55th EAAP Annual meeting

Bled, Slovenia, 5-8 September 2004

Commission on Animal Genetics

Session III Management of genetic variation
G3.4

Characterisation and conservation of genetic diversity between breeds

M.Toro[1] and A. Caballero[2]

[1] *Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Departamento de Mejora Genética, Carretera La Coruña km.7, 28040-Madrid, Spain,*

[2] *Facultad de Biología, Departamento de Bioquímica, Genética e Inmunología, Universidad de Vigo, Campus Universitario, 36310 Vigo, Spain*

Abstract

The different types and measures of genetic diversity and the different tools to analyze it in a subdivided population are reviewed. Emphasis is put on the classical genetic analysis coming from population genetics and on the more recent Weitzman's approach. The latter is questioned because it ignores within population genetic diversity. We discuss different reasons to emphasize either within or between breed variation and the methodology to establish some compromise. Finally, some directions of future research are presented.

Genetic diversity: Definition, Types and Measures

Genetic diversity has been defined as the variety of alleles and genotypes present in a population and that is reflected in morphological, physiological and behavioural differences between individuals and populations (Frankham *et al*. 2002). From a functional point of view the genetic diversity can be classified as neutral, deleterious or adaptive, but this classification is far from being simple. A variant is called neutral, in the population genetics tradition, if the selective coefficient is lower than $1/2N_e$, N_e being the effective population size, because, under such a condition, its destiny depends essentially on drift. The deleterious variation could be conditional on the actual environmental circumstances or unconditional if it is presumably undesirable under

most environments we can imagine. Finally, the adaptive variation is the interesting one responsible for fitness or economically important traits.

From a descriptive point of view the genetic information can refer to individual genes (either proteins, blood groups or DNA specific genes), chromosomes or quantitative genetic variation. Since the beginning of the 1990's the development of molecular data has propitiated a leading role for molecular markers (microsatellites, DNA-fingerprints, RAPDs, RFLPs, AFLPs, SNPs and DNA sequencing) in the characterisation of genetic diversity, but we should bear in mind that quantitative genetic variation is the basis of productive and reproductive traits and therefore of greatest concern in conservation biology.

At the molecular level, genetic diversity has been usually measured by the following parameters: 1) Frequencies of genotypes and alleles; 2) Proportion of polymorphic loci; 3) Observed and expected heterozygosity; 4) Allelic diversity.

The observed heterozygosity is simply the proportion of heterozygous individuals averaged over loci. The expected heterozygosity, or gene diversity, was defined by Nei (1973) as the probability that two alleles chosen at random from the population are different, and equals the proportion of heterozygotes in the population under Hardy-Weinberg equilibrium conditions. The expected heterozygosity reflects better the evolutionary or selective potential of the population because it is not a transient parameter as the observed heterozygosity, and it is less sensitive to sample sizes.

With pedigrees, the usual way to estimate diversity is to calculate $1 - F$ and $1 - f$, where F (inbreeding) and f (coancestry) are the probabilities that two genes taken at random form the same or different individuals are identical by descent (Malécot, 1948), but they correspond to the observed and expected heterozygosity in a model where all the alleles in the base or reference population are assumed to be different. On the other hand, with markers the usual estimated parameters are the observed and expected heterozygosity, but we would obtain the same results applying Malécot (1948) definition but substituting identity-by-descent by identity-by-state (Caballero and Toro, 2002). Finally, in monitoring conservation programmes, changes of gene diversity over time are important as they can indicate that the population is undergoing bottlenecks, inbreeding or it is loosing evolutionary potential. The rate of change in heterozygosity ($\Delta f = \frac{H_t - H_{t-1}}{1 - H_{t-1}}$) is called rate of advance of coancestry (if we are dealing with the

expected heterozygosity) or rate of advance of inbreeding (if we are dealing with observed heterozygosity), the effective population size being the inverse of the latter, $N_e = \frac{1}{2\Delta f}$.

Allelic diversity is an alternative criterion to measure genetic diversity and some authors (Petit *et al*. 1998; Barker, 2001) consider that this parameter is the most relevant, as a high number of alleles imply a source of single-locus variation for important traits such as the major histocompatibility complex (MHC), which is responsible for the recognition of pathogens. It is also important for a long-term perspective because the limit of selection response is determined by the initial number of alleles (Hill and Rasbash, 1986). And because it is more sensitive to bottlenecks than expected heterozygosity it reflects better past fluctuations in population size. However, because 'the effective number of alleles' is, by definition, the inverse of the mean coancestry (Crow and Kimura, 1970, p. 324), with respect to the genetic management of a population, the strategy of maximising gene diversity keep levels of allelic diversity as high as strategies maximising allelic diversity itself, but with a better control of inbreeding (Fernandez *et al*., 2004).

