Section B

Molecular tools for exploring genetic diversity

1 Introduction

Recent advances in the field of genomic technology have constituted a major innovation in livestock production. The increasing availability of molecular tools is deeply affecting the ways in which livestock species are studied and managed. This section provides an overview of recent developments related to molecular tools and their use, focusing particularly on the period since the first report on *The State of the World's Animal Genetic Resources for Food and Agriculture* (first SoW-AnGR) (FAO, 2007) was prepared.

The first SoW-AnGR noted that the main roles of molecular technologies in the characterization of AnGR include:

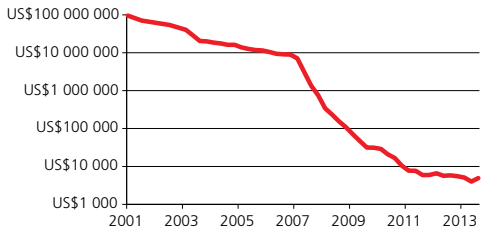
- assessing functional and neutral genetic variability within and between populations, including investigation of their history (domestication, expansion or reduction of the population size, migrations, introgression episodes, etc.);
- assessing the current state of a population in terms of risks related to inbreeding and genetic drift, using estimators such as effective population size; and
- genetic characterization of traits (e.g. physical appearance, productivity, disease resistance and other adaptability traits) specific to given populations.

The report highlighted the following three ongoing developments in molecular biology as being particularly relevant to AnGR management:

- the establishment of whole genome sequences for various livestock species;
- the development of technologies for measuring polymorphisms at loci spread across the entire genome; and
- the development of technologies for measuring gene transcription and expression on a large scale.

Since the first SoW-AnGR was prepared, the list of species whose genomes have been sequenced has continued to grow. It now includes chicken (2004), sheep (2010), cattle (2009), horse (2009), pig (2012), rabbit (2009), turkey (2009) and goat (2013). The costs of genotyping and sequencing have declined sharply during this period (Figure 4B1). High-density SNP arrays, allowing the simultaneous assay of several tens of thousands to several hundreds of thousands of SNPs, are available for use in livestock species at a cost of US$100 or less if a relatively large number of individuals are sequenced. Genomes

FIGURE 4B1
Change in cost per genome sequenced in humans



Note: Costs expressed on a logarithmic scale.
Source: Based on data from National Human Genome Research Institute (http://www.genome.gov/pages/der/sequencing_costs_jul2014.xlsx; accessed January 2015).

Box 4B1
From DNA to phenotype

The genome of livestock species is organized in pairs of chromosomes, each inherited from one of the parents and chemically made up of large molecules of DNA (deoxyribonucleic acid). Each gene in an individual, therefore, has two copies, known as alleles, one on each chromosome of a pair (with the exception of non-paired sex chromosomes). Chromosomes comprise genes and intergenic regions. The former encode proteins and other products. The latter, which represent the majority of the genome, are believed to play various regulatory roles (The ENCODE Project Consortium, 2012). Genes typically consist of coding sequences (exons) separated by introns and accompanied by regulatory elements. Like intergenic regions, introns carry no protein-coding information, but sometimes play a role in the regulation of gene expression.

Nucleic acids are strands of smaller molecules called nucleotides. DNA has four types of nucleotide (adenine, cytosine, guanine and thymine). The instruction encoded in a gene depends on the sequence in which these nucleotides are ordered. This sequence information is put into action in two steps. The first is the transcription (copy) of genetic information into another type of nucleic acid, RNA (ribonucleic acid). Both exons and introns are transcribed into a primary messenger RNA molecule. In the second step, these molecules (transcripts) are edited and eventually translated into proteins (particular three-nucleotide sequences correspond to particular amino acids, the molecules that constitute a protein). Gene expression is highly tissue-specific and time-dependent. Not all genes are translated into proteins; some express their function as non-coding RNA molecules that play important roles in protein synthesis (transfer RNA and ribosomal RNA) and various regulatory processes (microRNA and long non-coding RNA, two types of regulatory RNA that differ in terms of the number of nucleotides they consist of – approximately 20 and more than 200 nucleotides, respectively).

Phenotypic differences between individuals, populations and species are a consequence both of environmental effects (including epigenetic mechanisms – see Box 4B5) and of variations in DNA sequences. These variations may be caused by point mutations leading to the substitution of single nucleotides (single nucleotide polymorphisms – SNPs), insertions, deletions, duplications, copy number variations or inversions of DNA fragments. If SNPs are in exons, different alleles may lead either to the same amino acid (synonymous SNPs) or to a different amino acid (non-synonymous SNPs) being included in the resulting protein. DNA variations can be classified as "functional" or "neutral". In the case of functional variation, changes in the sequence of nucleotides in the DNA molecule induce changes in the phenotypic function of the organism. In the case of neutral variation, a change in the DNA sequence does not give rise to any change in function.

Because genes or SNPs that lie near each other on a chromosome (i.e. are physically "linked" to each other) tend be inherited together, a neutral variant can be associated with a functional one. In addition, this interdependence between linked genes means that the various combinations of their alleles are not distributed randomly (a phenomenon termed "linkage disequilibrium"). Linkage and linkage disequilibrium allow the use of one site of polymorphism as a "genetic marker" for polymorphism in a nearby region.

can be sequenced for less than US$3 000 each with moderate coverage (e.g. "eight-fold" coverage – meaning that, on average, each position in the genome is sequenced eight times). Sequencing smaller fractions of genomes (restriction site associated DNA sequencing – RAD-Seq) can be used directly in the characterization of individual animals (this is termed "genotyping by sequencing") (De Donato *et al.*, 2013). Similarly, the development of tools capable of assaying a high density of transcripts and even direct transcriptome sequencing (also known as "RNA-seq" – short for RNA sequencing), has increased capacity to study gene expression and hence to unravel

the complex physiological regulation of target traits (D'Alessandro and Zola, 2012).

2 Developments in the use of DNA markers

Progress in sequencing techniques and the opportunities offered by the development of high-density marker arrays have considerably improved the availability of DNA information over the last ten years, both in terms of the number of markers identified and in terms of the cost of genotyping.

Until recently, microsatellites remained one of the most popular types of marker in genetic characterization studies (Lenstra *et al.*, 2012), used for example in projects such as "GlobalDiv", which ran from 2007 to 2010 and combined microsatellite datasets from various diversity studies from different parts of the world (Ajmone-Marsan *et al.*, 2010). Microsatellite data continue to be used, especially in developing countries (e.g. Abdullah *et al.*, 2012; Azam *et al.*, 2012) and in the context of conservation and priority setting at regional level (e.g. Medugorac *et al.*, 2011; Ginja *et al.*, 2013). However, they are increasingly being superseded by the use of SNP marker arrays. With the advent of next-generation sequencing, mitogenomics (analysis of the whole mitochondrial genome rather than a limited fragment of mitochondrial DNA) can be routinely used in livestock species, including less intensively studied species such as goats (Doro *et al.*, 2014) and horses (Achilli *et al.*, 2012). The recent generation of whole genome reference sequences for many livestock species has allowed "population genetics" to become "population genomics". Population genomics uses large sets of SNPs to study specific variations across the genome and determine how they have been shaped by the history (e.g. changes in population size, selection, and cross-breeding) of livestock populations. SNPs can be assigned to various classes (neutral vs. genic, intron vs. exon or synonymous vs. non-synonymous), which provides opportunities for more detailed analysis of diversity. The past decade has also witnessed a revolution in sequencing technologies that has led to the development of various platforms for DNA and RNA sequencing, known collectively as next-generation sequencing technologies (see Metzker, 2010 and Davey *et al.,* 2011 for reviews). These tools can rapidly (in a few days or weeks) provide sequence data in the form of short reads (sequenced DNA fragments between 100 and 400 base pairs long on average) that collectively cover the whole genome of a sample (or the transcriptome of a particular organ) several times. Identifying SNPs from this type of data is relatively easy, provided that a reference sequence has been established (Nielsen *et al.*, 2011), which is the case for most livestock species. Methods have also been developed for SNP discovery in newly sequenced species (Norman *et al.*, 2013) and these approaches may prove useful for less common livestock species.

