned as breeding pairs at random, without regard to genetic relationship or performance.

Recessive – Applies to one member of an allelic pair that lacks the ability to manifest itself when the other, dominant member is present.
**Reciprocal cross** – A breeding scheme where males of breed A are mated to females of breed B and males of breed B are mated to females of breed A.

**Reciprocal recurrent selection** – A method of selection for combining ability or heterosis. Selection within two lines is based on the performance of crossbred progeny produced by crossing the two lines.

**Recombination** – The observed new combinations of DNA segments, or loci, or traits, which are different from those combinations exhibited by the parents.

**Recurrent selection** – A method of selection for combining ability or heterosis. Selection within one line is based on performance of crossbred progeny from matings with a ‘tester’ line.

**Repeatability** – The proportion of total phenotypic variation that is attributable to variations caused by genetic and permanent environmental effects. It is a measure of the degree to which early measures of a trait can predict later records of the same trait.

**RNA** – Ribonucleic acid, involved in the transcription of genetic information from DNA.

**Segregation** – The separation of paired alleles at loci during germ cell formation.

**Selection** – Any natural or artificial process favouring the survival and propagation of certain individuals in a population.

**Selection criteria** – The character(s) upon which selection decisions are based, with the intent of changing the character(s) in the selection objective.

**Selection differential** – The difference in mean performance of the selected group of animals relative to the mean performance of all animals available for selection.

**Selection index** – The combining of measurements from several sources into an estimate of genetic value; when more than one measurement on a trait, and/or measurements of the trait on relatives, and/or measurements of more than one trait are combined into a single estimate of overall genetic value.

**Selection intensity** – The proportion of animals selected to be parents relative to the total number available for selection. The smaller the proportion selected, the higher the selection intensity.

**Selection objective** – The character(s) that are intended to be modified by selection.

**Sex chromosomes** – The X or Y chromosomes.

**Sex-influenced** – Traits for which the expression depends on the sex of the individual.

**Sex-limited** – A trait that can be expressed only in one sex, such as milk production.

**Sex-linked** – Genes that are located on the sex (X or Y) chromosomes.

**Synthetic breed** – See composite breed.

**Zygote** – The cell produced by the union of mature gametes (egg and sperm) in reproduction.
Appendix II  Pig Trait Glossary

The pig traits are mainly abridged from the Animal Trait Ontology (ATO) database (http://www.animalgenome.org/bioinfo/projects/ATO/).

Abdominal fat (ABDF) – Abdominal fat measured by weight and/or converted to percentage.

ACTH basal level (ACTH1) – Level of adrenocorticotrophic hormone produced by the anterior lobe of the pituitary gland that stimulates the secretion of cortisone and other hormones by the adrenal cortex.

Age (AGE) – Age at certain physiological or production stage. Often measured at times of farrowing (AGELF), puberty (AGEP), slaughter (ENDAGE) or other times.

Androstenone, laboratory (ANDR) – Laboratory measurement of androstenone concentration.

Anti-K88 E. coli IgG level (K88Ab) – Laboratory detected or quantified anti-K88 antibody level (IgG) to *Escherichia coli*. Also denoted as IgG levels to K88 antigen (IGG2K88).

Anti-O149 E. coli IgG level (O149AbR) – Laboratory detected or quantified anti-O149 antibody level (IgG). Also denoted as IgG levels to E. coli O149 (IGG2O149).

Aspartate amino transferase activity (AST) – Laboratory measurement of AST enzymatic activity.

Average chain length (FA-ACL) – Average number of carbon atoms comprising the fatty acids in a fat sample.

Average daily feed intake (FEEDIN) – Daily feed intake by a pig averaged over a period of time.

Average daily gain (ADG) – Body weight gain measured during various time points, e.g. birth to weaning, weaning to marketing (also called fattening ADG), or specific test periods, such as between 0 and 3, 4, 5, 7, 10, 18 or 26 weeks of age, etc.

Average glycolytic potential (GLYPO) – An estimator of resting glycogen content measured *in vivo* on porcine muscle.

Average lactate (LACT) – Laboratory measurement of lactate concentration in muscle.

Backfat thickness (BF) – Thickness of the layer of fat that lies under the skin along the back of the pig. The measurement may vary depending on the method (e.g. Fat-O-Meter, ultrasound, ruler), anatomical location for taking the measure, the age and body weight (both represent physiological stage of the animal), etc. Also called fatback, subcutaneous fat thickness (SCFT).

Backfat weight (BFW) – Backfat amount, measured by dissected weight. May also be converted to a percentage (BFP).

Band-formed neutrophil number (BFNEUT) – Number of immature neutrophils undergoing granulopoiesis. Nuclei are unsegmented.

Basal cortisol level (CORT) – Blood cortisol concentration in the absence of external intervention.

Basal glucose level (GLU) – Blood glucose concentration in the absence of external intervention.

Base excess (BASEEX) – Refers to the amount of acid required to return the blood pH of an individual to the reference interval (pH 7.35–7.45), with the amount of carbon dioxide held at a standard value.

Basophil number (BASO) – Number of basophilic granulocytes.

Belly fat area (BFA) – The belly fat size as measured by area.

Belly meat content (BYLEAN) – Meat content measured by weight and/or converted to percentage (ratio).

Belly weight (BELLYWT) – Weight of belly; an indirect indication of fat deposition.

Bilirubin level (BILIR) – Level of this waste product formed from old red blood cells; used as an indicator of carcass quality.

Blood pH (BLPH) – pH measured from blood samples.

Body length (BODYL) – Measured from the tail head to the point of the shoulder when the head is down.

Body mass index (BMI) – A number calculated from weight and height, as an indicator of body conformation.

Body weight (BW) – Live body weight measured at various time points, for example, at birth (birth weight, BW), at weaning (weaning weight, WWT), at slaughter (slaughter weight, SWT), etc.

Bone weight (BONENWT) – Total bone weight in carcass. May also be expressed as a percentage (BP).

Calcium level (BCAL) – Laboratory measurement of blood calcium concentration.

Carbon dioxide level (CO2) – Laboratory measurement of blood carbon dioxide level.

Carcass fat-free weight (CWNOFAT) – Carcass weight minus weight of fat. Also referred to as carcass lean weight.

Carcass length (CRCL) – Length of carcass, measured post-mortem.

Carcass temperature (CARTEMP) – Usually measured at 45 min or 24 h post-mortem.

Carcass traits – Criteria used in assessing the quality of a carcass. Important in determining the price, suitability of breeding programme and value of sire. This may include length, weight, proportion of fat and lean, distribution of fat, relative size of valuable cuts, etc.
Carcass weight (CWT or HCWT) – Carcass weight measured either right after slaughter (HCWT) or 24 h after the slaughter (CWT).

Carcass width, minimum (CRCW) – Width of the carcass measured at the narrowest point.

Carcass yield (CYIELD) – Proportion of the pig's live weight salvaged at the point the carcass is harvested. Also called dressed weight, dressing percentage, carcass dressing or killing out percentage.

Chew score (CHEW) – A subjective score for chewiness as judged by a human on a taste panel. Also called chewiness score (CHEW).

Cholesterol (CHOL) – Blood cholesterol concentration, usually in mg/dl.

Chroma (CHROMA) – A measure of meat colour intensity, calculated based on a/a* and b/b* colour coordinates.

CO₂ partial pressure (PCO₂) – A blood parameter to describe haemoglobin oxygen-binding capacity as an indicator of health status. (A rise in the partial pressure of CO₂ or a lower pH will cause offloading of oxygen from haemoglobin.)

Coat color (CC) – Colour of the coat/hair.

Color density (COLORD) – Total colour density derived from the carcass image analysis.

ConA-induced cell proliferation (ConA) – Measurement of lymphocyte proliferation upon stimulation with concanavalin A. An indicator of immune capacity.

Conductivity post-mortem (COND) – Conductivity in muscle, typically measured at 45 min or 24 h post-mortem.

Conformation – Visual or measurable variation in body shape or proportions. Often assessed with a subjective scoring system.

Congenital – A condition present at birth but not necessarily hereditary.

Cooking loss (COOKL) – Loss by weight during the cooking process.

Cooling loss (CLOSS) – Loss by weight during the post-mortem carcass cooling process.

Cortisol level variations (dCORT) – Differences in blood cortisol levels. It is an indirect indicator of stress response.

Creatine kinase level (CK) – Measurement of creatine kinase concentration, typically as a stress response parameter. May also be referred to as creatine phosphokinase (CPK).

Creatinine level (CREAT) – Creatinine concentration, used as an alternative screening test for an abnormal hyperpyrexia response to halothane.

Diameter of muscle fibers – Average diameter of myofibres. Can include all fibre types, or specifically angular, giant, type I, type IIA, or type IIB (white; WFIBD) muscle fibres.

Diaphragm weight (DIAWT) – Dissected weight of the diaphragm.

Double-bond index (FA-DBI) – An assessment of the number of double bonds in a fat sample. Serves as an indicator of fat quality.

Dressing percentage (DRESSP) – See carcass yield.

Drip loss (DRIPL) – Fluid loss from fresh meat due to passive exudation. Also known as purge loss (PURGEL).

Ear erectness (EARER) – Determination of whether an animal's ears flop down, stand straight up or are intermediate; usually scored.

