

## SCIENTIFIC BASIS FOR POLICIES IN CONSERVATION OF FARM ANIMAL GENETIC RESOURCES

M.A. Toro<sup>1</sup>, J. Fernández<sup>1</sup> and A. Caballero<sup>2</sup>

<sup>1</sup> Departamento de Mejora Genética, INIA, Carretera de La Coruña km.7, 28040 Madrid, Spain

<sup>2</sup> Dep. Bioquímica, Genética e Inmunología, Universidad de Vigo, 36310 Vigo, Spain

### INTRODUCTION

During millennia, livestock holders have created a vast range of different breeds, each with specific characteristics adapted to specific environments. Erosion of this diversity, due to the intensification of production systems promoting the substitution or crossbreeding with higher-producing breeds, is well documented. Conservation of farm animal resources is important to cope with future breeding needs and to facilitate the sustainable use of marginal areas. In this paper we will review the tools for characterising genetic diversity in structured populations and its application for establishing conservation priorities, particularly in the context of livestock breeds. Some of these topics have been previously reviewed by Toro and Caballero (2005) where many of the relevant references can be found.

### PARTITION OF GENETIC DIVERSITY BETWEEN AND WITHIN BREEDS

The most widely used parameter to measure diversity within populations is the expected heterozygosity, or gene diversity. Since the development of appropriate tools in the early 1990's, molecular markers have played a leading role in the characterisation of diversity. They provide relatively rapid and cheap assays in the absence of quality phenotypic measures. In the context of structured populations, molecular measures of differentiation are based on genetic distances in allele frequencies among populations, as well as on the popular Wright's fixation index,  $F_{ST}$ . In a structured population (metapopulation), the total gene diversity ( $GD_T$ , or expected heterozygosity) can be partitioned into components within ( $GD_W$ ) and between ( $GD_B$ ) populations. Caballero and Toro (2002) summarised formulas that allow this partition using genealogical coancestries or marker allele frequencies from each population, and that can be used as a tool for establishing conservation priorities. As an example, the relative contribution of five populations to the Iberian pig breed based on data from 36 microsatellites (Fabelo *et al.* 2004) is given in Table 1.

Note that the Guadyerbas strain contributes the least to the within-population component of diversity. But because it shows the highest genetic distance to the other strains, it contributes the most to the between-population diversity. Another way of ascertaining contributions is, following Petit *et al.* (1998), to calculate the loss or gain of diversity if one or several groups are removed, and recalculating the gene diversity (Table 1). The removal of the Lampiño variety will cause the most damaging impact, decreasing the total genetic diversity, although it will increase the average genetic distance. The removal of the Guadyerbas strain will increase the total genetic diversity of the breed. This result could seem paradoxical, but arises from standard population genetics analyses (Caballero and Toro 2002). An alternative measure of between breed diversity, preferred by some authors (Ollivier and Foulley 2005), is the diversity measure of Weitzman given in the last column of Table 1, which indicates the reduction in global diversity when one of the strains is excluded, and that is highly correlated with  $GD_B$ .

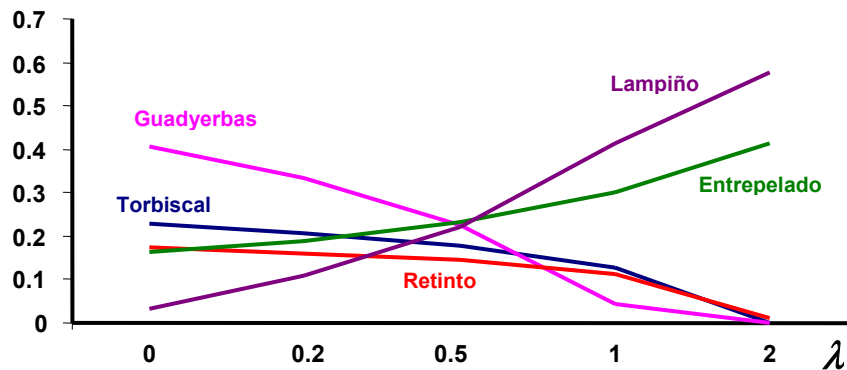
**Table 1. Relative contribution of five strains to gene diversity (GD) of the Iberian pig**