High-density SNP panels are now widely used for genome-wide association studies (GWAS), genomic prediction and population genomic analyses. However, the preliminary phase, i.e. SNP discovery or SNP selection from databases, is critical. If data have not been obtained randomly, standard estimators of population genetic parameters should be applied with caution. Non-random selection may occur if SNP sets are derived for use on a given set of breeds but later used on other breeds or if SNP sets are filtered to meet certain criteria (e.g. a minimum allele frequency).

Many current tools are affected by both these factors, as they have been developed primarily using widely used international transboundary breeds and with the use of SNP-filtering criteria. Such protocols bias the distribution of allelic frequencies relative to what would be expected in a random sample. The resulting inaccuracy in estimation of genetic parameters is known as "ascertainment bias". Bias caused by problems of this kind is probably present in most commercial and ready-to-use medium- and high-density SNP panels currently available for use in livestock species. Unbiased estimates of the absolute genetic diversity (i.e. the nucleotide diversity) of a population can, in theory, be

Box 4B2
Glossary: genetic markers

Genetic marker: a DNA sequence variation that is informative with respect to a specific location (locus) on a particular chromosome.

Microsatellites: segments of DNA characterized by a variable number of copies (typically 5 to 50) of sequence motifs of around two to five bases (referred to as a repeat unit). At any one locus (site in the genome), there are usually several different "alleles" in a population, each allele identifiable by the number of repeat units detected via polymerase chain reaction (PCR). Many microsatellites have a large degree of polymorphism. In many species, they were the first standard marker technology used to characterize diversity. However, due to their comparatively infrequent presence across the genome, inconsistent reproducibility across laboratories and genotyping platforms, and higher genotyping cost per locus, microsatellites are being replaced by other technologies.

Single-nucleotide polymorphism (SNP): a DNA sequence variation that results from a change in the nucleotide at a single location in the genome. SNPs usually have only two alleles. They may represent either neutral or functional genetic diversity and generally occur throughout the genome. In most species, SNPs occur, on average, once in every 100 to 300 positions in the DNA sequence. For most major livestock species, commercial arrays are available that allow substantial numbers of SNPs (from a few hundreds to over a million) to be genotyped in a single reaction at a low cost per marker. SNP arrays are now routinely used as more informative alternatives to microsatellite panels in genetic diversity studies.

Markers of sex-specific inheritance: certain parts of the genome have sex-specific inheritance. Mitochondrial DNA is passed from the mother to the offspring. The Y-chromosome in mammals is inherited from father to son, while the W-chromosome in birds is inherited from mother to daughter. This class of markers can include both SNPs and other sequence variations and has been instrumental in identifying wild ancestors, localizing domestication centres and reconstructing colonization and trading routes.

Sequence variation: with the emergence of whole-genome sequencing, the entire variation present in the DNA sequence is now available as a potential source of marker information. This variation comprises SNPs and insertions and deletions (InDels) (loss or gain of one or more nucleotides relative to the species reference sequence), duplications, copy number variations (CNVs) (variation in the number of copies of sections of the DNA) or inversions of DNA fragments. Sequencing can be performed on a whole genome basis or only for specific parts of the genome (e.g. the exome or genome fractions used for genotyping by sequencing). Whole-genome sequences will be the ultimate source of genetic diversity information, as they harbour the motifs responsible for genetic differences between breeds. However, efficient management, use and storage of this information will require, in addition to sequencing capacities, substantial development of resources in bioinformatics (methods, tools and hardware).

obtained only via whole genome sequencing. Statistical approaches that explicitly account for the methods used in SNP discovery and sample preparation have been developed for use when undertaking various kinds of population genetics analyses with SNPs (Nielsen *et al.,* 2011; Kofler *et al.,* 2011). Large-scale projects have also started to harvest genome-wide information for use in characterizing livestock populations at national or international scale, including studies on cattle (Gauthier *et al.,* 2010), sheep (Kijas *et al.*, 2012), horses (McCue *et al.* 2012; Orlando *et al.,* 2013), pigs (Groenen *et al.,* 2010), chickens (Weigend *et al.,* 2015) and goats (Dong *et al.,* 2013).

It is important to note that although cost per SNP is low relative to microsatellites (and decreases with the number of SNPs analysed) the costs of high-density assays – currently (2015)

US$50 to US$200 and depending heavily on the number of arrays purchased – are nonetheless prohibitive for many applications. Costs continue to decline, however, and financially realistic options are likely to eventually become available for most situations. This being said, even if lower cost genotyping assays become available, the bioinformatic infrastructure in most developing countries will still require further development. Both the sheer amount of raw data and the complexity of analytical models are several orders of magnitude larger than those associated with microsatellite-based analyses. This is true for work with SNP array data, but even more so for work with sequence data.

Further studies are in the process of identifying millions of SNPs and haplotypes (specific allelic combinations for a given set of loci) and also other sequence variants such as insertion–deletion polymorphisms (InDels) and copy number variants (CNV) (see Box 4B2 for explanations of these terms). Novel sequencing technologies are continuously evolving, accompanied by a drop in cost per sequenced genome (see Figure 4B1). Allele frequency differences and diversity measures derived from them can be obtained inexpensively by sequencing pooled DNA from multiple individuals from a population (e.g. Qanbari *et al*., 2012). Sooner or later, sequence-based approaches will become the standard methodology for generating data for use in livestock diversity studies.

Marker information will become even more useful when linked to biological background information available in specialized databases. Information about marked genes and their functions is available in the Ensembl database[1] (among others) for many livestock species. Information on quantitative trait loci (QTL) is collected in the AnimalQTL database[2] and genomic pathway information is available through KEGG.[3] In human genetics, the Encode project[4] is systematically annotating functional elements in the genome, and similar initiatives are emerging in other species, including livestock (Andersson *et al.*, 2015). On this basis it can, for example, be judged whether observed between-breed diversity in a given genomic region is purely neutral and has been generated by genetic drift or is of functional relevance and may have been caused by selection. Making systematic use of such information will allow a shift from a purely statistical assessment of genetic diversity to a more informative functional approach.

3 Characterization of within-population diversity

Classical estimators of genetic variability (heterozygosity, F-statistics, etc.) are still commonly used. However, some are not adapted for use with biallelic markers (e.g. number of alleles per locus, which is invariably equal to two for biallelic markers). As the use of SNPs has become more common, so has the use of individual and population genomic indicators of diversity and similarity, such as coancestry and inbreeding (Meuwissen and Goddard, 2001; Keller *et al.*, 2011; Saura *et al.*, 2013; Curik *et al.*, 2014). Some of these indicators make it possible to test whether inbreeding effects are more or less important in specific genomic locations, or whether inbreeding comes from a more or less distant common ancestor (e.g. Ferenčaković *et al*., 2013). Estimators of genetic variability can also be used in conservation decision-making as a means of optimizing the choice of breeding animals so as to minimize the loss of genetic variability (Oldenbroek, 2007).