Ear size (EARSZ) – Determination of the outer ear size.

Eosinophil number (EOS) – Number of one of the types of immune cells responsible for combating infection and parasites.

Epididymis weight (EPIDW) – Weight of the epididymis by dissection.

Estimated carcass lean content (ECLC) – Estimated lean meat content of the carcass.

Estrone, laboratory (ESTR) – Laboratory determination of oestrone concentration.

Exploration during stress (EXPL) – Time spent rooting, sniffing the floor or the sides of the test arena during test period.

External fat (EFAT) – External fat measured on ham, loin or shoulder (shoulder fat thickness, subcutaneous fat depth at shoulder).

Fat area (FATAREA) – Fat area on musculus lattissimus dorsi.

Fat content (FCON) – Fat content of meat, determined based on laboratory methods. May also be referred to as total lipid (TOTLIP).

Fat content, total – Fat content of the carcass by weight (FATWT), by percentage (FPIC) or by ratio to lean (FP). May also be referred to as fat weight (FATWT) or total body fat tissue (FATTIS).

Fat-cuts percentage (FATCP) – Dissectible fat cut percentage.

Feed conversion ratio (FCR) – A measure of an animal’s efficiency in converting feed mass into increased body mass. Also called feed conversion rate, feed conversion efficiency (FCE).

Feed intake (FEEDIN) – Feed consumption, usually over a defined period of time.

Feed intake per feeding – Feed consumed during each visit to the feeder, usually measured during the fattening period.

Feet and legs score (FLSCORE) – A subjective conformation score of feet and legs, often during movement.

Feet score, front (FSCOREF) – A subjective front feet conformation score.

Feet score, rear (FSCORER) – A subjective rear feet conformation score.

Feet weight (FEETWT) – Weight of the feet.

Fiber type ratio – Ratio of myofibre types, based on myosin isoform ratio, for type 2 (FIB2R) or type 1 (FIBER1) fibres.

Firmness (FIRM) – Subjective firmness score; an indicator of meat texture.

Flavor score (FLAV) – Subjective flavour score, as judged by a human on a taste panel.

Foreloin weight (LWT) – Weight of the foreloin.

Glucose level variations (dGLU) – Changes in blood glucose concentration.

Glucose-6-phosphate content (GLU6P) – Laboratory measurement of glucose-6-phosphate.

Half carcass weight (HALFCW) – Weight of a half carcass.

Ham fat (HFAT) – Fat in ham, expressed as weight (HFATW), thickness (HFT) or percentage (HFATP).

Ham meat (HMEAT) – Meat in ham, measured by weight or expressed as a percentage.

Heart weight (HWT) – Weight of the heart.

Height (HEIGHT) – A general representation of body height measured at the shoulder.

Hematocrit (HCT) – Proportion of blood volume that is occupied by red blood cells. Also called packed cell volume (PCV).

Hemoglobin (HGB) – Laboratory measurement of blood haemoglobin concentration.

Humerus length (HUML) – Forelimb bone length measured on dissected carcass.

IgG levels to K88 antigen (IgG2K88) – See Anti-K88 E. coli IgG level.

Indole, laboratory (INDO) – Laboratory determination of indole concentration.

Internal fat ratio (IFR) – Fat deposited around internal organs, expressed as a percentage.

Intestinal fat – Intestinal fat by weight (IFW), or expressed as a percentage (IFP).

Intramuscular fat content (IMF) – Amount of fat present within muscle. Usually expressed as a percentage of total muscle weight.

Jowl weight (JOWLWT) – The weight of the cheek meat of a hog.

Juiciness score (JUICE) – A subjective sensory score, as judged by a human on a taste panel.

Knuckle ham weight (KHAM) – Weight of vastus intermedius, vastus lateralis, tensor fasciae and vastus medialis muscles.

Lactate level (BLACT) – Laboratory measurement of lactate concentration in the blood.

Leaf fat – Leaf fat, expressed as weight (LFW) or percentage (LFP).

Lean mass (LEANWT) – Amount of lean meat, measured by weight.

Lean thickness (RIBLT) – Meat depth measured between the 3rd and 4th or at the last rib.

Leg angularity (LANG) – Leg angularity relative to the vertical; on front (LANGF) or hind (LANGH) legs.

Leg length (LEGLEN) – Length of the legs.

Leg pastern (LPAST) – Leg scores as part of body conformation assessment; for front (LPASTF) or hind (LPASTH) pasterns.

Leg score (LSCORE) – Leg scores as part of body conformation assessment; for front (LSCOREF) or hind (LSCOREH) legs.

Leg set view (LSETV) – An assessment of leg placement in relation to the body.

Leg torsion (LTORS) – Degree of twisting or torsion of the legs; on front (LTORSF) or hind (LTORSH) legs.

Length of small intestine (LSI) – Length of the small intestine. This usually reflects the digestive capacity of an individual.
Lipid accretion rate (LAR) – Rate at which lipids are deposited over time.

Litter size – Number of offspring in a litter. Also referred to as total number born (TNB).

Liver weight (LIVWT) – Weight of the liver.

Locomotion (LOC) – A behavioural trait, recording the number of test arena sections entered during a stress test.

Loin eye area (LEA) – See longissimus dorsi muscle area.

Loin meat weight (LOINMWT) – Weight of the dissected loin with fat removed.

Loin muscle depth (LMDEP) – Depth or height of the loin; usually measured by ultrasound, at the last rib.

Loin muscle width (LMWID) – Loin width; usually measured by ultrasound, at the last rib.

Loin weight (LOINWT) – Total weight of dissected loin. May also be expressed as a percentage of carcass weight (LOINP).

Longissimus dorsi muscle area (LMA) – Cross-sectional area of the longissimus dorsi muscle. Also referred to as loin muscle area, loin eye area (LEA), ribeye area (REA).

Longissimus muscle length (LOINLENGTH) – Length of the longissimus dorsi muscle.

Lymphocyte number (LYMPH) – Number of a type of white blood cell. Used as a measure of immune capacity.

Marbling (MARB) – Visible intramuscular fat in a cut of meat, which gives it an appearance similar to a marble pattern. Degree of marbling is a meat quality trait.

Mean corpuscular volume (MCV) – A measure of the average red blood cell volume.

Mean platelet volume (MPVOL) – Average platelet volume, as measured with capillary tubes.

Meat color-a (MCOLOR-a) – Meat redness, measured by Minolta, Hunter or CIE coordinate.

Meat color-b (MCOLOR-b) – Meat yellowness, measured by Minolta, Hunter or CIE coordinate.

Meat color-L (MCOLOR-L) – Meat lightness, measured by Minolta, Hunter or CIE coordinate.

Meat color score (MCOLOR) – A subjective score for observed meat colour.

Meat-to-fat ratio (MFR) – Ratio between dissected meat and fat, by weight.

Melanoma susceptibility (MELAN) – Susceptibility to melanoma, a dark-pigmented, usually malignant tumour arising from a melanocyte; occurs most commonly in the skin.

Melting point (MP) – A characteristic feature of fat, serves as an indication of fat quality in terms of the amount of saturated versus unsaturated fat.

Muscle fiber type percentage (AFIBP) – Percentage of various fibre types in a muscle cross-section; includes angular, giant, Type I, Type Ila, Type IIB muscle fibres.

Neck weight (NECKWT) – Weight of the dissected neck.

Nipple number (NN) – See teat number.

Nonfunctional nipples (NNIP) – Nipples through which milk cannot be drawn. May also be referred to as pin nipples or inverted nipples.

Nose score (NOSESC) – Nose conformation score.

Number of stillborn (NSB) – Total number of stillborn in a farrowing.

Off-flavor score (OFFFLAV) – A sensory flavour score used to describe the intensity of any flavour not associated with normal pork flavour. Also called subjective abnormal flavor score, in fat (AFLFAT) or lean (AFLLEAN).

Ovary weight (OW) – Weight of the ovary.

Ovulation rate (OVRATE) – Frequency of ovulation, determined by the number of corpora lutea in the ovary.

Parasite load (PARAS) – Number of parasites present in a sample. Also called parasite burden.

pH for [tissue] (pH) – Muscle pH, where [tissue] can be biceps femoris, longissimus dorsi, ham, semimembranosus, semispinalis capitis, semispinalis dorsi, etc. Measured at different time points, e.g. 45 min, 24 h or 48 h post-mortem.

Plasma FSH concentration (FSH) – Follicle-stimulating hormone (FSH) concentration in blood plasma.

Platelet count (PLTCT) – Blood platelet counts.

Pork flavor (PFLFAT) – Subjective score used in a taste panel to describe the intensity of characteristic pork flavour.

Production traits – Characteristics of animals, such as the quantity or quality of the milk, meat, fibre, eggs, draught, etc. they (or their progeny) produce, which contribute directly to the value of the animals for the farmer, and that are identifiable or measurable at the individual level. Production traits of farm animals are generally quantitatively inherited, i.e. they are influenced by many genes whose expression in a particular animal also reflects environmental influences.
Protein content (PC) – Amount of protein in muscle, usually determined by chemical methods; measured by weight.

Purge loss (PURGEL) – See drip loss.

Relative area of muscle fibers (RAFIB) – Average muscle fibre area, includes Type I, Type Ila and Type Iib muscle fibres.