Strain	Contribution to GD			Loss (+) or Gain (-) ( $\times 100$ )			Weitzman
	$GD_W$	$GD_B$	Total	$GD_W$	$GD_B$	Total	
Torbiscal	0.117	0.018	0.135	+0.54	-0.73	-0.20	0.139
Guadyerbas	0.089	0.027	0.116	+4.81	-3.48	+0.60	0.249
Retinto	0.131	0.015	0.146	-1.26	+0.44	-0.82	0.047
Entrepelado	0.134	0.015	0.148	-1.57	+0.48	-1.09	0.039
Lampiño	0.136	0.014	0.150	-1.78	+0.50	-1.28	0.080

**Importance of within vs. between-population genetic diversity.** The important point that arises from above is that the results obtained either using between-population diversity or total diversity will produce different and sometimes opposite conservation priorities. An over-emphasis on between-population variation may result in favouring inbred populations even though they might not contain specific interesting alleles, but an over-emphasis on within-population variation will favour the largest breeds, of current commercial value and, therefore, the less endangered ones. Thus, some compromise should be attempted.

Caballero and Toro (2002) and Eding *et al.* (2002) suggested that a way to prioritise populations is to calculate the contribution of each population to a pool of animals or a germplasm bank that would maximise its genetic diversity. These optimal contributions can also be applied considering a weighted combination of the within and between-population components of gene diversity,  $\lambda GD_W + GD_B$ . Ollivier and Foulley (2005) considered a similar aggregate diversity index that makes use of the Weitzman approach as a measure of the between-breed diversity. They showed that, when the between-breed component of this aggregate is weighted by  $F_{ST}$  and the within-breed component by  $(1 - F_{ST})$ , it produces an outcome highly correlated with that using  $\lambda = 1$ . In some contexts, however, other weights would be desirable. For example, Piyasatian and Kinghorn (2003) suggested giving five times more weight to the variation between populations than to that within populations ( $\lambda = 0.2$ ), reflecting the speed by which genetic change can be made across populations compared with selection within one large mixed population. In practice, the weights will depend on the scenario imagined for the medium term use of the genetic diversity. Variation between populations may be more important because the most valuable characteristics are likely to be those for which genes are fixed or at high frequencies within the population displaying these characteristics (Reist-Marti *et al.* 2003). It will be also more important in the context of animal breeding, because it plays an essential role in the benefits derived from heterosis and complementarity (Ollivier and Foulley 2005). Bennewitz and Meuwissen (2005) also proposed an automatic weight based on maximizing the total genetic variance of a hypothetical quantitative trait, what is equivalent to using a weighting factor  $\lambda = 0.5$ .

An illustration of the optimal contributions of the five strains of Iberian pigs to a germplasm bank (Fabuel *et al.* 2004) using different weights for within and between-breed diversity is shown in Figure 1. For maximising global genetic diversity ( $\lambda = 1$ ) the populations that should contributed more are Lampiño and Entrepelado, whereas if the objective were to maximise the genetic distance ( $\lambda = 0$ ), Guadyerbas and Torbiscal strains should be given priority. Interestingly, for maximising the total genetic variance of a hypothetical quantitative trait ( $\lambda = 0.5$ ), contributions from all strains would be close to one another in this instance.



**Figure 1. Optimal contributions of strains of the Iberian pig to a germplasm bank using different weights ( $\lambda$ ) to the within and between-population genetic diversity,  $\lambda GD_W + GD_B$**

**Allelic diversity.** Allelic diversity is an alternative criterion to measure genetic diversity and to identify populations for conservation (Petit *et al.* 1998; Foulley and Ollivier 2006) because a high number of alleles implies a source of single-locus variation for important traits such as the major histocompatibility complex, which is responsible for the recognition of pathogens. It is also important from a long-term perspective, because the limit to selection response is determined by the initial number of alleles and, because it is more sensitive to bottlenecks than expected heterozygosity, it reflects better past fluctuations in population size. From the point of view of the genetic management of a population, although it is known that the population size required to maintain several alleles is much larger than the needed to keep inbreeding at acceptable rates, Fernández *et al.* (2004) have shown that, within a single population, the strategy of maximising gene diversity keeps levels of allelic diversity as high as strategies maximising allelic diversity itself, but with a better control of inbreeding.

