Recent Developments in the Genetic Characterization of Animal Genetic Resources

P.J. Boettcher[*], P. Ajmone-Marsan[†] and J.A. Lenstra[‡]

Introduction

Characterization is one of the first activities to be undertaken in the development of a plan for the management of animal genetic resources (AnGR) and thus has a prominent role in Strategic Priority 1 – Characterization, Inventory and Monitoring of Trends and Associated Risks – of the Global Plan of Action for AnGR (FAO, 2007a). The goal of Strategic Priority Area 1 includes improving characterization, which is the understanding of the characteristics of all aspects and components of AnGR to facilitate and enable decision-making. In addition to the evaluation of phenotypes and production systems, characterization involves collection and molecular analysis of biological samples in order to yield insight into breed history and to guide breed development, utilization and conservation decisions. This process can be broadly described as "genetic" characterization and therefore contributes to the achievement of Strategic Priority 1 of the Global Plan of Action. Biotechnological advances have continually increased the amount and precision of information that can be obtained via genetic characterization. Here we summarize the recent developments in genetic characterization of AnGR and describe future perspectives.

Domestication and breed history

The present diversity of livestock species is the result of a combination of various processes, including domestication, migration, genetic isolation, environmental adaptation, selective breeding, introgression and admixture of subpopulations. Molecular characterization can help unravel the genetic history of a species, which is most relevant for managing the present and future genetic diversity. Different categories of genetic markers are informative for different aspects of a species' genetic history (Groeneveld et al., 2010). The maternally transmitted mitochondrial DNA (mtDNA) allows the identification of the ancestral wild species and to trace back the origin of individuals and breeds through the analysis of phylo-geographic patterns of variation of populations at the continental scale. Y-chromosomal variation shows male introgression of other breeds or related wild species. Autosomal variation reveals patterns of diversity within or across continents. By combining these elements, one can obtain a detailed picture of the genetic processes that accompanied the spread of livestock across the globe.

The first livestock were probably the goat, believed to have been domesticated in the Fertile Crescent 10,000 years ago (FAO 2007b, Driscoll et al., 2009). About 5500 years later the horse was the last of the major livestock species to be domesticated. In addition to the goat,

[*] Animal Genetics Resources Branch, FAO-AGA, 00153 Rome, Italy
[†] Istituto di Zootecnica, Università Cattolica del S. Cuore di Piacenza, 29100 Piacenza, Italy
[‡] Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands

the Fertile Crescent was also the cradle of domestication for sheep, taurine cattle and pigs (Driscoll et al., 2009). Other species came from the Arabian Peninsula (dromedary), South Asia (zebu, river buffalo), China (swamp buffalo, yak, duck), Central Asia (horse), Southeast Asia (chicken, gayal, Bali cattle), East Africa (donkey), Central America (turkey) and South America (llama, alpaca). Many of these species have now a world-wide distribution. The entry of sheep, goat, cattle and pigs marked in Europe the start of the Neolithic.

For all species, sequence analysis of mtDNA identified the wild ancestor (Bruford et al., 2003; Driscoll et al., 2009). In addition, and often in combination with Y-chromosomal haplotyping and microsatellite genotyping, mtDNA has revealed several cases of species hybridization. For instance, the dispersal of zebu from Asia to other continents was found to have occurred by exclusively (Africa) or almost exclusively (America) male introgression (Bradley, 1998; Meirelles et al., 1999). The South-American Criollo cattle combines a Spanish descent with zebu paternal introgression (Ginja et al., 2010). The Chinese yellow cattle forms a continuous gradient from completely taurine in the North to almost completely zebu in the South (Lai et al., 2006). Introgression of yak into cattle (Yu et al. 1999; Kikkawa, 2003) and vice versa (Qi, 2010) have been observed in South-Asia, while Indonesian zebu breeds show varying degrees of banteng (*Bos javanicus*) introgression (Mohamad et al., 2009).