Resistance to *E. coli* K88 infection (K88R) – Ability of a pig to exhibit resistance to K88 *E. coli* infection.

Resistance to pseudorabies (PrV) – Ability of a pig to exhibit resistance to pseudorabies virus.

Rib and vertebral number – Number of ribs and vertebrae.

Ribeye area – See longissimus dorsi muscle area.

Rib weight (RIBWT) – Weight of all ribs from a carcass.

Segmented neutrophil number (SNEUT) – Number of fully mature neutrophils.

Shear firmness (SHFRM) – A texture measurement usually tested with shear force.

Shear force (SHEAR) – An instrument test for meat resistance during shearing; an indicator of tenderness.

Shear force at first peak (SF1P) – The shear force measured at the first peak only.

Spotted coat (CCOLOR) – Presence of different coloured spots on the coat.

Stress-induced leukocyte proliferation (LEUKPRO) – Increase in leucocyte proliferation when the subject is under stress.

Subcutaneous fat thickness (SUBCFAT) – Thickness of the layer of fat directly beneath the skin.

Subjective abnormal odor (ABODOR) – Subjective scores describing the intensity of abnormal meat odour.

Subjective overall acceptability score (OVACCS) – A combined taste panel score describing overall quality.

Teat number (TN) – Number of teats on a sow in total (TN), on the left side (LTN) or on the right side (RTN). Also referred to as nipple number (NN).

Tenderness score (TEND) – A subjective assessment of meat tenderness, determined by a human on a taste panel.

Thoracic vertebra number (TVNUM) – Vertebra number in the thoracic section.

Total number born alive (NBA) – Number of offspring, minus the number of stillborn animals. Also called prenatal survival.

Total number of fibers (TNMF) – Number of myofibres in a section of muscle.

Total shear work (TSW) – Amount of work required to shear a meat sample; equal to shear force times displacement (joules).

Unsaturated index (FA-UI) – Measurement of the degree of fatty acid unsaturation.

Unsaturated to saturated fatty acid ratio (USR) – The ratio of unsaturated to saturated fatty acids in a sample of fat.

Uterine capacity (UC) – An indicator of the ability of the uterus to provide the necessary environment to maintain fetuses until farrowing.

Uterine horn size – Can be measured by length (UHL) or by weight (UHW).

Water holding capacity (WHC) – The ability of meat to retain water during cutting, heating, grinding and pressing. A quality trait that affects product processing, appearance and shelf life. Measured by weighing filter paper after it has been placed in contact with meat and collected moisture.

White blood cell counts (WBC) – The number of white blood cells per volume of blood. A high WBC count can be an indicator of infection.
Appendix III  Pig Disease and Defect Glossary

Pig diseases and defects were primarily obtained from the Online Mendelian Inheritance in Animals (OMIA) database (http://omia.angis.org.au/).

Agenesis of anal sphincter – Congenital absence of anal sphincter.
Agnathia – Congenital absence of the lower jaw.
Anal atresia (ANATR) – An abnormality where congenital absence of an opening at the bottom end of the intestinal tract is observed.
Anemia – Abnormally low number of red blood cells (erythrocytes) or abnormally low quantity of haemoglobin.
Anophthalmia – Congenital absence of one or both eyes, or the presence of rudimentary eyes.
Aplasia of tongue – Congenital absence of the median portion of the apex of the tongue.
Artery, anomaly of – Congenital abnormality of an artery.
Arthritis – Inflammation of a joint. Used interchangeably with arthrosis, which refers to degenerative disease of joints.
Arthritis deformans – Erosion of articular cartilage and destruction of sub-chondral bone. Regarded as an autoimmune disease.
Arthrogryposis – Persistent flexion of a joint.
Asymmetric hindquarter syndrome – Asymmetry of the hindquarters.
Ataxia, progressive – A progressive failure of muscle coordination, resulting in perverse movements.
Atherosclerosis – A common form of arteriosclerosis in which deposits of yellowing plaques (atheromas) containing cholesterol and other lipid material are formed within the arteries.
Atresia ani – Congenital absence of the anus, causing a build-up of faeces and consequent distension of the abdomen.
Atresia ilei – Congenital absence of the ileum (the distal portion of the small intestine, extending from the jejunum to the caecum).
Atrophic rhinitis – A chronic inflammation of the nose, characterized by atrophy of nasal mucosa, including the glands, turbinate bones and the nerve elements supplying the nose.
Blood group system – A set of blood types, each corresponding to a particular antigen on the surface of red blood cells. Includes blood groups A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, S.
Brachydactyly – Abnormal shortness of the digits.
Brachygnathia superior – Congenital abnormal shortness of the maxilla (upper jaw), resulting in protrusion of the mandible (lower jaw).
Cardiomyopathy, dilated – A disorder characterized by cardiac enlargement (especially of the left ventricle), poor myocardial contractility and congestive heart failure.
Cardiomyopathy, hypertrophic – Increase in volume of the muscle tissue of the heart due to an increase in the size of muscle cells, primarily in the left ventricle and ventricular septum.
Cataract – Opacity of the lens of the eye.
Cephalothoracophagus – A twin ‘monster’ united at the head, neck and thorax.
Cerebellar anomaly, congenital – Incoordination, ataxia, dizziness and unthriftiness, resulting in death. Associated with misshapen nuclei in some of the Purkinje’s cells in the cerebellum.
Cerebellar hypoplasia – Underdevelopment of the cerebellum.
Chondrodysplasia – Abnormal growth of cartilage, leading to disproportionate dwarfism.
Cleft palate – Congenital fissure (split) that involves the hard or soft palate (roof of the mouth).
Cloaca – Absence of the anus, causing faeces and urine to be voided through the vulva. Named after the structure in birds through which faeces and urine are normally voided together.
Cranioschisis – A cleft in the skull involving one or more of the bones of the cranial cavity.
Cryptorchidism – Failure of one (unilateral) or both (bilateral) testes to descend.
Cyclopia – A congenital developmental disorder characterized by a single orbital fossa (eye socket). Named after the race of one-eyed giants of Greek mythology.
Cystic bile ducts and renal tubules – Animals appear normal at birth, but develop distended abdomens and show signs of distress between 24 and 48 h. Death follows within 10 days.
Dicephalus – Developmental disorder resulting in two heads.
Dwarfism, chondrodystrophy – A skeletal disorder caused by one of the myriad genetic mutations that can affect the development of cartilage; leads to small body size.
Epididymal aplasia – Total or partial failure of development of the epididymis.
Epitheliogenesis imperfecta – Congenital absence of the skin.
Fragile site – A site on a chromosome that does not stain, at which a break in the chromosome often occurs.
Gangliosidosis, GM2 – A lysosomal storage disease in which there is a build-up (storage) of GM2 gangliosides (a type of glycolipid) in various tissues, due to the lack of the enzyme hexosaminidase, whose task is to break down the GM2 ganglioside into its constituents.
Goiter, familial – Enlargement of the thyroid gland, causing a swelling in the front of the neck – a goitre.
Harelip – A congenital disorder comprising a cleft in the upper lip, resulting from failure of fusion between the maxillary and medial nasal processes.
Hemophilia A – Impaired coagulability of the blood, with a consequential strong tendency to bleed, resulting from a deficiency of clotting factor VIII.
Hernia – Protrusion of part of an organ or tissue through the structures normally containing it. Depending on the locations and degrees, the types of hernia include diaphragmatic hernia (the displacement of abdominal organs into the thorax), inguinal hernia (part of the intestine protrudes through the abdominal wall; IHERN), scrotal hernia (inguinal hernia that has passed into the scrotum), umbilical hernia (protrusion of part of the intestine or other abdominal organs through the abdominal wall at the umbilicus), etc.
Heterochromia iridis – Difference in colour of the iris in the two eyes, or in different areas of one iris.
Hind limb paralysis – Paralysis of the hindlimb.
Hip dysplasia – Laxity of the hip joint, resulting from a shallow acetabulum and/or a small, misshapen head of the femur. In severe cases, affected animals are lame.
Histiocytosis, malignant – Also known as histiocytic sarcoma (HS).
Holoprosencephaly – Developmental failure of cleavage of the forebrain (prosencephalon), with a deficit in midline facial development and with cyclopia in the severe form.
Hydrocephalus – Enlargement of the cranium caused by accumulation of fluid.
Hydrops fetalis – Accumulation of fluid in the whole body of the newborn.
Hypercholesterolemia – The presence of high levels of cholesterol in the blood.
Hyperlipidemia – An elevation of lipids (fats) in the bloodstream.
Hyperostosis – An excessive growth of bone.
Hypocatalasia – A genetic disorder caused by deficiency of catalase enzyme.
Hypomyelinogenesis, congenital – Inadequate synthesis of myelin; includes failure of formation of myelin, plus incomplete and delayed myelination of axons.
Hypotrichosis, dominant – Dominant form of hair loss.
Hypotrichosis, juvenile with age-dependent emphysema – Juvenile hairlessness and age-dependent emphysema.
Hypotrichosis, recessive – Recessive form of hair loss.
Infertility – Inability to conceive. Has a wide variety of causes, some of which are at least partly genetic.
Intersex – A mixture of male and female characteristics, often due to an abnormality of the sex chromosomes.
Joint abnormalities (JOINTAB) – An assessment of abnormality in the joints, usually scored.
Joint lesions (JOINTL) – Presence of lesions on the joint.
Kartagener syndrome – Congenital defect in functioning of the cilia.
Knobbed acrosome – The presence of a protrusion (knob) on the acrosome (cap) of a sperm. Associated with sterility.
Leg weakness – Evidence of mis-formed legs (front or rear). May be a congenital defect of the limbs that prevents standing.
Lymphosarcoma – Malignant neoplastic disorder of lymphoid tissue.
Macrophthalmia – Abnormal enlargement of the eyeball.
Malignant hyperthermia – A progressive increase in body temperature, muscle rigidity and metabolic acidosis, leading to rapid death.
Mannosidosis, alpha – A lysosomal storage disease in which there is a build-up of mannose-rich compounds.
Megacolon – Also known as Hirschsprung disease. A disorder in which the large intestine undergoes a large dilation and fills with faecal mass.
Membranoproliferative glomerulonephritis type II – A progressive inflammation of the capillary loops in the glomeruli of the kidney.
Microphthalmia – Abnormal smallness in all dimensions of one or both eyes.
Miniature – A type of pig that is small in body size; usually less than 100 kg at maturity.
Muscular hypertrophy – Abnormal increase in muscular tissue caused entirely by enlargement of existing cells.
Neonatal diarrhea – Diarrhoea in newborn piglets, caused by E. coli K88 or K99.
Nephropathy – Disease of the kidneys.
Nipples, asymmetrical numbers – Asymmetrical numbers of nipples.
Nucleoside transport defect – Defect in the transport of nucleosides (purine or pyrimidine base attached to a ribose or deoxyribose sugar) across erythrocyte membranes.
Osteochondrosis – Abnormal differentiation of growth cartilage.
Otitis interna, susceptibility to – Susceptibility to inflammation of the inner ear, i.e. the vestibule, cochlea and semicircular canals.
Otitis media, susceptibility to – Susceptibility to inflammation of the middle ear.
Persistent frenulum preputii – A close attachment of the penis to the prepuce, resulting in inadequate protrusion of the penis.
Polydactyly – Having more than the normal number of digits.
Porcine stress syndrome (PSS) – A syndrome, commonly initiated by extreme physical stress such as fighting, marked by difficult breathing, increased rate of respiration, blanching and reddening of the skin, and followed by cyanosis and acidosis. The symptoms include total collapse, muscle rigidity and extreme hyperthermia, and generally result in death. It is caused by a single autosomal recessive gene, called the ryanodine receptor gene.
Porphyria, congenital erythropoietic – A rare inborn error of porphyrin-haem synthesis inherited as an autosomal recessive trait.
Porphyria cutanea tarda – The most common subtype of porphyria. The disorder results from low levels of the enzyme responsible for the fifth step in haem production.
PSE – Refers to pork that is pale, soft and exudative (watery surface) and may result from porcine stress syndrome.
Renal hypoplasia – An abnormally small kidney. May be bilateral or unilateral.
Respiratory distress syndrome – A disorder characterized by breathing problems in newborn pigs.
Retinitis pigmentosa – A group of genetic eye disorders.
Rhabdomyolysis – Disintegration of muscle fibres, with consequent excretion of myoglobin in the urine.
Rhinitis (RHIN) – Inflammation of the nasal mucosa.
Syndactyly – Fusion of the claws (cleats or digits). Occurs in one, two, three or four legs.
Testicular feminization – An abnormality of sexual development in which affected individuals have an XY chromosomal constitution, undescended testes and female secondary sexual characteristics (including female external genitalia).
Thrombocytopenic purpura, autoimmune – Characterized by a low platelet count, normal bone marrow and the absence of other causes of thrombocytopenia.
Thrombopathia – A disorder of blood coagulation resulting from a failure of ADP release from platelets on stimulation by aggregating factors such as thromboplastin.
Twinning, conjoined – Congenital disorder in which twins are partly joined.
Urolithiasis – Formation of calculi (‘stones’) in the urinary tract, or the condition associated with the presence of such calculi.
Ventricular septal defect – A congenital heart defect characterized by persistent patency of the ventricular septum, permitting flow of blood directly between ventricles, bypassing the pulmonary circulation and resulting in various degrees of cyanosis.
Vitamin D-deficiency rickets, type I – Inherited deficiency of the enzyme 1-hyroxylase. The most noticeable effects include a failure of calcification of bones (leading to bowing of limbs) and delayed dentition.
Von Willebrand disease – The most common hereditary coagulation abnormality.
Wilms tumor – A rapidly developing malignant mixed tumour of the kidneys.
Wilson disease – A disorder of copper metabolism due to a deficiency of ceruloplasmin, which forms a complex with copper. The excess copper is deposited in the brain, causing mental retardation, or the liver, causing jaundice and cirrhosis.
Index