**Probabilities of extinction** The genetic information can be refined introducing extinction probabilities. Note, however, that including extinction probabilities in the Weitzman method (that only considers between-population diversity) will give an extra weight to the inbred populations, exacerbating its problems (Eding *et al.* 2002). Gandini *et al.* (2004) used different criteria to assess the degree of endangerment of a population and proposed to estimate the number of years needed to reach a critical population size, taking into account both demographic and genetic factors. Simianer (2005) suggested the use of the criterion of the expected number of alleles segregating in the metapopulation after a given time period (usually a planning horizon of 50 years). He argued that this quantity would account for both the extinction probability of the populations and genetic drift. The extinction probability is assumed to be directly proportional to the rate of increase in inbreeding ( $1/2N_e$ , where  $N_e$  is the effective population size), and the drift was calculated from the probability of loss of alleles. For 13 European cattle breeds he assessed the effect of reducing genetic drift and changing extinction probabilities (for example paying a premium to keepers of animals of endangered breeds). Interestingly, he noticed that reducing drift within breeds is much more important than preventing breed extinction. That means that it might be better to focus on the management of diversity within a reasonably safe breed than to “rescue” breeds with high risk of extinction. He obtained a similar result by showing that the optimal allocation of funds to a larger set of African cattle breeds was 60% more efficient in terms of conserving diversity than spending the entire sum to save the three most endangered breeds from extinction (Simianer 2002).

Extinction probabilities could also be estimated in a more elaborated way. For example, in the analysis of 49 African cattle breeds, Reist-Marti *et al.* (2003) calculated extinction probabilities

using four variables related with the population (population size, change over time, distribution of the breed and risk of indiscriminate crossing), four related with the environment (organisation among farmers, existence of a conservation scheme, political situation and reliability of the information) and two related with the value of the breed (presence of special traits and cultural value).

#### **ESTABLISHING THE HIDDEN STRUCTURE OF A METAPOPOPULATION**

Traditional estimators of population structure rely on the *a priori* definition of population. Recently, methods have been developed that try to divide the total sample of genotypes of a metapopulation into an unknown number of populations (clusters). The individuals are assigned (probabilistically) to populations, or jointly to two or more populations if their genotypes indicate that they are admixed. The algorithms are based on multi-locus genotypes, solved adopting a Bayesian approach computed using Markov Chain Monte Carlo methods, and assuming multi-locus genotypes in Hardy-Weinberg (HW) and linkage equilibrium within each subpopulation. Three programs, STRUCTURE, PARTITION and BAPS, are currently available (Pearse and Crandall 2004). The approach was first applied by Rosenberg *et al.* (2001) to 20 chicken breeds. As an example, consider the five populations of the Iberian breed given in Table 1 (Fabuel *et al.* 2004). The clustering of individuals with BAPS without fixing the number of clusters suggests 12 panmictic subpopulations. All Torbiscal individuals are classified in one cluster and all Guadyerbas individuals in other. The remaining three populations are distributed over the other 9 clusters. When the alleles were randomised across individuals within strains, only three clusters appear corresponding to Torbiscal, Guadyerbas and the other three populations pooled, indicating that departures from HW and linkage equilibrium are in part responsible for the sub-structuring of the three less defined populations. When, with the original data, we impose the same number of clusters as populations (five), we obtain that on average > 98.5% of the Torbiscal genomes and > 99.5% of the Guadyerbas genomes are classified as two separated clusters. However, the results are less clear for the other populations, whose genomes are attributed to diverse clusters.

Rosenberg *et al.* (2001) have argued that genetically distinctive populations can be identified on the basis of how difficult is to separate them from others. Therefore, they suggest that the relative number of loci required for the correct clustering of several populations can be used as a way of identifying those that are genetically distinctive with respect to a collection. In the example of the Iberian pigs only six independent loci are enough to separate the two more distinct strains.