In European pigs, ancient mtDNA in fossils showed that the domestic immigrants of Southwest-Asian origin were soon replaced by domesticates of European origin (Larson et al., 2007). However, many modern European pigs have Asian mtDNA as the result of crossing Chinese pigs with English breeds during the 19th century. Cattle migrating from the Fertile Crescent also encountered their wild relatives in Europe, but ancient mtDNA indicates that introgression from wild aurochs has been sporadic (Stock et al., 2009; Achilli et al., 2009), at least from the female side. However, male introgression has so far not been excluded (Bollongino et al., 2008).

The Neolithic immigration of livestock in Europe may be considered as a series of successive founder effects when new territory was populated by animals from neighboring regions. This observation explains the cline in microsatellite diversity from southeast to northwest in cattle (Cymbron et al., 2005; European Genetic Cattle Diversity Consortium, 2006), sheep (Peter et al., 2007) and goats (Cañón et al., 2006) and in cattle mtDNA (Troy et al., 2001).

Development of agriculture in various environments has, via adaptation, selection and isolation by distance, led to a marked differentiation of both appearance and production traits, most notably for cattle, sheep and goats. This development process underlies the present diversity of livestock and has since the 18th century been accentuated by the formation of isolated breeds with well-defined breeding characteristics (FAO 2007b). Initially, breeds emerged as landraces and were, as often indicated by their name, indigenous to a region. However, isolation was not always strict and both the establishment and further development of a breed often involved crossing with sires from other breeds (Lenstra, 2009). Occasionally, discrepancies occur between the common perception of the origin of a breed and its actual genetic constitution as revealed by molecular studies. For instance, Ayrshire cattle imported in the Nordic countries in the 19th century have been crossed into breeds still considered to be of local origin. We expect that detailed molecular studies (e.g., see Decker et al., 2009) will eventually lead to an accurate decomposition of the history of breeds and provide a comprehensive description of the diversity of livestock.

Sustainable use and development

Genetic markers have been discussed and promoted for many years as a tool for increasing the accuracy of selection and thus enhancing the sustainable use and development of AnGR (e.g. Hansen et al., 1967). Nevertheless, for a variety of reasons, their commercial applicability has been relatively limited until recently (Dekkers, 2004). Due to limitations of the technologies available, genetic characterization for sustainable use and development was mainly focused on analysis of candidate genes or linkage analysis of quantitative trait loci (QTL). The candidate gene approach has successfully identified genes responsible for several simple genetic defects (e.g. Schuster et al., 1992) and has revealed variation in genes directly relevant for production traits such the casein complex in dairy species (Caroli et al., 2009). Linkage analyses consisting of genome scans with hundreds of markers have identified numerous QTL for many traits in all livestock species. However, these genome scans relied almost exclusively on anonymous markers such as microsatellites and the resulting QTL were typically localized within confidence intervals of up to tens of centiMorgans (cM), precluding a fast identification of the causative mutation and widespread use for selection across breeds or even across families within the same breed.

Over the last decade, the costs for DNA sequencing have decreased dramatically. Sequencing technology initially allowed the characterization of specific genomic regions, but entire genomes are now available for several livestock species. In concert with comparative genomics and other advances in bioinformatics, advances in sequencing have also resulted in the discovery of millions of single nucleotide polymorphisms (SNP) in the major livestock species (see www.ncbi.nlm.nih.gov/snp/). The increased quantity of genomic information considerably facilitates candidate gene analysis and makes genomic scans more accurate. SNP analysis allows the fine mapping of QTL regions previously localized by microsatellite genome scans. Candidate genes can be chosen based on both their biological function and their genomic position (e.g. Rincón et al., 2009; Schulman et al., 2009). Monogenic traits can now be localized with high accuracy by descent mapping (e.g. Becker et al., 2010). Although this approach has identified various SNP with statistically significant associations with multigenic traits, verifying that the SNP are causative mutations is not trivial and only a few successes have been described (e.g. Grisart et al., 2004).

For genome scanning, high-density SNP chips have been developed for most of the major livestock species (cattle, pig, chicken and sheep). These chips allow for the simultaneous typing of tens and even hundreds of thousands of SNP. The distribution and density of the SNP and the linkage disequilibrium within the populations is expected to reveal statistical associations with phenotypes even if the SNP on the chip are not causative mutations. This approach has successfully revealed genomic regions influenced by selective breeding.