Tools for the analysis of genetic diversity in subdivided populations

Partition of gene diversity in a subdivided population

In a subdivided population, gene diversity is partitioned into components between and within populations (breeds in this case). Here we follow closely the development of Caballero and Toro (2002) who expressed the average global coancestry as

$\bar{f} = \tilde{f} - \overline{D}$ or

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}}{n^2} = \frac{\sum_{i=1}^{n} f_{ii}}{n} - \overline{D} = \sum_{i=1}^{n}\frac{1}{n}\left[f_{ii} - \frac{\sum_{j=1}^{n} D_{ij}}{n} \right] \qquad (1)$$

where n is the number of populations, f_{ij} is the average coancestry between populations i and j, $\bar{f}$ is the average global coancestry and D_{ij} is Nei's minimum distance between subpopulations i and j.

Equation (1) shows how the average global coancestry $\bar{f}$ depends on the within-subpopulation coancestry (first term in the brackets) and the average distance among subpopulations (second term in the brackets). Other way of expressing (1) is as genetic diversity

$$(1-\bar{f}) = (1-\tilde{f}) + \bar{D} \qquad (2)$$

The last expression represents the partition of the total gene diversity (expected heterozygosity), $GD_T = 1-\bar{f}$, into two components: the gene diversity within subpopulations $GD_{WS} = 1-\tilde{f}$ and the gene diversity between subpopulations $GD_{BS} = (\tilde{f} - \bar{f})$.

The most popular measure of population divergence is the fixation index of Wright´s (1969), that can be written as $F_{ST} = \frac{GD_{BS}}{GD_T}$ or $F_{ST} = \frac{\tilde{f} - \bar{f}}{1-\bar{f}} = \frac{\bar{D}}{1-\bar{f}}$.

As an example, the proportional contribution of each of five strains of Iberian pig to the Iberian pig breed is given in Table 1 (Fabuel, 2004). The Guadyerba strain contributes most due to its own coancestry but, because it shows the highest genetic distance to the other strains, its total contribution is lower than the Retinto strain.

Table 1. Proportional contribution of each strain or variety to the global coancestry of the Iberian breed.

	Contribution to $\bar{f}$		
Strain	*Due to within population coancestry*	*Due to the distance to other populations*	*Total*
Torbiscal	0.074	0.017	0.058
Guadyerbas	0.103	0.026	0.077
Retinto	0.099	0.019	0.080
Entrepelado	0.057	0.012	0.045
Lampiño	0.056	0.012	0.043
	$\tilde{f} = 0.393$	$\bar{D} = 0.090$	$\bar{f} = 0.304$

One way of studying the relevance of the different Iberian strains and varieties to the breed diversity as a tool for establishing conservation priorities is, following Petit *et al.* (1998), to calculate the loss or gain of diversity if one or several groups are removed, and recalculating the global average coancestry (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 although it arises from a standard population genetics analysis (Caballero and Toro, 2002). We must realise that we are considering a theoretical model in which subpopulations contribute to an infinite pool of genes. If, as a consequence of the removal of one subpopulation, gene frequencies become more equalised, this will increase the expected heterozygosity. A similar argument explains that the variability of a population will increase if a group of the most related individuals (a group of clones, for example) are eliminated and substituted by randomly chosen individuals.

Table 2. Total genetic diversity and loss (-) or gain (+) of diversity when one or two Iberian pig subpopulations are removed.

Strain	*Within strain genetic diversity*	*Average genetic distance*	*Total genetic diversity*	*Optimal contributions*
All strains	0.6067	0.0895	0.6962	
Torbiscal (T)	+0.0054	-0.0073	-0.0019	0.128
Guadyerbas (G)	+0.0408	-0.0348	+0.0060	0.044
Retinto (R)	-0.0126	+0.0044	-0.0082	0.113
Entrepelado (E)	-0.0157	+0.0048	-0.0109	0.302
Lampiño (L)	-0.0178	+0.0050	-0.0128	0.413
T + G	+0.0616	-0.0584	+0.0032	
E + L	-0.0447	+0.0098	-0.0349	
R + E	-0.0377	+0.0025	-0.1315	
G + R	+0.0376	-0.0333	0.0042	

When two subpopulations are simultaneously removed, the results agree with the previous ones. The removal of Torbiscal and Guadyerbas will hardly affect to the total diversity, whereas that of Retinto and Entrepelado will produce the maximum depletion of diversity.