#### **RELATIONSHIP BETWEEN GENOMIC AND MARKER HETEROZYGOSITY**

The relationship between inbreeding calculated from pedigree and homozygosity calculated from molecular markers has been recently investigated. It is well known (M. Nei, quoted by Chakraborty 1981) that there are problems in estimating genomic heterozygosity using only a few molecular markers, because the expected correlation between the heterozygosity of the genome and of a sample is approximately  $\sqrt{r/n}$  where  $n$  is the number of loci in the genome and  $r$  the number of loci assayed. For example, a set of 20 markers from a genome of 20,000 genes would lead to an expected correlation of 0.03. This simple relationship seems to have been overlooked until the recent empirical reviews and theoretical updates (reviewed by Pemberton 2004). Toro *et al.* (2002) compared inbreeding of 62 Iberian pigs from two related strains calculated from a pedigree going back 20 generations with molecular coancestries estimated from 49 microsatellites. The correlation was negative for Guadyerbas (-0.32) and low for Torbiscal (0.19) but substantial for all animals together (0.69). Furthermore, the attempt to infer coancestries from molecular markers give results severely biased because the inference requires information on the true allelic frequencies of markers in the base population that are usually not known. Slate *et al.* (2004) examined 590 sheep of the Coopworth sheep with known pedigree for seven generations and genotyped 101 microsatellites. The correlation was

remarkably low (0.17) concluding that, for the correlation between the genealogical and the molecular inbreeding being substantial, a considerable number of loci and, more important, a high variance of the genealogical inbreeding values is required. This cast serious doubts about the methods proposed to estimate heritability using relationship inferred from markers in the absence of pedigree data (Ritland 2000). A second message is that it should be preferable to use pedigree information whenever available, and limiting the use of markers to verify, correct, complete or even implement pedigree recording (Fernández *et al.* 2005).

#### **RELATIONSHIP BETWEEN MOLECULAR AND QUANTITATIVE MEASURES OF GENETIC DIVERSITY**

Quantitative genetic variation is the basis of productive and reproductive traits and therefore, monitoring quantitative genetic variation (heritability or additive genetic variance), may yield more interesting information. The empirical relationship between molecular variability and morphological, behavioural or life-history variability seems to be generally low. For example, the meta-analysis carried out by Reed and Frankham (2001) suggests that the mean correlation between molecular and quantitative estimates is weak ( $0.217 \pm 0.05$ ), indicating that molecular measures only explain 4% of the variation in quantitative traits.

The analogous to  $F_{ST}$  for quantitative traits is  $Q_{ST}$ , obtained from the additive within and between-population components of the genetic variance for the considered trait. For genes that are neutral for fitness and with additive gene action, it is expected that  $F_{ST} = Q_{ST}$ . For traits under divergent selection pressure between populations,  $Q_{ST}$  is expected to be greater than  $F_{ST}$  whereas  $Q_{ST} < F_{ST}$  would indicate that selection acts on the trait towards the same optimal phenotype. Empirical data involving studies on a variety of plant and animal species indicated that  $Q_{ST}$  was generally larger than  $F_{ST}$  (Merilä and Crnokrak 2001), what is interpreted in the sense that a considerable part of the observed population divergence for quantitative traits should be attributed to differential selection pressures imposed by local environmental conditions. Notice also that the precision of estimates obtained from molecular markers and quantitative traits can be different (see Carvajal-Rodríguez *et al.* 2004). A single biallelic molecular marker gives estimates of  $F_{ST}$  with the same precision as  $Q_{ST}$  for a neutral polygenic quantitative trait, but the precision of the marker increases with the number of alleles. Using computer simulations, Le Corre and Kremer (2003) observed that, contrary to what might be expected *a priori*, estimates of variation obtained by directly studying the loci controlling the quantitative trait (QTL), are not necessarily closer to direct estimates obtained from the quantitative measures than to those from neutral variation. This is a consequence of covariances of allelic effects generated by linkage disequilibrium among selected loci and contributing to differentiation for the quantitative trait that are not expressed in single locus estimates. Thus, differentiation for QTLs might be not more informative than differentiation for neutral markers.