Selection, whether natural or artificial, changes the allelic frequency of not only the causative mutation, but also of SNP located nearby on the same chromosome. This process creates "selection signatures" within these genomic regions, which can be detected statistically as outliers among the genome-wide distribution of SNP allelic frequencies. Selection signatures have been identified both across entire species (e.g. Bovine HapMap Consortium, 2009) and between breeds selected for opposing phenotypes (e.g. MacEachern et al., 2009; Stella et al., 2010). By performing bioinformatics analyses of the genomic regions identified, the genes responsible for the effects observed can be proposed.

The development of these high-density SNP assays has also made genome-wide selection (Meuwissen et al., 2001) a reality. For "genomic selection", predictive equations are

developed relating SNP genotypes to phenotypes for production traits. Then, breeding values for other animals can be accurately predicted without the need for phenotypic data. These equations generally ignore any biological role that the SNP may have. Genomic selection has great potential to increase genetic response to selection and/or decrease the cost of genetic evaluation. For example, progeny testing can be decreased by orders of magnitude. Several countries are already applying some form of genomic selection for dairy cattle (Hayes et al., 2009).

Genomic and bioinformatics advances have also created opportunities to characterize livestock in terms of the function of their genes. In particular, microarrays consisting of cDNA strands or oligonucleotides allow a comprehensive measurement of gene expression. Quantitative PCR has been employed to verify results from microarray analyses. Studies have been undertaken to compare gene expression of breeds (e.g. MacKinnon et al., 2009) or selected lines (e.g. McCarthy et al., 2009) that are widely divergent for a given phenotype. Measures of gene expression are phenotypes, so their study could be regarded as a convergence of phenotypic and genetic characterization.

Conservation

Genetic diversity is a prerequisite for genetic improvement and environmental adaptation. Diversity of AnGR has been in a continual state of decline (FAO, 2007b). Because conservation of all AnGR is not feasible, conservation priorities are imperative. Obviously, conservation is on the level of breeds and genetic characterization of breeds will be most essential (Boettcher et al., 2010). Weitzman (1992, 1993) proposed the use of molecular markers to obtain a measure of diversity and to combine this with estimated risk of extinction to yield the "conservation potential" for each population under consideration. Many researchers have used or modified this approach in order to prioritize livestock breeds, at least for research purposes (e.g. Laval et al., 2000; Simianer et al., 2003). However, Weitzman (1992, 1993) developed his approach for wild species, and based diversity on genetic distances. This measure tends to emphasize distinctiveness, without consideration of within-breed diversity, and gives undue priority to small populations that diverged mainly by genetic drift. This approach has thus been criticized as inappropriate for within-species prioritization (e.g. Caballero and Toro, 2002; Eding et al., 2002). Eding et al., (2002) developed a method of prioritization that uses molecular markers to minimize kinship within a selected "core set" of breeds. Considering total diversity as a combination of across- and within-breed variability, the approaches of Weitzman (1992) and Eding et al. (2002) are polar opposites, with the former considering only across-breed diversity and the latter only within-population kinship. Piyasatian and Kinghorn (2003) and Bennewitz and Meuwissen (2005), presented methods that consider both across-breed and within-breed diversity, with proportional weights on within-breed variability of 0.2 and 0.5, respectively. The merits of these approaches have been the subject of considerable discussion, with the Weitzman (1992) diversity measure being generally dismissed for livestock. Meuwissen (2009) recently presented an elegant argument favouring the intermediate methods (i.e. Piyasatian and Kinghorn, 2003 and Bennewitz and Meuwissen, 2005).