ABO gene 54
activation-induced cytidine deaminase (AID) 117
adeno-associated virus (AAV) vectors 250
ADG see average daily gain
adrenocortical function 208, 210–211
AFLP see amplified fragment length polymorphism
African pigs
  domestic 31, 307, 311, 454
wild 3, 5–6
aggression 211–213
  maternal 205–206, 207
Agouti (A) locus 39, 45–46
Al see artificial insemination
Albino (C) locus 39, 46
allantois 281
allele nomenclature 475
allelic diversity 315
allelic exclusion 105
allelic richness 314–315
altrnogenest 243
American pigs
  domestic 307, 311, 455–456,
    Plates 4–12
  wild 2
AmiGO ontology browser 478
amnion 281
amplified fragment length polymorphism (AFLP) 316, 317
Anatolia 19, 20–21
androstenone 223, 224, 227, 359, 372–373
aneuploidy 140, 146–147
Angeln Saddleback breed 450, Plate 17
angiogenesis 284
animal models of human disease 427–429, 432–433
  pigs as 246, 253–254, 429–432, 433–434, 436–437
cancer 434–436
  retinitis pigmentosa 250
Animal Trait Ontology (ATO) database 476, 478
antibodies see immunoglobulins
aplasia of the tongue 54
Arapawa Island breed 454
archaeology, evidence of domestication 18, 19
Asia 24, 26–27, 28
Europe 16–17, 22–23
Near East 20–21
ARE-1P/ARE-2P (artiodactyl repetitive element-1/-2 porcine) 80
array-CGH 139
arthrogryposis multiplex congenita (AMC) 54
artificial insemination (AI) 220, 243–244, 396
Asian pigs
  breeds 452–453, 457–458, Plates 22–27
crout colour 42, 44, 47
domestication 17, 18, 19, 20–21, 23–31
genetic resources 307, 311
wild 3, 4–5, 6, 8–9
see also Meishan breed
ASIP gene (agouti signalling protein) 39, 45–46
assisted reproduction 220, 242–245, 396–397
ataxia, progressive 54
ATO (Animal Trait Ontology) database 476, 478
ATP1A1/ATP1B1 genes (ATPase) 85
Australia 28, 454
average daily gain (ADG)
  correlation with other performance traits 202, 330–335
correlation with reproduction traits 342, 343
heterosis and 347
selection experiments 335–336, 336–337
B cell receptors (BCRs) see immunoglobulins
Ba Xuyen breed 452
Babyrousa (babirusa) spp. 3, 4–5, 148, 150
back test 210
backcrossing 409
backfat thickness
correlation with reproduction traits 343, 344–345
feed intake and 328
heritability/correlations with other
performance traits 330, 331, 332, 333, 334, 335
as measure of carcass composition 326, 357, 358, 360
QTLs 366–369
selection experiments 335–336, 336–337
variations in terminology 477
background selection 409
BACs (bacterial artificial chromosomes) 154–155, 156, 185
contig maps 155–156, 185–186
Plate 3
in genome sequencing 187–188
Bantu breed 454
basal cell carcinoma 435
Bazna breed 450
BCRs (B cell receptors) see immunoglobulins
bearded pig (Sus barbatus) 8, 149
behaviour genetics 200–213
aggression 205–206, 207, 211–213
breeding programmes 202, 405
emotional behaviours 208–211
feeding and drinking 202–204
maternal behaviours 205–208
pig well-being and 201–202
QTLs 201–202, 203–204, 206, 211
sexual behaviours 204–205
Beijing Black breed 452
Belarus Black Pied breed 450
Belt (Ibre) allele 41, 46
Bentheim Black Pied breed 450, Plate 18
Berkshire breed 44, 376, 449, 460, Plate 13
best linear unbiased prediction (BLUP) 400, 402
biodiversity see genetic diversity
biomedical sciences 426–437
animal models of human disease 427–429, 432–433
cancer 434–436
pigs as 246, 253–254, 429–432, 433–434, 436–437
retinitis pigmentosa 250
production of heterologous proteins 245–246
research uses 146, 432, 436–437
xenotransplantation 78–79, 253, 431
birth weight 225, 226
Black Slavonian breed 450
black spotting (E) allele 42–44
blastocysts 268, 269, 276–277
blood groups 54–55
blood parameters, QTLs 122
BLUP see best linear unbiased prediction
boar taint 359, 372–373, 405
boars
performance testing stations 400
reproduction traits 222–223, 223, 224
correlation with growth rates 227
in crossbred animals 221
QTLs 228, 229
reproductive biology 205, 219–220, 265–266
sire lines 394, 397, 400–402
Brachyury (T) gene 278, 279, 285
breed organizations
breed societies 394–395, 396, 446
breeding companies 395, 396, 459–460
breeding 390–417
behavioural traits 202, 405
breeding objectives 326, 392, 397–399, 404
breeding pyramid 326, 392, 393–394, 404–405
environmental influences 326–327, 404
GS 414–416
MAS 407–414
phenotypic selection, limitations of 406–407
principles of genetic improvement 391–392
programme designs 400–406
selection indexes 325–326, 392, 398–400, 410–411
see also breed organizations; breeds; crossbreeding
breeds 445–461
allelic diversity 315
coil colour as a marker 46–47
conservation 309–311, 315, 446–447
definition of 445–446
degree of endangerment 308–309, 315
descriptions of 447, 448–454, Plates 4–27
distinctiveness (D) 315
genetic diversity 312–314
history of 316, 445
identification and traceability 47, 316–317, 373–374, 414
inventories 307–308, 394–395, 446
meat quality/carcass composition traits 374–376
reproduction traits 220–221
research information sources 447, 455–459
synthetic (composite) 394, 409–410, 459–460
taxonomic (phylogenetic) trees 316, 317, 460, 461
see also breed organizations; breeding; crossbreeding