In parallel, several methods of estimating present and past effective population sizes have been developed or improved, based either on the correlation between allele frequencies (linkage disequilibrium) or on runs of homozygosity (Sved, 1971; Hill, 1981; Hayes *et al.*, 2003; Waples, 2006; Li and Durbin, 2011; Hillestad *et al.*, 2014). These approaches have been increasingly applied in livestock, including cattle (de Roos *et al.*, 2008;

[1] http://www.ensembl.org
[2] http://www.animalgenome.org/cgi-bin/QTLdb/index
[3] http://www.genome.jp/kegg
[4] http://www.genome.gov/encode

Flury *et al.*, 2010), sheep (Kijas *et al.*, 2012), pigs (Uimari and Tapio, 2011), chickens (Qanbari *et al.*, 2010) and horses (Corbin *et al.*, 2010). It should be noted, however, that the widely used approach suggested by Sved (1971) has some methodological shortcomings (Sved, 2008) and is especially sensitive to non-random samples of SNPs (Corbin *et al.*, 2012; Ober *et al.*, 2013).

At the time the first SoW-AnGR was prepared, it was generally considered that because of the limited number of markers available it was more efficient to use genealogical information than molecular information in conservation decision-making (Fernandez *et al.*, 2005). This appears no longer to be the case. Commercial SNP arrays are now affordable and provide estimates of genetic relationships that account for the random segregation and recombination of chromosomes that occur during inheritance from parents to offspring. Because marker-based information provides better estimates of genetic relationships than pedigree data, inclusion of genomic data is likely to increase the efficiency of conservation schemes (Hasler *et al.*, 2011; Toro *et al.*, 2014).

4 Characterization of between-population diversity

Relationships between populations have long been assessed through the estimation of genetic distances, which are often used to construct phylogenetic trees to visually infer genetic relationships. However, a major drawback of reconstructing phylogenetic trees is that the evolution of lineages is assumed to be non-reticulate, i.e. it is assumed that while lineages may diverge, they never result from crosses between lineages. There is therefore a tendency for these methods to be replaced by alternative graphical networks or other approaches such as Bayesian clustering methods or multivariate analysis (Bertorelle *et al.*, 2004). One of the most popular model-based Bayesian approaches in current use is the model-based clustering method developed by Pritchard *et al.* (2000) (STRUCTURE software), although alternatives are available (e.g. Alexander *et al.*, 2009). The approach uses Monte Carlo Markov chain simulation to assign individuals to a chosen number of clusters (populations), inferring genetic origins without *a priori* knowledge or assumptions. This is a particularly important consideration in livestock populations, where unsuspected admixture may have occurred. The approach is, however, not without limits. For example, inferred clusters may not always be ancestral, but rather related to highly inbred populations ("inbreeding bias") or to populations over-represented in the dataset ("sampling bias") (Lenstra *et al.*, 2012). Multivariate analysis approaches are interesting alternatives to model-based approaches, as they are generally assumption-free methods and are specifically designed for summarizing large and complex datasets into a small number of synthetic variables (Jombart *et al.*, 2010). These various approaches are usually applied to microsatellite or SNP marker information. They have been extensively used in livestock studies, either independently or (because of the complementary information they may provide) in parallel (Muchadeyi *et al.*, 2007; Leroy *et al.*, 2009; Gautier *et al.*, 2010; Kijas *et al.*, 2012). Methods have been developed over the last few years that use dense haplotype data to unravel fine-scale population structure (Lawson *et al.*, 2012) or apply advanced admixture analysis in order to infer the presence and historical timing of admixture events among human populations (Patterson *et al.*, 2012; Pickrell and Pritchard, 2012; Hellenthal *et al.*, 2014).

Recently, a growing number of methods for combining genomic information with information from other sources, often related to the environments where animals are raised, have been developed (Pariset *et al.*, 2012). Landscape genomics is an approach that aims to use various methods (e.g. estimation of molecular distance, Bayesian and multivariate analyses) to identify environmental factors that shape genetic variability. For example, a study on Vietnamese goats showed that social organization and husbandry practices were as important as geographical distance in shaping genetic structure (Berthouly *et al.*, 2009). The increasing density of markers genotyped may also allow these approaches to be used to identify chromosomal

regions and genes likely to be subject to positive selection linked to the environment. Finally, knowledge of the history of livestock populations has greatly increased in the last ten years, based on the development of new methods and the increasing availability of large sets of markers (see Part 1 Section A). For instance, a recent study on horse breeds (Wallner *et al.*, 2013) showed that the diversity of the paternally inherited Y-chromosome was very low in comparison to that of maternally inherited mitochondrial DNA haplotypes, a finding consistent with the disproportionate use of a limited number of popular stallions over recent centuries. Genome-wide panels of markers also make it possible to discriminate areas of the genome whose variation has increased or decreased though history in relation to specific gene function.

5 Molecular tools for targeting functional variation

Recent advances in genomics have clearly improved our capacity to characterize functional variation in livestock species. Detection and mapping of QTLs, i.e. markers physically linked to a genomic variant that underlies variation in a quantitative trait, have benefited from increased genome coverage, as well as from the development of new methodologies. In particular, the use of sequence data may allow causative polymorphism to be targeted directly instead of via QTLs. The molecular background of various breed-specific traits has been the subject of numerous investigations (Table 4B1) (see also Box 4B3). The molecular analysis of adaptive variation has also improved knowledge of the possible adverse effects of selection on the health and productivity of animals. For example:

- Several gene variants are pathogenic or confer sterility in homozygous animals. For instance, myostatin deficiency caused by mutations in the MSTN (myostatin) gene hinders the delivery of calves (Bellinge *et al.*, 2005).
- Gene variants exhibiting clear antagonism between milk yield and fertility (increasing the former while decreasing the latter) have been identified in cattle (Kadri *et al.*, 2014).
- The use of only a few top sires promotes inbreeding and thus increased homozygosity. This effect, which inevitably increases the proportion of offspring that have recessive genetic defects, can be assessed using neutral genetic markers (Lenstra *et al.*, 2012). Several pathogenic mutations in livestock species, most of which are recessive, have been identified. They surpass in number the gene variants known to be involved in economic traits (Nicholas *et al.*, 2012).
- Significant deficiency or complete absence of individuals homozygous for a given haplotype may indicate the presence of a recessive genetic defect causing early embryonic mortality. This concept has been successfully used in the identification of possible causes of reduced fertility in various cattle populations (Fritz *et al*., 2013).

International consortia have provided large amounts of data on SNPs and other variants. For example, the "1 000 bull genomes project" (Daetwyler *et al.*, 2014) identified 28.3 million variants, related, *inter alia*, to coat colour, embryonic loss and production traits. However, it is still difficult to obtain genome sequences for a large number of animals at an affordable price. Methods have therefore been developed that can be used to "impute" or infer the genotypes of individual animals for which information is sparse (e.g. obtained using low- or medium-density SNP chips) from information on a subset of individuals that have been sequenced (e.g. those studied in the above-mentioned 1 000 bull genomes project) (Jansen *et al.*, 2013).

Although the study of animal genetic diversity has typically concentrated on direct differences in genomes, the impact of genetic diversity on the expression of genes may be relevant, especially as interest grows in functional genetic diversity relative to neutral genetic diversity. Since the mid-1990s, the widespread use of DNA microarrays and serial analysis of gene expression (SAGE), both of which provide a snapshot of actively expressed

Box 4B3

How genetic tools helped to solve the mystery of the origin of the Booroola gene

Exceptionally high litter size in an Australian Merino flock kept at the Booroola Estate in Cooma, New South Wales, attracted the attention of scientists from the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO). Initial analysis of ewes' pedigrees and performance records led to a hypothesis regarding the segregation of a major gene affecting this trait and increasing litter size in carriers. This hypothesis was first substantiated by analysis of litter-size segregation in families (Piper and Bindon, 1982). The origin of the high-fecundity gene in the low-prolific Merino population, however, remained a mystery until Professor Helen Newton Turner found evidence that ancestors of the Booroola flock could have had some admixture of Indian Bengal sheep brought to Australia from Calcutta in the 1790s (Turner, 1983). The hypothetical major gene increasing litter size was named Fecundity Booroola (*FecB*).