Caballero and Toro (2002) also considered the following question: if we had to pool the different subpopulations to produce a single one (a synthetic population or a germplasm bank), what would be the contribution of each subpopulation to the pool in order to maximise its genetic diversity? If the different subpopulations were imposed to give different contributions (c_i) to the next generation the genetic diversity could be obtained as

$$GD_T = 1 - \bar{f} = 1 - \sum_{i,j=1}^{n} f_{ij} c_i c_j = 1 - \sum_{i=1}^{n} c_i \left[f_{ii} - \sum_{j=1}^{n} \mathsf{D}_{ij} c_j \right] \qquad (3)$$

This question can be answered by obtaining the values of c_i in equation (3) that maximise genetic diversity, with the restrictions $c_i \geq 0$ and $\sum_{i=1}^{n} c_i = 1$. These optimal contributions are given in the last column of Table 2, indicating that the strains that contributed most are Lampiño and Entrepelado. With these optimal contributions the genetic diversity will increase up to 0.7070.

Phylogenetic reconstruction based on genetic distances

Genetic distances estimated from polymorphic microsatellite markers have been the most popular method of choice to assess genetic diversity of livestock breeds. The main difference between the application of genetic distances between livestock and natural populations is that the first have been domesticated and improved by man and, therefore, the divergence period is short and the role of mutation in creating differences will be small. Another important difference, emphasised by SanCristobal *et al*. (2003) is that, when applied to breeds, genetic distance is a measure of *distinctiveness* at a given time, without reference to any model that have generated the differences but, in contrast, in the population genetics approach, genetic distance is an estimate of parameters of the model underlying the generation of differences observed.

The behaviour of the different measures of genetic distances has been reviewed by Laval *et al*. (2002). They conclude that the Reynolds distance is the best method for closely related breeds, but all of them strongly depend on the number of generations since the divergence and on the effective population size of the breeds and, therefore, no

phylogeny can be inferred from the tree in the case of closely related breeds exhibiting different effective sizes. For this reason, it is generally admitted that dealing with breeds of farm animals the interpretation of trees in terms of phylogeny can be misleading (Felsenstein, 1980; SanCristobal *et al*.,2002).

However, some authors (e.g., Barker, 1999) have argued that phylogenetic diversity will provide the best objective criterion for making conservation decisions, i.e. breeds that are taxonomically distinct should be favoured for conservation. This approach present several problems: 1) Genetic variation within populations is completely ignored; 2) Construction of trees using admixed populations, as often happens in livestock, contradicts the principles of phylogeny reconstruction (Felsenstein, 1982); 3) It fails to take into account the fact that genetic distances vary greatly according to the marker used and the recent demographic history of the breed (whether it has passed though a population bottleneck); 4) Markers used for estimating genetic distances are assumed to represent neutral loci but natural and artificial selection have been crucial in the formation and evolution of domestic breeds.

Multivariate consensus representation of genetic relationship among populations

Among the many multivariate analysis methods principal component analysis is a simple and powerful one that has been advocated by Moazami-Goudarzy and Laloe (2002). It present some advantages: 1) It is less sensitive for data where admixtures are known to have occurred; 2) It is independent from the mutation model assumed; 3) It can be applied to various types of markers (microsatellites, AFLPs, proteins, blood groups, phenotypical traits, ...)

The analysis is carried out in a two-step process. The first consist of performing single-marker analyses and studying if they are congruent by their Mantel correlations. If that is the case, a principal component analysis can by done on the entire data with the possibility of evaluating the relative contribution of each marker in the structure of the principal components.

Clustering analysis

Recently, a clustering method has been proposed (Pritchard *et al*., 2000; Dawson and Belkhir, 2001, Rosenberg *et al*., 2001, Corander *et al*., 2003) that constructs genetic clusters from a set of individual multilocus genotypes estimating, for each individual, the fraction of its genome that belongs to each cluster without any prior information on

the structure of the population. Thus, the individuals are assigned (probabilistically) to populations, or jointly to two or more populations if their genotypes indicate that they are admixed. The algorithm is solved adopting a Bayesian approach computed using Markov Chain Monte Carlo methods. It constitutes a most flexible alternative to cluster methods based on genetic distances. It can separate a set of individuals in several populations if their genetic origin is unknown beforehand or, as in the present situation, to study the correspondence between inferred genetic clusters and known predefined population categorisations.