British Lop breed 449
Brown (B) locus 39, 46
BTG gene family (B-cell translocation) 289
Bulgarian White breed 450
bush pig (Potamochoerus larvatus) 5–6, 148

C-value paradox 74
CAGE (cap analysis gene expression) 184
calcium transient 266
calpastatin 364
Canada 393, 395
cancer
lymphosarcoma in the pig 58
pig as a model for human disease 434–436
cannibalism 205–206
Cantonese breed 452
carcass length 358, 360, 369
carcass traits 357
breed differences 374–376

correlation between carcass traits 359, 360

correlation with meat quality traits 360, 362

correlation with reproduction traits 227
crossbreeding and 376
genes affecting 364–365, 413
heritability of 357, 358, 359
QTLs 365–370
traceability 47, 373–374, 414
see also backfat thickness

cardiovascular disease, pig as a model for 430, 434
casein gene family 86
CAST gene (calpastatin) 364
castration 405
cathepsins 365
cattle–pig zoo-FISH mapping 159
Çayönü Tepesi (Anatolia) 20
CCKAR gene (cholecystokinin-A receptor) 341
cDNA probes 181–182, 184
CDR3 (complementarity determining region) 116, 117
Cdx2 gene 277, 280
cell polarity 274, 277, 279–280
centric fusion (Robertsonian) translocations 144, 147
centromeres 76–77, 137
Cetartiodactyla 1–2
CGH see comparative genetic hybridization
chemokines 283
Chester White breed 448, Plate 4
chimerism 145
Chinese pigs
breeds 452–453, 457–458, Plates 22–27
colour 47
domestication 19, 23–27
see also Meishan breed

cholecystokinin-8 (CCK) 338–339, 340, 341
cholesterolaemia 57
chorion 281
chromosome maps 150–160, 185–186, 187
immunoglobulins 111, 112
MHC classes I and II 106, 107–108
TCR genes 120

chromosomes
abnormalities 139–147, 192–193, 293
breeding out 406
banding studies 135–137
cloned pigs 147
comparative studies
human–pig 158–159, 189–190, Plate 2
Sus scrofa subspecies 7
Swiidae 147–150
FISH analysis 137–139, 157–159
karyotypes
domestic pig 136, 146
translocation in a feral pig 150
NORs 135–137
telomeres 83–84, 137, 147
clonal libraries 154–155, 156
cloning
biomedical research 432
cytogenetics 147
as method of reproduction 245, 247–248, 397
tissue samples 183

C-Mos knockout mice 266
CNVs see copy number variants

coagulation disorders 64, 67
coat colour 32, 38–47, 39, 374, Plate 1
colostrum 219, 226–227
colour of meat 357, 370, 375
commercial organization of the pig industry 392–395, 396, 459–460
comparative chromosome maps 158–159, 189–190, Plate 2
comparative genetic hybridization (CGH) 139
composite (synthetic) breeds 394, 409–410, 459–460
conception rates 219, 221, 227
congenital tremor syndromes 65
conservation of genetic diversity 309–311, 315, 446–447
consered synteny 158–159, 189–190, Plate 2
contig maps 155–156, 185–186, Plate 3
cooking loss 360, 361
copy number variants (CNVs) 87, 139, 192–193
corpus luteum 243, 282
cortisol 211
CRC locus (calcium release channel) 87
Cre recombinase 251
crossbreeding
breeding programmes 392, 403–404, 405
marker-assisted selection and 408–410, 411
crossbreeding (continued)
meat quality and 376–377
performance traits and 346–347
reproduction traits and 221–222
cryopreservation of gametes and embryos 244, 310–311, 315
cysts, renal 63
cytochrome P450 gene family 66, 372, 434
cytogenetics see chromosomes
cytokines 282–283
cytoskeleton, in the embryo 277
Czech Improved White breed 450

Da Min breed 452, Plate 22
dam lines 39, 394, 397, 402–403
DAXI gene 293
DDX4 gene 290
deepCAGE 184
degree of endangerment (DE) 308–309, 315
demethylation of DNA 251, 266, 275
Denmark 393, 395
Dermantsi Pied breed 450
dermatosis vegetans 55
developmental genetics 263–294
embryogenesis 266–271
gametogenesis 219, 264–266
 genetic control
implantation and placental development 281–285
post-implantation 285–289
pre-implantation 272–281
sex differentiation 64, 145, 289–293
diarrhoea, neonatal 61
Dilution (D) locus 39, 46
disease resistance 109–110, 122, 253
distinctiveness of breeds (D) 315
diversity see genetic diversity
DNA
cDNA probes 181–182, 184
GC content 75–76, 91
methylation/demethylation 91, 251, 266, 275, 276
mitochondrial 92–93
repetitive sequences 75–76, 186–187, 188
centromeric repeats 76–77, 137
CNVs 87, 139, 192–193
expressed sequences 84–87
microsatellites 82–83, 191, 317, 319
minisatellites 81–82
telemeric repeats 83–84, 137, 147
transposons 77–81
single-copy 87
DNA-combing 157
domestication 15–16
of Sus scrofa 14, 16–33, 445

Dominant black (Eb) allele 42, 43
Dominant white (I) locus 39–42, 41, 46
DRB locus (MHC class II) 86
dressing percentage 357, 358, 360, 369–370
drinking behaviour 204
drip loss 360, 361, 363
Duroc breed 394, 448, Plate 5
production traits 336, 375, 376
reproduction traits 221
dwarfism 56

eating quality traits 357, 358, 360, 361, 364, 371–372
EBVs see estimated breeding values
economic considerations
economic weights of breeding traits 326, 398–399, 403
marker-assisted selection 411–412
ectoderm, embryonic development 285
EDNRB gene (endothelin receptor beta) 47
effective population size (N) 308
Egypt 31
Eif41b gene 273
electroporation 248
embryonic genome activation (EGA) 273, 274–275
embryonic stem (ES) cells 248, 252
embryos

cryopreservation 244
cytogenetic analysis 147
death of 219, 270–271
embryogenesis 266–271

genetic control of development
implantation and placentaion 281–285
post-implantation 285–289
pre-implantation 272–281
sex differentiation 64, 145, 289–293
sexing of 76–77
transfer (ET) 244–245, 396–397
emotional behaviours 208–211
endoderm, embryonic development 278, 279, 285
environmental influences on breeding outcomes 326–327, 404
Eomes gene 277, 280
epigenetics 76, 91
in early embryos 275–276
in gametogenesis 266
imprinting 276, 412, 413
in SCNT 251–252
epitheliogenesis imperfecta 56
Escherichia coli F4 (formerly K88) 61
ESR (estrogen receptor) gene 231, 413
estimated breeding values (EBVs) 400, 402, 408, 410