The first genetic markers linked to the *FecB* locus were discovered by a New Zealand team led by Professor Grant Montgomery (Montgomery, 1993). Further research led to the conclusion that the Booroola gene is located on the sixth chromosome. The first molecular test, devised to enable the introgression of the *FecB* mutation into the Romney breed, was based on the polymorphism of three microsatellite sequences (Lord *et al.*, 1998).

The real breakthrough with respect to the physiological basis for increased fecundity happened in 2001, when teams from AgResearch (New Zealand), INRA (France) and Edinburgh University (United Kingdom) independently discovered that carriers of the Booroola gene have a mutation in the bone morphogenetic protein receptor IB gene (*BMPR-IB*). The Booroola gene (*FecB*) is a dominant autosomal gene with an additive effect on ovulation rate.

Garole sheep of Bengal

Photo credit: Kanhaiya M. Chavan.

The discovery of the mutation and the development of the molecular test enabled the identification of the mutation in the Garole sheep of Bengal, a breed that is well known for its large litter sizes – thus supporting Professor Turner's theory. At present, the *BMPR-IB* mutation has been found in a number of breeds that have high fecundity. The list includes Javanese Thin Tail sheep (Davis, 2009) and some Chinese breeds such as the Huyang, Small Tail Han (STH), Cele, Duolang and Chinese Merino (Hua and Yang, 2009). It seems that the original mutation took place in Mongolian Fat Tail sheep and was introgressed into Chinese breeds and later into the Indian Garole and Javanese breeds as a result of the movement of people and animals along the Silk Road.

So the mystery was solved thanks to the persistence of scientists and development of technology. Over time, the *FecB* mutation has been introgressed into about 40 breeds, all around the world (Walkden-Brown *et al.*, 2008).

Provided by Elžbieta Martyniuk.

genes and transcripts in a biological sample, has facilitated high-throughput molecular studies of the transcriptome. Microarray experiments provide a cost-effective means of studying the transcriptome, and the bioinformatic and statistical analyses (referred to as "analysis pipelines")

TABLE 4B1

Examples of non-disease phenotypes specific to one or more livestock breeds

Species/breed(s)	Phenotype	Gene or locus	Reference
Cattle			
French	Dairy traits	Several candidate genes	Flori *et al.*, 2009
Several	Dairy and beef traits	Several candidate genes	Rothammer *et al.*, 2013
Danish Red	High milk yield, low fertility	Deletion removing *RNASEH2*#	Kadri *et al.*, 2014
Several	Milk protein content	*ABCG2*+	Braunschweig, 2010
Several	Muscular hypertrophy	*MSTN*#* (different mutations)	Nicholas and Hobbs, 2012; O'Rourke *et al.*, 2013
Holstein	Stature	*PLAG1-CHCHD7* intergenic	Karim *et al.*, 2011
Dexter	Short stature	*ACAN*#	Cavanagh *et al.*, 2007
Dutch Belted Galloway Swiss Brown	Belted pattern	*HES1* (candidate gene)	Drogemuller *et al.*, 2010
Sheep			
Several	Litter size	*GDF9*# (*FecG*, different mutations)	Vage *et al.*, 2013
Several	Litter size	*BMP15*# (*FecX*, different mutations)	Nicholas and Hobbs, 2012
Several	Litter size	*BMPR1B* (Booroola, *FecB*)	Davis *et al.*, 2006
Lacaune	Litter size	B4GALNT2 (*FecL*)	Drouilhet *et al.*, 2013
Texel and others	Muscular hypertrophy	*MSTN* (= *GDF8*)	Clop *et al.*, 2006
Dorset	Muscular hypertrophy	*CLPG**	Braunschweig, 2010
Pig			
European	Muscle growth	*IGF2*	Braunschweig, 2010
Horse			
Quarter	Type I muscle fibres	*MSTN*	Petersen *et al.*, 2013
Several	Endurance	*GYS1*#	McCoy *et al.*, 2014
Chicken			
Several	Naked neck	*BMP12*	Mou *et al.*, 2011
Several	Frizzle feather	*KRT75*	Ng *et al.*, 2012
Several	Silky feather	*PDSS2*	Feng *et al.*, 2014
Several	Comb shape	*MNR2*	Imsland *et al.*, 2012

Note: Several mutations may already have played a role in more general adaptation to domestication (see Part 1, Section A, Table 1A2).
causative gene variant is pathogenic or confers sterility if homozygous; * recent gene mutation; + ATP-binding cassette, sub-family G (WHITE), member 2.
Sources: Braunschweig, 2010; Nicholas and Hobbs, 2012.

used to transform raw microarray data into interpretable results are now well established (Ritchie *et al.*, 2015). Since the time the first SoW-AnGR was prepared, the development of high-throughput sequencing in the field of transcriptomic analysis (RNA sequencing or RNA-seq) has led to radical changes (Mortazavi *et al.*, 2008), primarily because RNA-seq approaches do not necessarily require prior knowledge of a genome sequence or annotation (identification of locations and functions of coding regions within a genome) and can therefore be used even in poorly characterized organisms. In addition, it enables a wide range of novel applications, including detection

of weakly expressed genes and alternative splicing isoforms (variations in the proteins translated from the same gene) (Wang *et al.*, 2008; Pan *et al.*, 2008), variable assembly of transcripts (Trapnell *et al.*, 2010; Guttman *et al.*, 2010; Robertson *et al.*, 2010; Grabherr *et al.*, 2011; Schulz *et al.*, 2012) and allele-specific expression (Skelly *et al.*, 2011).

Recent comparisons have indicated good overall agreement among results obtained using microarrays, quantitative PCR (polymerase chain reaction) and RNA-seq across different sequencing platforms (Zhao *et al.*, 2014; Trapnell *et al.*, 2013; Nookaew *et al.* 2012; Liu *et al.*, 2011). However, although microarrays and RNA-seq are both used to characterize transcriptional activity, the experimental, bioinformatic and analytical steps associated with the two differ considerably (Oshlack *et al.*, 2010). In particular, RNA-seq experiments generate much more data than alternative transcriptomic approaches and require more sophisticated analyses and therefore greater technical capacity in bioinformatics and biostatistics (e.g. Langmead and Salzberg, 2012; Grabherr *et al.*, 2011; Oshlack and Wakefield, 2009; Zhou *et al.*, 2014). The analytical processes of transcriptomics constitute a major area of research in bioinformatics and statistics.

In recent years, studies using RNA-seq to examine genetic variation in gene expression have been undertaken in cattle (Li *et al.*, 2011), chickens (Endale Ahanda *et al.*, 2014; Davis *et al.*, 2015) and pigs (Corominas *et al.* 2013; Fischer *et al.*, 2015). The objectives of these studies have included the identification of candidate genes influencing phenotypic differences and the study of differences in gene expression associated with specific SNPs.