As an example (Fabuel *et al*., 2004), Table 3 shows data from 36 microsatellites from 213 pigs belonging to 5 populations of Iberian and one population of Duroc, which classified in two clusters by the STRUCTURE algorithm of Pritchard *et al*. (2002). The results indicate that most of the genomes of all the Iberian strains and varieties fall into the same cluster, with the genome of the Duroc breed constituting the other. Both the Torbiscal and the Guadyerbas strains are the subpopulations whose genomes are differentiated the most unambiguously from Duroc.

Table 3. Proportion of membership of each predefined population in each of either two or five possible clusters.

	Two clusters assumed		**Five clusters assumed**				
Population	*1*	*2*	*1*	*2*	*3*	*4*	*5*
Torbiscal	0.001	0.999	0.004	0.003	0.002	0.985	0.006
Guadyerbas	0.001	0.999	0.001	0.001	0.001	0.002	0.995
Retinto	0.011	0.989	0.449	0.451	0.009	0.084	0.007
Entrepelado	0.050	0.950	0.527	0.419	0.008	0.030	0.016
Lampiño	0.010	0.990	0.321	0.223	0.351	0.024	0.081
Duroc	0.997	0.003					

On the other hand, when the STRUCTURE algorithm is applied to the Iberian breed assuming the same number of clusters and subpopulations (five), we obtain the results presented in the right-hand side of Table 3. They indicate that, on average, 98.6% of the Torbiscal genomes and 99.5% of the Guadyerbas genomes are classified as two separate

clusters. However, the results are less clear for the other subpopulations, whose genomes are attributed to diverse clusters. This again emphasises that the first two strains constitute more defined populations than the others.

Rosenberg *et al*. (2001) have argued that genetically distinctive populations can be identified on the basis of how difficult it is to separate them from other breeds. That is, if some breeds were easier to separate into clusters than others with only a small number of markers, this could indicate the presence of distinctive multilocus genetic combinations in the breeds that were easier to separate. Therefore, they suggest that the relative number of loci required for the correct clustering of several breeds can be used as a way of identifying populations that are genetically distinctive with respect to a collection.

The Weitzman approach

Thaon d'Arnoldi *et al.* (1998) proposed to set conservation priorities through the analysis of genetic distances by the Weitzman (1992) approach to measure the global diversity and the marginal loss of diversity attached to each breed. From a genetic distance matrix, Weitzman (1992) proposed a method to construct hierarchical trees based on a form of maximum likelihood phylogeny conditional on the model. Thus, the contribution of an element to group diversity is proportional to the reduction in tree length caused by the removal of the element from the group. It is computationally intensive, limiting its use to sets of 25 or fewer populations. Laval *et al.* (2000) applied this method to analyse the genetic diversity of 11 pig breeds from six European countries, Cañón *et al.* (2001) to 18 European beef cattle breeds, Aranguren-Méndez *et al.* (2002) to five endangered Spanish donkey breeds and Reist-Marti *et al*. (2003) to 49 African cattle breeds.

Several authors have criticised the Weitzman approach (Caballero and Toro, 2002; Eding *et al.*, 2002). This method does not have a clear interpretation in terms of the most widely accepted measure of genetic variability, Nei's (1973) expected heterozygosity and, therefore, has properties, such as that the removal of an element always decreases the variability, or the calculation of marginal diversity, that are at variance with classical population genetics ideas. Besides, it does not have a way of including the population size if desired, and most important of all, it ignores within-population variability, which is a crucial component of total variability.

The ignorance of the within-group variability is a drawback not only of the Weitzman method but also all methods based only on genetic distances. In fact, one of the properties of the method (monotonocity in distance) is that the diversity in a set of populations should increase if the distance between populations increases. Thus, it will favour inbred populations with extreme allele frequencies, whereas the coancestry approach would favour non-inbred populations with an even distribution of gene frequencies. Thaon d'Arnoldi *et al.* (1998) also suggest the inclusion, together with the Weitzman method, of the probability of extinction of each population. But, as Eding *et al.* (2002) pointed out, this will make things worse because inbred populations will get an even higher weight. On the other hand, an over-emphasis on within-breed variation will favour the largest breeds, of current commercial value, and therefore the less endangered ones.