genoic (G-EBV) 415, 416
mothering ability 205
pig survival 208
ESTs see expressed sequence tags
ET see embryos, transfer
ethical concerns 201, 398–399, 405
Eurasian wild boar (Sus scrofa) 6–7, 9, 10, 148–149
domestication 14–33
European pigs
breeds 449, 450–451, 455–456,
Plates 13–21
coat colour 44, 47
domestication 17–18, 19, 20, 21–23
genetic resources 307, 311, 316
see also individual breeds
evolution 2, 3, 6, 7, 10, 316
molecular genetics 80–81, 159
evolutionary break-point regions (EBRs) 143
expressed sequence tags (ESTs) 180
expression profiling 180–184
Extension (E) locus 39, 42–45, 43
extinction risk 308–309, 315
extra-embryonic membranes 281–282
extrapolation (calculation of allelic richness) 314–315
eye development 288
farrowing 219, 226
fat content of meat 357, 358, 359, 360, 361, 365
fatty acid composition 372
marbling 371, 375
QTLs 366–369
see also backfat thickness
feed conversion rate (FCR) 331, 333, 334, 335, 343
feed intake (FI) 331, 333, 334, 335
physiology of high versus low FI lines 339
relationship with lean tissue gain 327–330
selection experiments 335, 336, 337
see also residual feed intake
feeding behaviours 202–204
females see sows
Fengjing breed 452, Plate 23
fertilization 219, 266
FI see feed intake (FI)
fibre-FISH 157
fibroblasts
in vitro transformation 435, 436
SCNT 251
fingerprinted contig (FPC) maps 155–156, 186
FISH see fluorescence in situ hybridization (FISH)
flow sorting of chromosomes 137–138, 158
fluorescence immunocytochemistry 139
fluorescence in situ hybridization (FISH) 137–139, 156–159
foliculogenesis 264, 265, 290
foreground selection 409
forest hog (Hylochoerus meinertzhageni) 5, 148
fragile sites (FS) 144–145
freezing of gametes and embryos 244, 310–311, 315
G-banding 136
G-EBV (genomic estimated breeding values) 415, 416
G-value paradox 74–75
gametes
cryopreservation 244, 310–311, 315
gametogenesis 219, 264–266
see also oocytes; sperm
gametic imprinting 276
gangliosidosis 56
gastrointestinal disease, pig as a model for 430, 434
gastrulation 269, 278–279
GC content of DNA 75–76, 91
gene banks 310–311, 315
Gene Expression Omnibus (GEO), microarrays 181–182
genetic hitchhiking 318
genome knockout/knock-in (gene targeting) 248–251
gene mapping 150–160, 185–186, 187
immunoglobulins 111, 112
MHC classes I and II 106, 107–108
TCR genes 120
gene nomenclature 473–475, 480
gene silencing 292
genetic diversity 306–320, 392
breed inventories 307–308, 394–395, 446
conservation 309–311, 315, 446–447
extinction risk 308–309, 315
genomic variation 190–193
markers 311–317
marker–trait associations 318–320
genetic engineering see cloning; transgenics
genetic evaluation 391–392, 400, 402
genetic improvement 390–417
behavioural traits 202, 405
breeding objectives 326, 392, 397–399, 404
breeding programme designs 400–406
breeding pyramid 326, 392, 393–394, 404–405
GS 414–416
MAS 407–414
phenotypic selection, limitations of 406–407
principles of 391–392
role of assisted reproduction 242–245, 396–397
role of breeding organizations 394–395, 396, 459–460
genetic improvement (continued)
  selection indexes 325–326, 392, 398–400, 410–411
  see also crossbreeding; domestication
  genetic lag 392, 394, 397
  genomic imprinting 276, 412, 413
  genomic libraries 154–155, 156
  genomics 179–193
  in biomedical research 436–437
  comparative 158–159, 189–190, Plate 2
  databases 180, 181, 186, 188, 192, 228, 480
  eQTLs 373
  ESTs 180
  expression profiling 180–184
  genomic postulates 427
  GS 414–416
  ncRNA 74, 89–90, 184–185
  reproduction traits 232
  sequencing 155–156, 179, 186–189, 193
  variation within the species 190–193
  see also chromosome maps; molecular genetics
  genotype nomenclature 475
  gilts, reproductive biology 219, 227, 243
  glomerulonephritis 60
  Gloucestershire Old Spot breed 449, Plate 14
  glycogen metabolism 357, 363
  gonadotrophins 243
  Goosecoid (GSC) gene 278
  GPI pseudogene (glucosephosphate isomerase) 81
  growth factors see insulin-like growth factors; vascular endothelial growth factor
  growth hormone 87–88, 91
  growth rates
    correlation with reproduction traits 227, 342–346
    covariation of performance traits 330–335
    heterosis and 347
    model of relationships between feed intake and tissue gain 327–330
    physiology and molecular biology of 338–342
    selection experiments 335–338, 341, 344–345
    transgenic pigs 252–253
  GS (genomic selection) 414–416
  Guinea breed 454
  Hausdorff distance 312
  heart, embryonic development 288
  Hereford breed 448, Plate 7
  heterochromia iridis 56
  heterosis 221–222, 346–347, 376–377
  heterosynapsis, early 145–146
  Heude’s pig (Sus bucculentus) 4, 150
  Hezuo breed 452
  hind limb paralysis 57
  histones 76, 252, 275–276
  HOX (homeobox) genes 286, 287
  human approach test 210
  human–pig comparative maps 158–159, 189–190, Plate 2
  hybrid vigour (heterosis) 221–222, 346–347, 376–377
  hybridization, between Sus species 9–10, 28, 30
  Hylochoerus meinertzhageni (giant forest hog) 5, 148
  hyperacute rejection 253, 431
  hypercholesterolaemia, spontaneous 57
  hypothalamic–pituitary–adrenal (HPA) axis 208, 210–211
  hypotrichosis 57–58
  Iberian pigs 47, 374, 450
  ICSI see intracytoplasmic sperm injection
  IGFs see insulin-like growth factors
  IGH locus (heavy chain immunoglobulins) 86, 111–114, 115, 116–117
  IGL/IGK loci (light chain immunoglobulins) 88–89, 115–116
  Illumina 60K Beadchip 193
  immotile short-tail sperm (ISTS) defect 64
  immunization 110
  immunogenetics 101–124
    BCRs see immunoglobulins
    innate immunity 103–105, 121
    MHC (SLA) 86, 102, 105–110, 186
    quantitative trait loci (QTLs) 109, 121–124
    TCRs 89, 102, 105, 117–121
    transplant rejection 253, 431
  immunoglobulins 89, 102
    heavy chains (IGH genes) 86, 111–114, 115, 116–117
    light chains (IGL and IGK genes) 88–89, 115–116
    response associated with SLA genes 110
  somatic recombination 88–89, 105, 115–117
  subtypes 111
  IMMpRH2 panel 154
  implantation 281–285
  IMpRH panel 153–154
  imprinting 276, 412, 413
  inbreeding 308, 310, 392, 402, 407
  H-FARP gene (heart fatty-acid binding protein) 371
  hair, curly 67
  hairlessness 57–58
  HAL locus (halothane sensitivity) (RYR1 gene) 59, 209–210, 232, 361–363, 369, 374
  ham, variation in processing yield see RN locus
  Hampshire breed 363, 375–376, 448, Plate 6
index selection 325–326, 392, 398–400, 410–411
India 9, 30–31
Indochina 19, 25, 27–29
Indonesia 5, 8, 10, 19, 29–30
infanticide 205–206, 207
inherited disorders 51–53, 54–67
inner cell mass (ICM) 269, 277
INRA (Institut National de la Recherche Agronomique) 140, 457
insulin-like growth factors (IGFs)
  performance traits and 338, 339, 339–340
  \( IGF2 \) in carcass composition 364–365, 413
  in pregnancy and the fetus 283–284, 289
interchromosomal effects (ICEs) 146
interferon 282
interferon-stimulated genes (ISGs) 282–283
Internet resources
  breeds 307, 447
  gene nomenclature 474, 475
  genome databases 180, 181, 186, 188, 192, 228, 480
  inherited traits and disorders 52–53, 476
  ISGs 282
  pig models of human disease 436
  SLA polymorphisms 108
  trait nomenclature 476, 478
intersexuality 145, 293
intracytoplasmic sperm injection (ICSI) 247
interregression 409, 413
inventories of genetic resources 307–308, 394–395, 446
inversions (chromosomal) 144
ISEA (Island South-east Asia) 5, 8, 10, 19, 29–30
isochores 75–76, 91
Italy 19, 22, 23

Javan warty pig (\( Sus \) verrucosus) 8, 149
Jinhua breed 452

kallikreins 86
karyotype
  domestic pig 136, 146
  translocation in a feral pig 150
Kele breed 452
keratinocytes 436
kidney disorders 60, 63
kinetochores 76
\( KIR \) gene (killer immunoglobulin receptor) 102
\( KIT \) locus (tyrosine kinase receptor) 40–42, 41, 374
Krskopolje breed 450
Kunekune breed 454

Lacombe breed 448
Lacorie breed 450
lactation 219, 226–227
transgenic sows 252
Landrace breed 394, 450, Plate 19
  feeding behaviour 203
  production traits 333, 375
  reproduction traits 220–221, 221
Lanyu (Taiwan) 26
Large Black breed 449
Large Black-White breed 452
Large White breed 394, 449
  crossbred with Meishan 222
  production traits 333, 336, 376
  reproduction traits 220, 221
see also Yorkshire breed
LD see linkage disequilibrium markers
LE see linkage equilibrium markers
lean tissue, relationship to feed intake 327–330
lean tissue feed conversion (LTFC) 329, 331, 333, 348
  correlation with reproduction traits 345, 346
  selection experiments 337–338, 341, 345
lean tissue growth rate (LTGR) 329, 331, 333
  correlation with reproduction traits 345, 346
  physiology of high versus low LTGR lines 339
  selection experiments 336, 337–338, 341, 345
lean-to-fat ratio 360, 362, 448–454
left–right asymmetry 280
legless 58
\( LEPR \) gene (leptin receptor) 229, 369
leptin 339, 340
Lewontin–Krakauer (LK) test 318
libido, male 205
\( LILR \) gene family (leucocyte immunoglobulin-like receptor) 102
limb abnormalities 54, 57, 58
limb morphogenesis 288
LINEs see long interspersed nuclear elements
linkage disequilibrium (LD) markers 320, 407–408
linkage equilibrium (LE) markers 407, 408
linkage maps see chromosome maps
Lithuanian Native breed 451
litter size 218
  breed differences 220
  breeding programmes 402, 403
  chromosomal abnormalities and 140, 144
  correlation with other traits 227, 342–343, 344–345, 403
  genetic effects 222, 223–224, 225–226
  QTLs 229–230
litter survival see piglet survival
locus nomenclature 474–475
loin muscle dimensions 326, 357, 360, 370
long interspersed nuclear elements (LINEs) 78, 79, 187
long terminal repeat (LTR) elements 76, 186
LTFC see lean tissue feed conversion
LTGR see lean tissue growth rate
LTR see long terminal repeat elements
Luchuan breed 452, Plate 24
lymphosarcoma 58