6 The role of bioinformatics

The successful use of high-throughput technologies in the study of genetic diversity is largely contingent on the availability of support and expertise in bioinformatics and statistics. Increasingly large and complex datasets need to be understood, organized, quantified, and analysed. Developing and applying the methods and software tools needed to do this requires appropriate computing resources (including sufficient computational power and memory to store and manipulate large data files) and programming skills. For example, genome sequencing and RNA-seq studies often require the services of a dedicated bioinformatics team to pre-process the data, including raw-data quality control and sequence alignment or assembly, in addition to biostatisticians for eventual data analysis. Bioinformatic support is often also an integral part of the development, maintenance and interrogation of biological databases.

An increasing number of well-documented and open-access bioinformatics and statistical tools are available online. For example, the Bioconductor project[5] is an open-source open-development software project that develops and provides widespread access to a diverse set of well-documented statistical and graphical analysis tools (written in the R programming language) for high-throughput genomic data. In addition, an increasing number of free and publicly accessible resources (e.g. the Galaxy project,[6] an open web-based platform) are available to facilitate bioinformatic analyses without the need for extensive programming knowledge.

It is highly desirable that when researchers gather large-scale genomic data for a given project they make them freely available to other researchers once the initial analyses have been completed. Increasingly, scientific journals and research-funding organizations request that data underlying publications or generated in publicly funded projects be deposited in open repositories. This kind of open-source policy will generate a large quantity and variety of reference data, across species and breeds, that can be used for increasingly comprehensive and informative diversity studies.

[5] http://master.bioconductor.org
[6] http://galaxyproject.org

7 Conclusions and research priorities

The world of genetics has been revolutionized over the last decade with the advent of massive parallel sequencing and high-throughput genotyping technologies. Other technologies and opportunities are on the way (see Box 4B4). These developments have opened many opportunities to utilize molecular techniques in the management of AnGR. However, while these technologies facilitate the sequencing of complete genomes or the genotyping of high-density SNP panels at moderate cost, they have not completely replaced traditional molecular markers such as microsatellites, mainly because of their still relatively high costs and the additional skills needed to analyse the enormous amount of data they produce. Low-cost alternatives, such as low-density SNP panels, that allow genetic variants scattered across the genome to be queried and can feasibly be used in small and medium-sized laboratories are in development, but remain to be implemented in practice.

Understanding of genetic diversity needs to be improved, even in the most widely used livestock species. For example, comprehensive assessments of genetic diversity using molecular genetic markers need to be extended to locally adapted breeds, particularly those with small population sizes. The value of the large quantities of data that currently exist in fragmented form needs to be maximized (e.g. by undertaking meta-analyses and by making as much data as possible publicly available for use by breeders, researchers and policy-makers). Improvements in sequencing and genotyping technologies have already provided standards that can be used as references for further genotyping and sequencing studies. Reference genomes, biological background information and population genotypic data are still not available for some species, but sequencing efforts currently underway in laboratories around the world will soon fill these gaps. For most populations and production systems, taking full advantage of the opportunities that advances in genomics have created for the study of genetic diversity will also require new and additional phenotypic data.

Understanding of the genetic basis of adaptive traits also needs to be improved. Potential approaches include the use of new technologies, such as genome sequencing and geographic

Box 4B4
What are the promises of the post-genomic era?

Over the last twenty years, the use of molecular tools has acquired paramount importance in animal breeding through the development of genetic tests, as well as the implementation of genomic selection in a growing number of species. The role of molecular tools is expected to continue expanding. Potential developments include:

- increased use of whole-genome sequencing for genomic selection, identification of new functional variants (allowing selection on new traits) and analysis of genetic diversity;
- the use of epigenetics (see Box 4B5) in the study of environment × genome interactions to provide insight into complex traits, especially those related to development;
- the use of meta-genomic studies that consider the gut microbiome to enable the optimization of the rumen microbial ecosystem for better feed-conversion efficiency; and
- combining genomics with other advanced biotechnologies, such as *in vitro* embryo transfer (selecting breeding candidates at the embryo stage) and genetic engineering (introducing genes of interest into the genome or even directly editing the genome through novel technologies such the CRISPR/Cas system), which may bring about major changes in the way animals are raised and selected.

Some of these developments would, clearly, raise social and ethical concerns that would need to be addressed before putting them into practice.

Sources: Gonzalez-Recio, 2012; Hayes *et al.*, 2013.

Box 4B5

The reality and promises of epigenetics for animal production

The term "epigenetics" refers to mechanisms that cause variation of gene activity and are based not on variation in the nucleotide sequence but on chemical modifications of DNA and chromatin, for instance, affecting the degree of chromatin compaction or the accessibility of regulatory sequences to transcription factors.

Epigenetic processes occur during cell, tissue and organismal differentiation processes such as gametogenesis, embryo genome activation, X-chromosome inactivation and genomic imprinting (the differential expression of certain genes depending on which parent they come from). Epigenetic modifications driven by environmental factors, such as diet, stress and disease, form a molecular memory (transient or permanent) that adapts the organism and programmes signalling and metabolic pathways appropriate for particular conditions. Epigenetic marks are meiotically and mitotically inherited and the epigenome represents an additional level in genotype–phenotype mapping.

As epigenetic variation explains part of the so-called missing heritability (i.e. the part of genetic variation not explained by individual genes), there may be interest in using epigenetics in livestock breeding, for instance through genomic imprinting and metabolic programming (stimulation of early development to influence later physiological outcome). Particularly in cross-breeding schemes, imprinting offers new opportunities to establish dam and sire lines with enhanced production performances and maternal abilities and to produce cross-breeds according to market requirements. The interaction between genetic (breed) and environmental factors such as diets and management practices is key to the use of epigenetics in animal breeding. Extensive knowledge of epigenetic mechanisms and gene regulation will, in future, offer new opportunities in livestock breeding in relation to environment–genome–epigenome interactions.

Provided by Klaus Wimmers.

information systems, in combination with new data-capture methods (e.g. remote sensing, image analysis and mid-infrared technology) and analytical approaches (e.g. landscape genetics approach), to facilitate the identification of signatures of natural selection reflecting local adaptation to diseases and other environmental conditions. This is of particular importance in the light of climate change. There is a need to develop methods for integrating molecular information into conservation and breeding programmes, and these methods need to be adapted to different environmental, agricultural and socio-economic circumstances. Tackling this task will be a challenge and will require substantial additional data (on genotypes, phenotypes and production environments). Greater international collaboration in data collection, analysis and interpretation will be essential.

References

Abdullah, M.A., Martojo, H., Noor, R.R., & Solihin, D.D. 2012. Genetic characterization of the Aceh cattle using phenotypic, mitochondrial DNA of D-loop region and microsatellite DNA analyses. *Reproduction of Domestic Animals*, 47 (Suppl 1): 15–17.

Achilli, A., Olivieri, A., Soares, P., Lancioni, H., Hooshiar, K.B., Perego, U.A., Nergadze, S.G., Carossa, V., Santagostino M., Capomaccio, S., Felicetti, M., Al-Achkar, W., Penedo, M.C.T., Verini-Supplizi, A., Houshmand, M., Woodward, S.R., Semino, O., Silvestrelli, M., Giolotto, E., Pereira, L., Bandelt, H.J. & Torroni, A. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences*, 109: 2449–2454.

Ajmone-Marsan, P. & GLOBALDIV Consortium. 2010. A global view of livestock biodiversity and conservation-GLOBALDIV. *Animal Genetics,* 41: 1–5

Alexander, D.H., Novembre, J. & Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19: 1655–1664.

Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S.C., Burt, D.W., Casas, E., Cheng, H.H., Clarke, L., Couldrey, C., Dalrymple, B.P., Elsik, C.G., Foissac, S., Giuffra, E., Groenen, M.A., Hayes, B.J., Huang, L.S., Khatib, H., Kijas, J.W., Kim, H., Lunney, J.K., McCarthy, F.M., McEwan, J.C., Moore, S., Nanduri, B., Notredame, C., Palti, Y., Plastow, G.S., Reecy, J.M., Rohrer, G.A., Sarropoulou, E., Schmidt, C.J., Silverstein, J., Tellam, R.L., Tixier-Boichard, M., Tosser-Klopp, G., Tuggle, C.K., Vilkki, J., White, S.N., Zhao, S., Zhou, H. & The FAANG Consortium. 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biology*, 16: 57.

Azam, A., Babar, M.E., Firyal, S., Anjum, A.A., Akhtar, N., Asif, M. & Hussain, T. 2012. DNA typing of Pakistani cattle breeds Tharparkar and Red Sindhi by microsatellite markers. *Molecular Biology Reports,* 39: 845–849.

Bellinge, R.H.S., Liberles, D.A., Iaschi, S.P.A., O'Brien, P.A. & Tay, G.K. 2005. *Myostatin* and its implications on animal breeding: a review. *Animal Genetics*, 36: 1–6.

Berthouly, C., Do Ngoc, D., Thévenon, S., Bouchel, D., Nhu Van, T., Danes, C., Grosbois, V., Hoang Thanh, H., Vu Chi, C. & Maillard, J.-C. 2009. How does farmer connectivity influence livestock genetic structure? A case-study in a Vietnamese goat population. *Molecular Ecology*, 18: 3980–3991.

Bertorelle, G., Bruford, M., Chemini, C., Cristiano, V. & Hauffe, H.C. 2004. New, flexible Bayesian approaches to revolutionize conservation genetics. *Conservation Biology*, 18: 1–2.

Braunschweig, M.H. 2010. Mutations in the bovine ABCG2 and the ovine MSTN gene added to the few quantitative trait nucleotides identified in farm animals: A mini-review. *Journal of Applied Genetics*, 51: 289–297.

Cavanagh, J.A., Tammen, I., Windsor, P.A., Bateman, J.F., Savarirayan, R., Nicholas, F.W. & Raadsma, H.W. 2007. Bulldog dwarfism in Dexter cattle is caused by mutations in ACAN. *Mammalian Genome*, 18: 808–814.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. & Georges, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38: 813–818.

Corbin, L.J., Blott, S.C., Swinburne, J.E., Vaudin, M., Bishop, S.C. & Woolliams, J.A. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics*, 41: 8–15.

Corbin, L.J., Liu, A.Y., Bishop, S.C. & Woolliams, J.A. 2012. Estimation of historical effective population size using linkage disequilibria with marker data. *Journal of Animal Breeding and Genetics*, 129: 257–270.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castello, A., Alves, E., Pena, R.N., Ballester, M. & Folch, J.M. 2013. Analysis of porcine adipose tissue transcriptome reveals differences in *de novo* fatty acid synthesis in pigs with divergent muscle fatty acid composition. *BMC Genetics*, 14: 843.

Curik, I., Ferenčaković, M. & Sölkner, J. 2014 Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science*, 168: 26–34.

D'Alessandro, A. & Zola, L. 2012. Meat science: From proteomics to integrated omics towards system biology. *Journal of Proteomics*, 78: 558–577.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Grohs, C., Jung, S., Baur, A., Esquerré, D., Gollnick, N., Legrain, S., Laurent, P., Djari, A., Rossignol, M.N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P., Coote, D., Chamberlain, A., VanTassel, C.P., Hulsegge, I., Goddard, M., Guldbrandtsen, B., Lund, M.S., Veerkamp, R., Boichard, D., Fries, R. & Hayes, B.J. 2014. Whole genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46: 858–865.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q. & Catchen, J.M. 2011. Genome-wide genetic marker discovery and genotyping using next generation sequencing. *Nature Reviews Genetics*, 12: 499–510.

PART **4**

Davis, G.H. 2009. The Booroola gene: origin, distribution, use and management of the Fec B mutation. *In* S.W. Walkden-Brown, J.H.J. van der Werf, C. Nimbkar & V.S. Gupta, eds. *Use of the FecB (Booroola) gene in sheep-breeding programs. Proceedings of the Helen Newton Turner Memorial International Workshop held in Pune, Maharashtra, India, 10–12* November *2008*. ACIAR Proceedings No. 133. Canberra, Australian Centre for International Agricultural Research (available at http://aciar.gov.au/files/node/11617/ACIAR_PR133(online).pdf).

Davis, G.H., Balakrishnan, L., Ross, I.K., Wilson, T., Galloway, S.M., Lumsden, B.M., Hanrahan, J.P., Mullen, M., Mao, X.Z., Wang, G.L., Zhao, Z.S., Robinson, J.J., Mavrogenis, A.P., Papachristoforou, C., Peter, C., Baumung, R., Cardyn, P., Boujenane, I., Cockett, N.E., Eythorsdottir, E., Arranz, J.J. & Notter, D.R. 2006. Investigation of the booroola (*FecB*) and inverdale (*FecX(I)*) mutations in 21 prolific breeds and strains of sheep sampled in 13 countries. *Animal Reproduction Science*, 92: 87–96.

Davis, R.V., Lamont, S.J., Rothschild, M.F., Persia, M.E., Ashwell, C.M. & Schmidt, C.J. 2015. Transcriptome analysis of post-hatch breast muscle in legacy and modern broiler chickens reveals enrichment of several regulators of myogenic growth. *PLoS ONE*, 10: e0122525.

De Donato, M., Peters, S.O., Mitchell, S.E., Hussain, T. & Imumorin, I.G. 2013. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE,* 8: e62137.

De Roos, A.P.W., Hayes, B.J., Spelman, R.J. & Goddard, M.E. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics*, 179: 1503–1512.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, X.J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J. & Wang, W. 2013. Sequencing and automated whole genome optical mapping of the genome of the domestic goat (*Capra hircus*). *Nature Biotechnology*, 31: 135–141.

Doro, M.G., Piras, D., Leoni, G.G., Casu, G., Vaccargiu, S., Parracciani, D., Naitana, S., Pirastu, M. & Novelletto, A. 2014. Phylogeny and patterns of diversity of goat mtDNA haplogroup reveals by resequencing complete mitogenomes. *PLoS ONE*, 9: e95969.

Drogemuller, C., Demmel, S., Engensteiner, M., Rieder, S. & Leeb, T. 2010. A shared 336 kb haplotype associated with the belt pattern in three divergent cattle breeds. *Animal Genetics,* 41: 304–307.

Drouilhet, L., Mansanet, C., Sarry, J., Tabet, K., Bardou, P., Woloszyn, F., Lluch, J., Harichaux, G., Viguie, C., Monniaux, D., Bodin, L., Mulsant, P. & Fabre, S. 2013. The highly prolific phenotype of Lacaune sheep is associated with an ectopic expression of the B4GALNT2 gene within the ovary. *PLoS Genetics*, 9: e1003809.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489: 57–74.

Endale Ahanda, M-L., Zerjal, T., Dhorne-Pollet, S., Rau A., Cooksey, A. & Giuffra, E. 2014. Impact of the genetic background on the composition of the chicken plasma miRNome in response to a stress. *PLoS ONE*, 9: e114598.

FAO. 2007. *The State of the World's Animal Genetic Resources for Food and Agriculture*, edited by B. Rischkowsky & D. Pilling. Rome (available at http://www.fao.org/docrep/010/a1250e/a1250e00.htm).

FAO. 2011. *Molecular genetic characterization of animal genetic resources.* FAO Animal Production and Health Guidelines. No. 9. Rome (available at http://www.fao.org/docrep/014/i2413e/i2413e00.htm).