Another application of molecular genetic analysis is the selection of individuals in genetic conservation programmes. For *in vitro* conservation, the approach taken will depend on the goal of the gene bank. For maximizing diversity in a future reconstituted population, the

approach of Eding et al. (2002) can be modified to identify a group of individuals within a breed with the minimum kinship, rather than a group of breeds within a population. Depending on how the results are interpreted and used, this approach could effectively determine the relative quantities of germplasm to be cryoconserved from each animal. For conservation of specific alleles, animals can be selected by analysis of single genes. For example, if a breeding programme includes the elimination of a given allele (e.g. causing a genetic defect), germplasm from animals carrying that allele may be conserved, so that allele (or latent alleles in linkage disequilibrium) can be available in the future, if needed. Animals in an *in vivo* conservation programme, may also be selected on the basis of molecular characterization. Saura et al. (2008) showed how molecular markers can be used in order to maintain allele frequencies and thus the diversity of the original population.

To date, most of the work relating to characterization and conservation has been done with microsatellites. However, SNP are rapidly becoming the marker of choice and will also be used for conservation purposes. The methods listed above for incorporating molecular diversity into conservation decisions are generally as amenable to SNPs as microsatellites. High-density SNP chips will theoretically improve the accuracy of those methods (Meuwissen, 2009). They also offer the opportunity to target specific genomic regions and can provide information on selected as well as neutral variation (Toro, 2006). One potential problem, however, is ascertainment bias. Most of the SNP on current high density chips have been identified and selected on the basis of a high minor allele frequency in cosmopolitan populations, such as the Holstein cow and commercial chicken lines. This will bias measures of breed-wise genetic variability and underestimate diversity in local breeds (The Bovine HapMap Consortium, 2009). Therefore, the proper use of SNP for conservation will require the development of panels that are free from ascertainment bias by selection of SNP in breeds that represent the unselected basis population.

Application of the current methods for incorporating molecular information into conservation decisions has so far remained on the level of research (Boettcher et al., 2010). In practice, the decision to conserve a breed is based mostly on the risk for extinction without considering its value for conservation. In other instances, breeds are prioritized for conservation because they are perceived to belong to the national heritage. Development of methods that are easier to understand and use may help increase the use of molecular data in prioritization. .We envisage that technological advances will allow the development of practical methods for assessing conservation values of breeds by linking genetic variation to breed characteristics. However, an increased interaction between researchers and policy-makers will be essential in order to increase awareness about objective approaches for breed priorization.

Future Perspectives

Since 2005, massively parallel DNA sequencing platforms have become available and have reduced the cost of DNA sequencing by over two orders of magnitude (Shendure and Ji 2008). In the next few years, even more effective sequencing systems will be available. This development will open the door for very low cost whole genome sequencing. As a consequence, the amount of genomic information is expected to grow exponentially. The focus is now moving from the sequencing of a single individual to hundreds or even thousands of individuals. The “1000 genome project” in humans (www.1000genomes.org),

which aims to sequence the genomes of approximately 1200 individuals from 3 major populations at approximately 4x coverage, and the "Genome 10K project" (genome10k.soe.ucsc.edu), that targets 10,000 vertebrate species' bio-specimens suitable for whole-genome DNA sequencing, are clear examples of the present trend.
Affordable resequencing will largely extend SNP identification. Analysis of breeds that have not been subject to selective breeding will foster the compilation of bias-free SNP panels for diversity studies. However, this will be of use only if SNP typing has a cost/benefit ratio than whole genome sequencing, including data management.
Transcriptome sequencing will likely substitute microarrays, permitting the discovery of alternative splicing and of new small RNAs. It is already applied to genome-wide detection of DNA methylation, the "methylome" (Lister, 2009). Such findings are likely to change fundamental insights. The widespread existence and importance of Copy Number Variation (CNV) on the level of gene expression (e.g. Beckman et al., 2007) and of microRNAs on gene regulation (e.g. Shivdasani, 2006) are recent examples of unexpected discoveries.
High throughput sequencing will be more and more integrated with new statistical approaches (e.g. Luikart et al., 2003; Joost et al., 2010). We expect that these new tools will stimulate the identification and understanding of variation underpinning important traits, including phenotypes relevant for adaptation and sustainable exploitation. With regard to conservation, whole-genome sequencing will also provide more objective indications of uniqueness than any marker panel. In addition, adaptive variation will be included in prioritization protocols in order to ensure conservation of unique adaptive variants, thus optimizing conservation efforts both *in vivo* and *in vitro*.
It is also envisaged that breeding and selection will be more and more guided by molecular analysis. Models are to be developed and customised to populations with different genetic structure (small vs. large breeds) and to different purposes (e.g. genetic improvement, control of inbreeding, maintenance of diversity). The implementation of velo-genetics and perhaps whizzo-genetics schemes was already anticipated some time ago (Georges and Massey 1991; Haley and Visscher, 1998) and may very well become reality given the significant progress in the control of cell differentiation and de-differentiation (Takahashi and Yamanaka 2006).
Summarizing, we are facing a challenging and stimulating time. Tools for molecular characterization of livestock breeds have increased faster than our capacity to process and understand this information. Storing, organizing, analyzing, and exploitation will be the future challenge for researchers and will more than ever benefit from networking and integration of expertise in different disciplines. Under this circumstances, any loss of biodiversity before characterization risks to turn into a loss of invaluable opportunities for both science and agriculture.