males see boars
malignant hyperthermia 59, 361–362
see also RYR1 gene
Mangalitsa breed 45, 46, 451, Plate 20
MAP kinase 266, 289
marbling 371, 375
marker-assisted selection (MAS) 407–414
maternal behaviours 205–208
maternal imprinting 276, 412, 413
MC1R gene (melanocortin-1 receptor) 32, 42–45, 374
MC4R gene (melanocortin-4 receptor) 341, 365, 366
meat quality 355–377
breed differences 374–376
coat colour and 39
crossbreeding 376–377
genes affecting 59, 360–364, 365, 413
marker-assisted selection 411
QTLs 365–366, 367–368, 370–373
reproduction traits and 228
traits associated with 356–357
correlation with carcass composition traits 360, 362
heritability 357, 358, 359
medical uses see biomedical sciences
meiosis 139, 264, 265–266, 266
meiotic linkage maps 151–152
Meishan breed 452, Plate 25
crossbred with Large White 222
production traits 375, 376
reproduction traits 220
melanocortin receptor genes
MC1R 32, 42–45, 374
MC4R 341, 365, 366
membranoproliferative glomerulonephritis
type II 60
mesoderm, embryonic development 278, 279, 285
methylation
of DNA 91, 251, 266, 275, 276
of histones 275–276
MHC (major histocompatibility complex) 86, 102, 105–110, 186
microarrays 180, 181–182
microdissection 138–139, 157–158
microRNA (miRNA) 74, 86, 90–91, 184, 251
micrornas 82–83, 191, 317, 319
Middle White breed 449, Plate 15
milk production 219, 227
transgenic sows 252
minisatellites 81–82
Minzhu breed 452
miRNA see microRNA
mitochondrial genome 92–93, 192
markers of domestication 18, 21, 25, 26, 30–31, 31–32
MLEs (Mariner-like elements) 77
molar tooth morphology, domestication and 27, 30
molecular genetics 73–93
behavioural traits 201–202, 203–204, 206, 211
breeds 460
coat colour 38–47
developmental 266–267, 272–294
gene structure and function 87–91
genetic diversity markers 311–320
genome size 74–75, 91–92, 188
identification and traceability 47, 316–317, 373–374, 414
immune system 86, 88–89, 102–124
marker-assisted selection 407–417
meat quality/carcass composition
360–373, 377
mitochondrial genome 92–93
ncRNA 74, 89–91, 184–185
performance traits 340–341
repetitive DNA 75–76, 186–187, 188
centromeric repeats 76–77, 137
CNVs 87, 139, 192–193
expressed repetitive sequences 84–87
microsatellites 82–83, 191, 317, 319
minisatellites 81–82
telomeric repeats 83–84, 137, 147
transposons 77–81
reproduction traits 228–232
single-copy DNA 87
see also genomics
Mong Cai breed 452, Plate 26
Mora Romagnola breed 451
morphology 51, 52
Babyrous spp. 4–5
Hylochoerus meinertzhageni 5
Phacochoerus spp. 5
Potamochoerus spp. 5–6
Sus spp. 8, 9
motor neuron disease 60
Moura breed 448
mRNA (messenger RNA) 88
oocyte 267, 272–274
RNAseq 183–184
MS see microsatellites
MTTP gene (microsomal triglyceride transfer protein) 365
Mukota breed 454
Mulefoot breed 448
muscles
conversion to meat 356
see also RN locus
growth and relationship with feed intake 327–330
myogenesis 288–289
progressive myopathy 62

Napole yield 357, 361
see also RN locus
national breeding programmes 393, 395
ncRNA (non-coding RNA) 74, 89–90, 184–185
Near East 19, 20–21, 23
Neijiang breed 452
Neolithic Age, pig domestication 16–33
neonatal diarrhoea 61
neuroendocrine responses to stress 208, 210–211
neuropeptide Y (NPY) 339
neurulation 269, 279
neutrality of markers, tests for 318
New Guinea 8, 20, 29, 30
next-generation sequence technology 183–184
Ningxiang breed 453
NOD genes (NLR (nucleotide-binding domain, LRR-containing) proteins) 104–105
Nodal signalling 279, 285
nomenclature 473–480
non-economic values of breeding traits 201, 399
NORs see nucleolus organizer regions
notochord 279
NPY see neuropeptide Y
NROB1 (DAX1) gene 293
nuclear mitochondrial sequences (numts) 93
nucleolus organizer regions (NORs) 85, 135–137
nucleoside transport defect 61
nucleus herds 326, 393–394
nursing behaviour 219

obesity, pig as a model for 431, 432
OBO-Edit ontology editor 478, 479
Oceania
breeds 454
pig domestication 28, 29, 30
OCT4 gene 278, 280
oedema (myxoedema) 61
oedema disease (gut oedema) 63
oestrogens, produced by the conceptus 282, 283
oestrus 205, 219, 225
synchronization of 243, 244
oncogenes 435, 436
proto-oncogenes 277
Online Mendelian Inheritance in Animals (OMIA) 52–53, 476
Online Mendelian Inheritance in Man (OMIM) 52
ontologies, for traits 478–479
oocytes
cytogenetic analysis 146–147
imprinting 276, 412, 413
mRNA and control of embryonic development 267, 272–274
oogenesis 219, 264–265, 266
transduction 247
organogenesis 285, 286–288
Ossabaw Island breed 448
ovary, embryonic development 289–290, 292
ovulation 219, 264
rate 225, 228–229, 230, 252
synchronization of 244
Oxford Sandy and Black breed 449
Pacific Clade 25, 28–30
PAGs see pregnancy-associated glycoproteins
painting probes 138
palaeogenetics 17–20, 21–22, 23, 24–26, 28–33
Pampa Rocha breed 448, Plate 11
Patch (P) allele 40
paternal imprinting 276
Pax genes 288
peccaries 2
pedigree analysis 309–311, 413
pedigree records 394–395
performance traits 325–348
by breed 448–454
correlations with reproduction traits 342–346, 403
crossbreeding and 346–347
genetic variation/covariation 330–335
heritability estimates 331–332
physiology and molecular biology 338–341, 342
selection experiments 335–338, 341, 344–345
selection methods 325–330, 399–400
see also individual traits
PERVs see porcine endogenous retroviruses
PG 600 243
pH of meat 370–371
as a quality indicator 357, 358, 360
Phacochoerus spp. (warthogs) 3, 5, 148
pharmaceutical research 433, 434
phenes 54–67
phenotype nomenclature 480
Philippine warty pig (Sus philippensis) 8, 150
phylogeny 1–2, 3–4, 7
physiology, pigs as biomedical models of 430
PI genes (protease inhibitors) 86
Piau breed 448, Plate 12
Piétrain breed 394, 451, Plate 21
meat quality traits 375, 376
Piétrain breed (continued)

reproduction traits 221
research information sources 459
piglet birth weight 225, 226
piglet survival
  behaviourial factors affecting 205, 207, 208, 219
  breed differences 220
  correlation with performance traits 343
  economic value in breeding programmes 403
  genetic effects 208, 222, 223–224, 226–227, 230
piglets, striped coats 45
PigQTL database 124, 228, 340, 366, 477
Pink-eyed dilution (P) locus 39, 46
placental development 281–282
Poland China breed 448, Plate 8
polarity of embryonic cells 274, 277, 279–280
polydactyly with otocephalic monster 62
porcine endogenous retroviruses (PERVs) 78–79, 253
porcine stress syndrome (PSS) 59, 361–362
PorcineSNP60 Beadchip 319, 377
Porcula salvanius (pygmy hog, Sus salvanius) 9, 149
pork value chain 392–393, 404
porphyria 62
post-weaning feeding 203, 327–330
see also performance traits
Potamochoerus spp. 3, 5–6, 148
PPAG genes (pregnancy-associated glycoproteins) 284–285
PRE-1 (porcine repetitive element-1) 79–81
pregnancy, maternal recognition of 282
pregnancy-associated glycoproteins (PAGs) 284–285
prenatal mortality 219, 226, 230, 270–271
PRKAG3 gene 363–364, 370–371, 375, 376
see also RN locus
processing meat quality traits see meat quality
production traits see carcass traits; feed intake; growth rate; meat quality
profit function method 398
progesterone receptor agonists (altrenogest) 243
prolificacy see litter size
pronuclear injection 246–247
proamine-2 deficiency 63
protease inhibitors 86
protectiveness, maternal 206
proto-oncogenes 277
provenance of meat products 47, 373–374, 414
PSE (pale, soft, exudative) meat condition 59, 361
Pseudofam database 81
pseudogenes 81, 86–87, 93, 107
puberty 219, 225, 227, 228, 230
correlation with performance traits 343–344
pygmy hog (Sus salvanius or Porcula salvanius) 9, 149
QTLs (quantitative trait loci)
  behavioural 201–202, 203–204, 206, 211
  expression QTLs (eQTLs) 373
  gene maps 152, 155
  health traits 109, 121–124
  imprinting 276, 412
  linkage with genetic markers 318–320, 407–408, 412
  meat quality and carcass composition 365–373
  performance traits 340
  PigQTL database 124, 228, 340, 366, 477
  reproduction traits 228–232
  radiation hybrid (RH) mapping 153–154
  rarefaction (calculation of allelic richness) 314
  rebreeding 225, 227, 344
  recombination
   homologous in construction of linkage maps 151, 152
   in genetic modification 248, 250
   site-specific 250–251
   somatic (immune system genes) 88–89, 102, 105, 115–117
  red coat colour 42
  red river hog (Potamochoerus porcus) 5, 148
  Red Wattle breed 454
  reflectance of meat 357, 361
  renal disorders 60, 63
  reproduction/reproduction traits 218–232
   assisted 220, 242–245, 396–397
   behaviours associated with 204–205
   biology of 219–220
   gametogenesis 219, 264–266
   sex differentiation 64, 145, 289–292
   cloning 245, 247–248, 397
   correlations between reproduction traits 224–227
   correlations with other traits 227–228, 342–346, 403
   genetic variation 220–224
   QTLs 228–232
   genomic studies 232
   transgenic pigs 252
   see also individual traits
  residual feed intake (RFI) 331, 333, 336, 339–340
  see also feed intake
  respiratory distress syndrome 64
Index