Feng, C., Gao, Y., Dorshorst, B., Song, C., Gu, X., Li, Q., Li, J., Liu, T., Rubin, C.J., Zhao, Y.,Wang, Y., Fei, J., Li, H., Chen, K., Qu, H., Shu, D., Ashwell, C., Da, Y., Andersson, L., Hu, X. & Li, N. 2014. A cis-regulatory mutation of PDSS2 causes silky-feather in chickens. *PLoS Genetics*, 10: e1004576.

Ferenčaković, M., Sölkner, J. & Curik, I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution*, 45: 42.

Fernandez, J., Villanueva, B., Pong-Wong, R. & Toro, M.A. 2005. Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics*, 170: 1313–1321.

Fischer, D., Laiho, A., Gyenesei, A. & Sironen, A. 2015. Identification of reproduction related gene polymorphisms using whole transcriptome sequencing in the large white pig population. *Genes Genomes Genetics*, 5: [Epub ahead of print].

Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. & Gautier, M. 2009. The genome response to artificial selection: A case study in dairy cattle. *PLoS ONE*, 4: e6595.

Flury, C., Tapio, M., Sonstegard, T., Drogemuller, C., Leeb, T., Simianer, H., Hanotte, O. & Rieder, S. 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics,* 127: 339–347.

Fritz, S., Capitan, A., Djari, A., Rodriguez, S.C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerre, D., Klopp, C., Rocha, D. & Boichard, D. 2013. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS ONE,* 8: e65550.

Gautier, M., Laloë, D. & Moazami-Goudarzi, K. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS ONE*, 5(9): e13038.

Ginja, C., Gama, L.T., Cortes, O., Delgado, J.V., Dunner, S., García, D., Landi, V., Martin-Burriel, I., Martinez-Martinez, A., Penedo, M.C.T., Rodellar, C., Zaragoza, P., Canon, J. & Biobovis Consortium. 2013. Analysis of conservation priorities of Iberoamerican cattle based on autosomal microsatellite markers. *Genetics Selection Evolution*, 45: 35.

Gonzalez-Recio, O. 2012 Epigenetics: a new challenge in the post-genomic era of livestock. *Frontiers in Genetics*, 2: 106.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, 29: 644–652.

Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S. *et al.* 2012. Analysis of pig genomes provide insight into porcine demography and evolution. *Nature*, 491: 393–398.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., Rinn, J.L., Lander, E.S. & Regev, A. 2010. *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnology,* 28: 503–10.

Hasler, H., Flury, C., Menet, S., Haase, B., Leeb, T., Simianer, H., Poncet, P.A. & Rieder, S. 2011. Genetic diversity in an indigenous horse breed: implications for mating strategies and the control of future inbreeding. *Journal of Animal Breeding and Genetics,*128: 394–406.

Hayes, B.J., Lewin, H.A. & Goddard, M.E. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics*, 29: 206–14.

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13: 635–643.

Hellenthal, G., Busby, G.B., Band, G., Wilson, J.F., Capelli, C., Falush, D. & Myers, S. 2014. A genetic atlas of human admixture history. *Science*, 343: 747–751.

Hill, W.G. 1981 Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, 38: 209–216.

Hillestad, B., Woolliams, J.A., Meuwissen, T., Våge, D.I. & Klemetsdal, G. 2014. Estimating rate of inbreeding and effective population size using genomic data in Norwegian Red Cattle. In *Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, Canada, August 17–27, 2014* (available at http://tinyurl.com/ngeyujq).

Hua, G-H. & Yandg, L-G. 2009. A review of research progress of *FecB* gene in Chinese breeds of sheep. *Animal Reproduction Science*, 116, 1–9.

Imsland, F., Feng, C., Boije, H., Bed'hom, B., Fillon, V., Dorshorst, B., Rubin, C.J., Liu, R., Gao, Y., Gu, X., Wang, Y., Gourichon, D., Zody, M.C., Zecchin, W., Vieaud, A., Tixier-Boichard, M., Hu, X., Hallböök, F., Li, N. & Andersson, L. 2012. The rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genetics,* 8: e1002775.

Jansen, S., Aigner, B., Pausch, H., Wysocki, M., Eck, S., Benet-Pagès, A., Graf, E., Wieland, T., Strom, T.M., Meitinger, T. & Fries, R. 2013. Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. *BMC Genomics*, 14: 446.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11: 94.

Kadri, N.K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Nielsen, U.S., Panitz, F., Aamand, G.P., Schulman, N., Georges, M., Vikki, J., Lund, M.S. & Druet, T. 2014. A 660-kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: Additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genetics*, 10: e1004049.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J.A., Baurain, D., Cambisano, N., Davis, S.R., Farnir, F., Grisart, B., Harris, B.L., Keehan, M.D., Littlejohn, M.D., Spelman, R.J., Georges, M. & Coppieters, W. 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics*, 43: 405–413.

Keller, M.C., Visscher, P.M. & Goddard, M.E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189: 237–249.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. & International Sheep Genomics Consortium Members. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, 10: e1001258.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C. & Schlötterer, C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE,* 6: e15925

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9: 357–359.

Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics*, 8:e1002453.

Lenstra, J.A., Groeneveld, L.F., Eding, H., Kantanen, J., Williams, J.L., Taberlet, P., Nicolazzi, E.L., Sölkner, J., Simianer, H., Ciani, E., Garcia, J.F., Bruford, M.W., Ajmone-Marsan, P. & Weigend, S. 2012. Molecular tools and analytical approaches for the characterization of farm animal diversity. *Animal Genetics*, 43: 483–502.

Leroy, G., Callèle L., Verrier E., Mériaux J.C., Ricard, A. & Rognon, X. 2009. Genetic of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genetics Selection Evolution*, 41, 5.

Li, H. & Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475: 493–496.

Li, R., Rinaldi, M. & Capuco, A. 2011. Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. *Veterinary Research*, 42: 114.

Liu, S., Lin, L., Jiang, P., Wang, D. & Xing, Y. 2011. A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Research*, 39: 578–588.

Lord, E.A., Davis, G.H., Dodds, K.G., Henry, H.M., Lumsden, J.M. & Montgomery, G.W. 1998. Identification of Booroola carriers using microsatellite markers. In *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia, Volume 27*, pp. 19–22.

McCoy, A.M., Schaefer, R., Petersen, J.L., Morrell, P.L., Slamka, M.A., Mickelson, J.R., Valberg, S.J. & McCue, M.E. 2014. Evidence of positive selection for a glycogen synthase (GYS1) mutation in domestic horse populations. *Journal of Heredity*, 105: 163–172.

Medugorac, I., Veit-Kensch, C.E., Ramljak, J., Brka, M., Marković, B., Stojanovic, S., Bytyqi, H., Kochoski, L., Kume, K., Grünenfelder, H.P., Bennewitz, J. & Förster, M. 2011. Conservation priorities of genetic diversity in domesticated meta-populations: a study in taurine cattle breeds. *Ecology and Evolution,* 1: 408–420.

Metzker, M.L. 2010. Sequencing technologies – the next generation. *Nature Reviews Genetics,* 11: 31–46.

Meuwissen, T.H.E. & Goddard, M.E. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution*, 33: 605–634.

Montgomery, G.W., Crawford, A.M., Penty, J.M., Dodds, K.G., Ede, A.J., Henry, H.M., Pierson, C.A., Lord, E.A., Galloway, S.A., Schmack, A.E., Sise, J.A., Swarbick, P.A., Hanrahan, V., Buchanan, F.C. & Hill, D.F. 1993. The ovine Booroola fecundity gene (*FecB*) is linked to markers from a region of human chromosome 4q. *Nature Genetics*, 4: 410–414.