References

Achilli, A., Bonfiglio, S., Olivieri, A, .et al. (2009) *PLoS One* 4:e5753.

Becker, D., Tetens, J., Brunner, A., et al. (2010) *PLoS One* 5:e8689.

Beckmann, J.S., Estivill, X., Antonarakis, S.E. (2007) *Nat. Rev. Genet.* 8:639-646.

Bennewitz, J., and Meuwissen, T.H. (2005) *Genet. Sel. Evol.* 37:315-37.

Boettcher, P.J., Tixier-Boichard, M., Toro, M., et al. (2010) *Anim. Genet.* 41(s1):64-77.

Bollongino, R., Elsner, J., Vigne, J., et al. (2008) *PLoS One* 3:e3418.

Bradley, D.G., Loftus, R.T., Cunningham, P., et al. (1998) *Evol. Anthrop.* 6:79-86.

Bruford, M.W., Bradley, D.G., and Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics 4(11):900-910.

The Bovine HapMap Consortium (2009) *Science* 324:528-532.

Caballero, A. and Toro, M.A. (2002) *Conserv. Genet.* 3:289-299.

Cañón, J., García, D., García-Atance, M.A., et al. (2006) *Anim. Genet.* 37: 327-334.

Caroli, A.M., Chessa, S., and Erhardt, G.J. (2009) *J. Dairy Sci.* 92:5335-5352.

Cymbron, T., Freeman, A.R., Isabel Malheiro, M., et al. (2005) *Proc. Biol. Sci*. 272:1837-1843.

Decker, J.E., Pires, J.C., Conant, G.C., et al. (2009) *Proc. Natl. Acad. Sci. USA*. 106:18644-18649

Dekkers, J.C.M. (2004) *J. Anim. Sci.* 82:E313-E328.

Driscoll, C.A., Macdonald, D.W., O'Brien, S.J. (2009) *Proc. Natl. Acad. Sci. USA*. 106 Suppl 1:9971-9978.

Eding, J.H., Crooijmans, R.P.M.A., Groenen, M.A.M., and Meuwissen, T.H.E. (2002) *Genet. Sel. Evol.* 34:613–33.

European Genetic Cattle Diversity Consortium (2006) *Anim. Genet.* 37:475-481.

FAO (2007a) *Global Plan of Action for Animal Genetics Resources*, Rome, Italy.

FAO (2007b) *The State of the World's Animal Genetic Resources for Food and Agriculture*, Rome, Italy.

Georges, M., Massey, J.M. (1991) *Theriogenology* 35:151-159.

Ginja, C., Melucci, L., Quiroz, J., Martinez Lopez, O.R., et al. (2010) *Anim. Genet.* 41:128-141.

Grisart, B., Farnir, F., Karim, L., et al., (2004) *Proc. Natl. Acad. Sci. USA* 101:2398–2403.

Groeneveld, L. F., Lenstra, J. A., Eding, H., et al., (2010) *Anim. Genet.* 41(s1):6-31.

Haley, C.S., Visscher, P.M. (1998) *J. Dairy Sci.* 81(Suppl 2):85-97.

Hayes, B.J. Bowman, P.J., Chamberlain, A.J., and Goddard, M.E. (2009) *J. Dairy Sci.* 92:433-443.