retinitis pigmentosa 250
retinoic acid (RA) 286
retrotransposons 78
retroviruses 78–79, 253, 436
reverse transcriptase 78, 79
RFI see feed intake; residual feed intake
RH (radiation hybrid) mapping 153–154
ribosomal RNA (rRNA) 84–85, 92, 135–137, 275
riptides 66
RN locus (Rendement Napole) 59, 232, 357, 359, 363–364, 369, 375
RNA see microRNA; messenger RNA; non-coding RNA; ribosomal RNA; transfer RNA
RNA-mediated transposons 77–81
Roan (IRn) allele 42
Robertsonian (rob) translocations 144, 147
rolling behaviour 205
RPSA pseudogene (ribosomal protein SA) 81, 87
rRNA see ribosomal RNA
RYR1 gene (ryanodine receptor 1) (HAL locus) 59, 209–210, 361–363, 369, 374
Saddleback breed 449
SAGE analysis see serial analysis of gene expression
SC analysis see synaptonemal complex analysis
SCH mapping see somatic cell hybrid mapping
SCNT see somatic cell nuclear transfer
segmentation 270, 285–286, 288–289
selection indexes 325–326, 392, 398–400, 410–411
selective sweep 318
semen 220, 244, 310–311, 315
see also sperm
sensory traits in meat quality 357, 358, 360, 361, 364
QTLs 370–372
serial analysis of gene expression (SAGE) 182–183
sex chromosomes 140, 290–292
sex differentiation 64, 145, 289–293
sex reversal 64, 145, 293
sexing
of embryos 76–77
of sperm 396
sexual behaviours 204–205
SINEs (short interspersed nuclear elements) 78, 79–81
single nucleotide polymorphisms (SNPs)
in genomic selection 414–415
identification 191–192, 193
as markers of genetic diversity 318–320
in TLR genes 103–104
sire lines 394, 397, 400–402
skatole 359, 372
skin diseases, pig as a model for 430, 435–436
skin-on meat products 39
SLA see swine leucocyte antigens
SNP chips 154, 193, 319, 377
SNPs see single nucleotide polymorphisms
social behaviour 201–202, 210, 405
societal concerns 201, 398–399, 405
somatic cell hybrid (SCH) mapping 152–153
somatic cell nuclear transfer (SCNT) 147, 245, 247–248, 249, 251–252
somatic hypermutation 102, 105, 117, 118
somatic recombination 88–89, 102, 105, 115–117
sows
dam lines 39, 394, 397, 402–403
maternal behaviours 205–208
reproduction traits 220, 221, 222, 223, 224
correlations between 225–227
crossbreeding 221, 222
molecular genetics 228–231
reproductive biology 219, 264–265
synchronization of oestrus 243, 244
spectratyping 117
sperm
cryopreservation 244, 310–311, 315
cytogenetic analysis 139, 145–146, 147
heritability of characteristics 222–223
imprinting 276
mRNA found in 267
numbers in ejaculate 220, 224
QTLs associated with 229
short-tail 64
sorting 396
spermatogenesis 219, 265–266
in transgenesis 247
spots
black spotting (E) allele 42–44
Spotting (S) locus 39
Spotted breed 448, Plate 9
SRY gene 290–291, 292, 293
stillborn piglets 219, 226, 230, 271
stress response 208, 210–211
stripes, in piglets 45
Suidae 2–10, 147–150
Suiformes 2
Suinae 5–10
Suini 3
Sulawesi warty pig (Sus celebensis) 8, 10, 19, 29–30, 149
super-traits 476–477
supply and distribution chain 392–393
Sus genus 6, 9–10
Sus barbatus (bearded pig) 8, 149
Sus bucculentus (Heude’s pig) 4, 150
Sus cebifrons (Visayan warty pig) 9, 149
Sus celebensis (Sulawesi warty pig) 8, 10, 19, 29–30, 149
Sus genus (continued)
  Sus philippensis (Philippine warty pig) 8, 150
  Sus salvanius (pygmy hog, Porcula salvanius) 9, 149
  Sus scrofa (Eurasian wild boar) 6–7, 9, 10, 148–149
domestication 14–33
  Sus verrucosus (Javan warty pig) 8, 149
Swabian-Hall Swine breed 451
swine leucocyte antigens (SLA) (MHC antigens) 86, 102, 105–110, 186
Swine Testing and Genetic Evaluation System (STAGES) 400
synaptonemal complex (SC) analysis 145–146
synteny, conserved 158–159, 189–190, Plate 2
synteny maps 152–153
synthetic (composite) breeds 394, 409–410, 459–460
systematics 1–10
  of domestic breeds 316, 317, 460, 461

T cells
  adaptive immune response 106–107
  antigen receptors (TCRs) 89, 102, 105, 117–121
T-box genes 279, 286–288
  Eomes 277, 280
  T (Brachyury) 278, 279, 285
Tamworth breed 449, Plate 16
tandem selection 410, 411
tandemly repeated elements 81–85
taxonomy 1–10
  of domestic breeds 316, 317, 460, 461
Tayassuidae (peccaries) 2
TCRs (T cell receptors) 89, 102, 105, 117–121
teat number 227, 228
telomerase 84
telomeres 83–84, 137, 147
tenderness of meat 357, 358, 361, 364, 371
  embryonic development 289–292
  size as a reproduction trait 224, 227, 229
testosterone 223, 224, 227
THA-banding 137
thrombopathia 64
Tuoch Nhieu breed 453
Tibetan breed 453
TLRs (Toll-like receptors) 103–104, 104
TNFA/TNFB genes (tumour necrosis factor alpha/beta) 85
Tongcheng breed 453, Plate 27
tongue aplasia 54
tonic immobility test 210
toxicology research 433, 434
  traceability of meat products 47, 373–374, 414
  trait nomenclature 475–479
  transcription 87–88, 89
  embryonic (EGA) 274–275
  mitochondrial 92
  transcription start sites 184
transgenes 243, 245–254
  agricultural uses 252–253
  biomedical uses
    disease models 246, 250, 253–254
    production of heterologous proteins 245–246
    xenotransplantation 253, 431
  methodologies 246–252
  translation, in mitochondria 92
tanslocations
    reciprocal (rcp) 140–143, 145–146
    Robertsonian (rob) 144, 147
transplantation 78–79, 253, 431
transposons 77–81
tremor syndromes 65
trophiocerderm 269, 277
Turopolje breed 451

UK, breeds originating in 449, 455–456,
  Plates 13–16
uniformity at slaughter 406
USA, breeds originating in 448, 455–456,
  Plates 4–10
vaccination 110
variation see genetic diversity
vascular endothelial growth factor (VEGF) 284
Vietnamese Pot Belly breed 453
Visayan warty pig (Sus cebifrons) 9, 149
vitamin D-deficiency rickets 66
von Willebrand disease 67
warthogs (Phacochoerus spp.) 3, 5, 148
water-holding capacity of meat 361, 375
wattles 67
websites see Internet resources
Wee1B gene 273
Weitzman method for analysis of genetic diversity 312–314
welfare issues 201, 398–399, 405
white coat colour 4, 39–42, 47, Plate 1
White head (He) locus 46
wild boar (*Sus scrofa*) 6–7, 9, 10, 148–149  
domestication 14–33  
woolly hair 67  
*World Dictionary of Livestock Breeds*  
*Types and Varieties* (Mason) 307  
YAC vectors (yeast artificial chromosome) 154  
Yellow River Valley 24, 26–27  
yolk sac 281  
Yorkshire breed 448, *Plate 10*  
see also Large White breed  
X chromosome 140, 292  
xenotransplantation 78–79, 253, 431  
Xist gene 292  
ZAR1 gene 273  
zinc finger nucleases  
(ZFNs) 250  
zona pellucida 269  
zo-FISH 158–159, *Plate 2*  
zygotes 267, 268, 273–274