Mou, C., Pitel, F., Gourichon, D., Vignoles, F., Tzika, A., Tato, P., Yu, L., Burt, D.W., Bed'hom, B., Tixier-Boichard, M., Painter, K.J. & Headon, D.J. 2011. Cryptic patterning of avian skin confers a developmental facility for loss of neck feathering. *PLoS Biology*, 9: e1001028.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods*, 5: 621–628.

Muchadeyi, F.C., Eding, H., Wollny, C.B.A., Groeneveld, E., Makuza, S.M., Shamseldin, R., Simianer, H. & Weigend, S. 2007. Absence of population substructuring in the Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Animal Genetics*, 38: 332–339.

Ng, C.S., Wu, P., Foley, J., Foley, A., McDonald, M.L., Juan, W.T., Huang, C.J., Lai, Y.T., Lo, W.S., Chen, C.F., Leal, S.M., Zhang, H., Widelitz, R.B., Patel, P.I., Li, W.H. & Chuong, C.M. 2012. The chicken frizzle feather is due to an alpha-keratin (KRT75) mutation that causes a defective rachis. *PLoS Genetics*, 8: e1002748.

Nicholas, F.W. & Hobbs, M. 2012. Mutation discovery for mendelian traits in non-laboratory animals: a review of achievements up to 2012. *Animal Genetics,* 45: 157–170.

Nielsen, R., Paul, J.S., Albrechtsen, A. & Song, Y.S. 2011. Genotype and SNP calling from next generation sequencing data. *Nature Reviews Genetics*, 12: 433–451.

Nookaew, I., Papini, M., Pornputtpong, N., Scalcinati, G., Fagerberg, L., Uhlén, M. & Nielsen, J. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 40: 10084–10097.

Norman, A.J., Street, N.R. & Spong, G. 2013. *De novo* SNP discovery in the Scandinavian brown bear (*Ursus arctos*). *PLoS ONE*, 8:e81012.

Ober, U., Malinowski, A., Schlather, M. & Simianer, H. 2013. The expected linkage disequilibrium in a finite population revisited. *arXiv*:1304.4856v2 [q-bio.PE].

Oldenbroek, K. 2007. *Utilisation and conservation of farm animal genetic resources*. Wageningen, the Netherlands, Wageningen Academic Publishers.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P.L.F., Fumagalli, M., Vilstrup, J.T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C.D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A.M.V., Cahill, J., Rasmussen, M., Wang, X. *et al.* 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499: 74–78.

PART **4**

O'Rourke, B.A., Greenwood, P.L., Arthur, P.F. & Goddard, M.E. 2013. Inferring the recent ancestry of myostatin alleles affecting muscle mass in cattle. *Animal Genetics,* 44: 86–90.

Oshlack, A., Robinson, M.D. & Young, M.D. 2010. From RNA-seq reads to differential expression results. *Genome Biology*, 11: 220.

Oshlack, A. & Wakefield, M.J. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, 4: 14.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40: 1413–1415.

Patterson, N.J., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschorek, T., Webster, T. & Reich, T. 2012. Ancient admixture in human history. *Genetics*, 192: 1065–1093.

Pariset, L., Joost, S., Gargani, M. & Valentini, A. 2012. Landscape genomics in livestock. *In* M. Çalişkan, ed. *Analysis of genetic variation in animals*, pp. 189–216. Rijeka, Croatia, INTECH.

Petersen, J.L., Mickelson, J.R., Rendahl, A.K., Valberg, S.J., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama P., da Câmara Machado, A., Capomaccio, S., Cappelli, K., Cothran, E.G., Distl, O., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C., Piercy, R.J., Raekallio, M., Rieder, S., Røed, K.H., Swinburne, J., Tozaki, T., Vaudin, M., Wade, C.M. & McCue, M.E. 2013. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genetics*, 9: e1003211.

Pickrell, J.K. & Pritchard, J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8: e1002967.

Piper, L.R. & Bindon, B.M. 1982. Genetic segregation for fecundity in Booroola merino sheep. *In* Burton & W.C. Smith, eds. *Proceedings of the 1st World Congress on Sheep and Beef Cattle Breeding, 1980. Volume 1*, pp. 395–400. Palmerston North, New Zealand, Dunmore Press.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–59.

Qanbari, S., Hansen, M., Weigend, S., Preisinger, R. & Simianer, H. 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC Genetics*, 11:103.

Qanbari, S., Strom, T.M., Haberer, G., Weigend, S., Gheyas, A.A., Turner, F., Burt, D.W., Preisinger, R., Gianola, D. & Simianer, H. 2012. A high resolution genome-wide scan for significant selective sweeps: An application to pooled sequence data in laying chickens. *PLoS ONE* 7: e49525.

Ritchie, M.E., Phipson, B., Wu, D., Law, C.W., Shi, W. & Smyth, G.K. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *doi:*10.1093/nar/gkv007.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A-L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J., Hoodless, P.A. & Birol, I. 2010. *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, 7: 909–912.

Rothammer, S., Seichter, D., Forster, M. & Medugorac, I. 2013. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics,* 14: 908-2164-14-908.

Saura, M., Fernández, A., Rodríguez, M.C., Toro, M.A., Barragán, C., Fernández, A.I. & Villanueva, B. 2013. Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. *PLoS ONE*, 8(10): e78314.

Schulz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28: 1086–1092.

Skelly, D.A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J.M. 2011. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research*, 21: 1728–1737.

Sved, J.A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology,* 2: 125–141.

Sved, J.A. 2008. Linkage disequilibrium and its expectation in human populations. *Twin Research and Human Genetics,* 12: 35–43.

Toro, M., Villanueva, B. & Fernandez, J. 2014. Genomics applied to management strategies in conservation programmes. *Livestock Science*, 166: 48–53.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. & Pachter, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31: 46–53.

Trapnell, C., Williams, BA., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. & Pachter, L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28: 511–515.

Turner, H.N. 1983. Origins of the CSIRO Booroola. *International Journal of Sheep and Wool Science*, 31(1): 10–13.

Uimari, P. & Tapio, M. 2011. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science*, 89: 609–614.

Vage, D.I., Husdal, M., Kent, M.P., Klemetsdal, G. & Boman, I.A. 2013. A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. *BMC Genetics,* 14: 1-2156-14-1.

Walkden-Brown, S.W., van der Werf, J.H.J., Nimbkar, C. & Gupta, V.S. (eds.). 2008. *Use of the* FecB *(Booroola) gene in sheep-breeding programs. Proceedings of the Helen Newton Turner Memorial International Workshop held in Pune, Maharashtra, India, 10–12 November 2008*. ACIAR Proceedings No. 133. Canberra, Australian Centre for International Agricultural Research (available at http://aciar.gov.au/files/node/11617/ACIAR_PR133(online).pdf).

Wallner, B., Vogl, C., Shukla, P., Burgstaller, J.P., Druml, T. & Brem, G. 2013. Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLoS ONE*, 8: e60015.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. & Burge, C.B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456: 470–476.

Waples, R. 2006 A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, 7: 167–184.

Weigend, S., Janssen-Tapken, U., Erbe, M., Ober, U., Weigend, A., Preisinger, R. & Simianer, H. 2015. Potentials of biodiversity in chickens. *Züchtungskunde,* 86: 25–41.

Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K. & Liu, X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T Cells. *PLoS ONE,* 9: e78644.

Zhou, X., Lindsay H. & Robinson, M.D. 2014. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*, 42: e91.