Hansen, M.P., Law, G.R.J., and Van Zandt, J.N. (1967) *Poult. Sci.* 46:1268.

Joost, S., Colli, L., Baret, P.V., et al. (2010) *Animal Genet.* 41(s1):47-63.

Kikkawa, Y., Takada, T., Sutopo, et al. (2003) *Anim Genet.* 34:96-101.

Laval, G., Iannuccelli, N., Legault,C., et al. (2000) *Genet. Sel. Evol.* 32:187-203.

Lai, S.-J., Liu, Y.-P., Liu, Y.-X., et al. (2006). *Molec. Phylogenet. Evol.* 38:46-154.

Larson, G., Albarella, U., Dobney, K., et al. (2007) *Proc. Natl. Acad. Sci. USA.* 104:15276-15281.

Lenstra, J.A. (2009). *GlobalDiv Newsletter* 8:1-3.

Lister, R., Pelizzola, M., Dowen, R.H., et al. (2009) *Nature* 462:315-322.

Luikart, G., England, P.R., Tallmon, D., et al. (2003) *Nat. Rev. Genet.* 4:981-994.

MacEachern, S., Hayes, B., McEwan, J., Goddard, M. (2009) *BMC Genomics* 10:181

MacKinnon, K.M., Burton, J.L., Zajac, A.M., and Notter, D.R.. (2009) *Vet Immunol Immunopathol.* 130:210-220.

McCarthy, S.D., Butler, S.T., Patton, J. et al., (2009) *J Dairy Sci.* 92:5229-5238.

Meirelles, F.V., Rosa, A.J.M., Lôbo, R.B., Garcia, J.M., Smith, L.C., Duarte, F.A.M. (1999) *Genet. Molec. Biol.* 22 :543-546.

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001) *Genetics* 157:1819-1829.

Meuwissen, T.H.E. (2009) *J. Anim. Breed. Genet.* 126:333-334.

Mohamad, K., Olsson, M., van Tol, H.T., et al. (2009) *PLoS One* 4:e5490.

Qi, X.B., Jianlin, H., Wang, G., et al. (2010) *Anim Genet.* 41:242-252.

Peter, C., Bruford, M., Perez, T., et al. (2007) *Anim. Genet.* 38:37-44.

Piyasatian, N. and Kinghorn, B.P., (2003) *J. Anim. Breed. Genet.* 120:137-149.

Rincón, G., Islas-Trejo, A., Casellas, J., et al., (2009) *J. Dairy Sci.* 92:758-764.

Saura, M., Pérez-Figueroa, A., Fernández, J., et al. (2008) *Conserv Biol.* 22:1277-1287.

Shendure, J., Ji H. (2008) *Nat. Biotechnol.* 26:1135-1145.

Shuster, D. E., Kehrli M.E. Jr., Ackermann, M.R., et al. (1992) *Proc. Natl. Acad. Sci.* USA 89:9225–9229.

Simianer, H., Marti, S.B., Gibson, J., *et al.* (2003) *Ecological Econ.* 45:377–392.

Schulman, N.F., Sahana, G., Iso-Touru, T., et al., (2009) *Anim Genet.* 40:509-519.

Shivdasani, R.A. (2006) *Blood* 108:3646-3653.

Stella, A., Ajmone-Marsan, P., Lazzari, B. and Boettcher, P.J., (2010) *Genetics* 184(in press).

Stock, F., Edwards, C.J., Bollongino, R., et al. (2009) *Anim Genet.* 40:694-700.

Takahashi, K., Yamanaka, S., (2006) *Cell* 126:663-676.

Toro, M. (2006) *J. Anim. Breed. Genet.* 123:289.

Troy, C. S., MacHugh, D. E., Bailey, J. F., et al. (2001) *Nature* 410:1088-1091.

Weitzman, M.L. (1992) *Quarterly J. Econ*., 107:363–405.

Weitzman, M.L. (1993) *Quarterly J. Econ*., 108:157–183.

Yu, Y., Nie, L., He, Z.Q., et al. (1999) *Anim. Genet.* 30:245-250.