#### **NEUTRAL VS ADAPTIVE OR DELETERIOUS VARIATION**

Strict neutral variation would be of prime interest to carry out genetic analysis of population structure or history in order to give priority to ancestral populations. It has also been used as a surrogate of non-neutral variation, either deleterious or adaptive. Although neutral and non-neutral diversity are expected to be correlated because of the disequilibrium generated by random drift or hitch-hiking effects, their comparison could be also misleading because adaptive variants might differ in mutation rates and selective regimes (Hedrick 2001). As a large number of highly polymorphic markers are now available, the statistical power to detect differentiation between groups is very high. The problem is to know whether such differentiation reflects meaningful differences. In parallel, there could be no significant differences based on molecular markers but some important loci might be highly differentiated because selective forces are strong enough at such loci to overcome the effects of low effective size, gene flow or short divergence time. Adaptive variation can provide new criteria and

measurements to back-up conservation decisions. Differences between populations that are functional rather than neutral can be required, either for individual loci or genome regions. One way of approaching the problem is to use the existing type I markers (markers associated to known functional genes) to characterise the populations, as is planned in recent biodiversity projects. The second is to identify regions that have been subject to selection or, in other words, identifying signatures of selection among molecular markers. Unlike demographic processes, which affect the entire genome, selection affects specific important loci but will also affect linked sites, leaving its signature in the region: the level of variability will be reduced, the level of linkage disequilibrium will be increased and the genetic differentiation between populations will also be increased. Based on the seminal test of neutrality proposed by R.C. Lewontin and J. Krakauer in 1973, more recent developments have addressed this question (see Luikart *et al.* 2003). The idea behind this is to compare the observed distribution of  $F_{ST}$  values or other population parameters from markers with that expected under the neutral hypothesis, so as to identify those loci that significantly deviate from neutrality. Putative adaptive markers could be removed from the computation of neutral differentiation but could be then used as indicators of adaptive differentiation. It seems likely that the characterisation of diversity in future works will include an increasingly high use of adaptive variation, through the analysis of specific genes, quantitative traits or outlier markers, in combination with neutral variation.

#### **COMBINING GENETIC AND OTHER SOURCES OF INFORMATION**

Here, we have been mainly dealing with the use of genetic information but the final decision, setting priorities in conservation, should take into account other criteria (Oldenbroek 1999) such as the species to which the breed belongs, the adaptation to specific environments or the disease resistance, the possession of specific traits of present or future economic or scientific value and the historical or cultural value of the breed. As pointed out by Rege and Gibson (2003), in the developed world the economic value lies predominantly in productivity, whereas in low input systems in the developing world little value lies on productivity, most value lying on adaptation, survival, disease resistance, traction and nutrient recycling. There have been several attempts to include different additional sources of information. Piyasatian and Kinghorn (2003) suggested a method to balance genetic diversity, population viability and genetic merit of the breed in an objective function. In a more formal way, Simianer (2005) reviewed the theory and practical implementation of a process of decision making in livestock conservation and, more practically, Simianer *et al.* (2003) proposed an objective containing diversity, specific features (like disease resistance) and specific breed values. Gandini and Villa (2003) suggested a methodological approach to the analysis of the cultural value of livestock breeds. Reist-Marti *et al.* (2005) analysed designs, costs and benefits of possible conservation schemes for a set of 23 African breeds, and Reist-Marti *et al.* (2006) provided an algorithm to optimally allocate funds to the same breeds, concluding that with rather small amounts of total funds special traits could be conserved, but that, even with unlimited funds, a maximum of 92% of the present diversity in a 50 years horizon could be conserved. Finally, landscape genetics has recently emerged as an attempt to combine molecular markers, socio-economics and geo-statistics to determine the spatial genetic pattern of diversity and correlate it with landscape or environmental features (Manel *et al.* 2003). The ECONOGENE project tries to apply this approach to address the conservation of sheep and goats in marginal rural agrosystems.

*This work was funded by project INIA-CPE03-004-C2 (Spain).*

#### **REFERENCES**

- Bennewitz, J. and Meuwissen, T.H.E. (2005) *Genet. Sel. Evol.* **37**: 315-337.  
Caballero, A. and Toro, M.A. (2002). *Conserv. Genet.* **3**: 289-299.  
Carvajal-Rodríguez, A., Rolán-Alvarez, E. and Caballero, A. (2004) *Biol. Conserv.* **124**: 1-13.  
Chakraborty, R. (1981) *Genetics* **98**: 461-466.

- Eding, H., Crooijmans, P.M.A., Groenne, M.A.M. and Meuwissen, T.H.E. (2002) *Genet. Sel. Evol.* **34**: 613-633.
- Fabuel, E., Barragán, C., Silió, L., Rodríguez, M.C., Toro, M.A. (2004) *Heredity* **93**: 104-113.
- Fernández, J., Toro, M.A. and Caballero, A. (2004) *Conserv. Biol.* **18**: 1-10.
- Fernández, J., Villanueva, B., Pong-Wong, R. and Toro, M.A. (2005) *Genetics* **170**: 1313-1321.
- Foulley, J.J. and Ollivier, L. (2006) *Livest. Prod. Sci.* (in press)
- Gandini, G.C. and Villa, E. (2003) *J. Anim. Breed. Genet.* **120**: 1-11.
- Gandini, G.C., Ollivier, L., Danell, B., Distl, O., Georgoudis, A., Groeneveld, E., Martyniuk, E., van Arendok, J.A. M. and Woolliams, J. A. (2004) *Livest. Prod. Sci.* **91**: 173-182.
- Hedrick, P.W. (2001) *TREE* **16**: 629-636.
- Le Corre, V. and Kremer, A. (2003) *Genetics* **164**: 1205-1219.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P. *Nature Rev. Genet.* **4**: 981-994.
- Merilä, J. and Crnokrak, P. (2001) *J. Evol. Biol.* **14**: 892-903.
- Manel, S., Schwartz, M.K., Luikart, G. and Taberlet, P. (2003) *TREE* **18**: 189-197.
- Oldenbroek, J.K. (1999) «Genebanks and the conservation of farm animal genetic resources». DLO Institute for Animal Sciences and Health, Lelystad, The Netherlands.
- Ollivier, L. and Foulley, J.J. (2005) *Lives. Prod. Sci.* **95**: 247-254.
- Pearse, D.E. and Crandall, K.A. (2004) *Conserv. Genet.* **5**: 585-602.
- Pemberton, J. (2004) *TREE* **19**: 613-615.
- Petit, R.J., El Mousadik, A. and Pons, O. (1998) *Conserv. Biol.* **12**: 844-855.
- Piyasatian, N. and Kinghorn, B.P. (2003). *J. Anim. Breed. Genet.* **120**: 137-149.
- Reed, D.H. and Frankham, R. (2001) *Evolution* **55**: 1095-1103.
- Rege, J.E.O. and Gibson, J.P. (2003). *Ecol. Econ.* **45**: 319-330.
- Reist-Marti, S.B., Simianer, H., Gibson, J., Hanotte, O. and Rege, J.E.O. (2003) *Conserv. Biol.* **17**: 1299-1311.
- Reist-Marti, S.B., Abdulai, A. and Simianer, H. (2005) *J. Anim. Breed. Genet.* **122**: 95-109.
- Reist-Marti, S.B., Abdulai, A. and Simianer, H. (2006) *Genet. Sel. Evol.* **38**: 99-126.
- Ritland, K. (2000) *Mol. Ecol.* **9**: 1195-1204.
- Rosenberg, N.A., Burke, T., Elo, K., Feldman, M.W., Freidlin, P.J. and Groen, M.A.M. (2001) *Genetics* **159**: 699-713.
- Simianer, H. (2002) *Proc. 7<sup>th</sup> WCGALP* **33**: 489-493.
- Simianer, H., Marti, S.B., Gibson, J., Hanotte, O., Rege, J.E.O. (2003) *Ecol. Econ.* **45**: 377-392.
- Simianer, H. (2005) *Ecol. Econ.* **53**: 559-572.
- Slate, J., David, P., Dodds, K.G., Veenvliet, B.A., Glass, B.C., Broad, T.E. and McEwan, J.C. (2004) *Heredity* **93**: 255-265.
- Toro, M.A. and Caballero, A. (2005). *Phil. Trans. R. Soc B* **360**: 1367-1378.
- Toro, M.A., Barragán, C., Óvilo, C., Rodríguez, J., Rodríguez, C. and Silió, L. (2002) *Conserv. Genet.* **3**: 309-320.