As an example, consider the analysis of genetic diversity carried out by Laval *et al*. (2000) for eleven European pig breeds using 18 microsatellites. Column 2 of Table 4 shows the marginal losses of diversity calculated by Laval *et al*. (2000, Table VI) with the Weitzman method, when each of the eleven breeds (column 1) is removed from the set. Column 3 of Table 4 gives the loss/gain of global genetic diversity when each of the breeds is removed, calculated as in the previous example. Again the first term of the sum refers to the loss/gain due to the average coancestry of the subpopulation, while the second term refers to the loss/gain due to its average distance with all the others.

According to Laval *et al*. (2000) (see column 2 of Table 4), the highest and lowest losses of diversity are incurred with the extinction of the French Basque (FRBA) and the Piétrain (BEPI) breeds, respectively. They also showed that the four French local breeds (FRBA, FRGA, FRLI and FRNO) altogether account for half of the total diversity, supporting the potential value of preserving local endangered breeds in the maintenance of species diversity. However, the analysis of genetic diversity using the global coancestry when each breed is removed (column 3 of Table 4) gives quite different results. Removal of the FRBA breed will produce on of the largest increases in diversity over the remaining pool, while removal of the BEPI breed will produce a slight increase in diversity. In addition, removal of the four French breeds would produce a substantial increase in diversity (7.02 – 3.81 = +3.21%) instead of a large decrease. Therefore, the conclusions that one can draw from the two analyses are very different and, in fact, can be opposite.

How important is within vs. between genetic diversity?

The important point that arises above is that the results obtained either using between-population diversity or total diversity will produce different and sometimes opposite conservation priorities. Therefore, some compromise should be attempted.

In the framework of the classical partition of gene diversity the simplest way to act is to carry out the classical analysis of gene diversity considering a weighted combination of the within-population gene diversity and the average genetic distances, $\lambda(1-\tilde{f})+\overline{D}$.

Table 4. Reanalysis of genetic diversity with the data of Laval et al. (2000).

Breed	*Weitzman*	*Loss/gain* GD_T	c_i *for max* GD_T
BEPI	–3.8	–0.80 +1.01 = +0.21	0.0005
DKSO	–10.6	–0.23 –0.22 = –0.45	0.1128
FRBA	–15.2	+2.62 –1.95 = +0.67	0.0228
FRGA	–7.9	+0.48 +0.12 = +0.60	0
FRLI	–10.8	+1.34 –0.66 = +0.68	0
FRNO	–9.5	+0.48 –0.05 = +0.43	0
DELR	–11.6	–1.23 –1.30 = –2.53	0.2832
DESH	–5.2	–1.80 –1.14 = –2.94	0.2019
NLLW	–12.1	+0.48 –0.58 = –0.10	0.1214
SELR	–4.4	–0.52 +1.16 = +0.64	0
SEWP	–9.4	–0.80 –0.02 = –0.82	0.2573

Fabuel *et al*. (2004) present an application to the calculation of the optimal contributions of the five strains of Iberian pigs to a possible synthetic of germplasm bank (Fabuel *et al*., 2004). lt appears in Table 5. The results agree with those of the analysis of genetic diversity. For maximising global genetic diversity (λ = 1) the strains that should contributed more are Entrepelado and Lampiño, whereas if the objective was to maximise the genetic distance (λ = 0) Guadyerbas and Torbiscal strains should be prioritised. For λ = 2, two of the subpopulations would have a null contribution. If, for whatever reason, we want to set up a minimum for the contribution of any strain or

variety, we can include a restriction in the quadratic programming solver and we would obtain the appropriate solutions ($\lambda^* = 2$, minimum contribution equal to 0.02).

In the same spirit, Eding *et al.* (2002) proposed to work always with optimal contributions. Their strategy is: 1) to rank breeds according to their optimal contributions (core set); 2) to calculate the gene diversity of a safe core set formed by commercial lines together with their optimal contributions; 3) to calculate the gain in gene diversity when one extra breed is added to the safe core. They illustrate the method by an example involving 45 Dutch poultry breeds.

Table 5. Optimal contributions to a synthetic line or to a germplasm bank for different weights of the within and between population variability ($\lambda\tilde{f}$ - $\overline{D}$).

Population	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 1$	$\lambda = 2$	$\lambda^* = 2$
Torbiscal	0.228	0.208	0.128	0	0.020
Guadyerbas	0.406	0.333	0.044	0	0.020
Retinto	0.173	0.161	0.113	0.012	0.020
Entrepelado	0.162	0.190	0.302	0.412	0.392
Lampiño	0.031	0.107	0.413	0.576	0.548
$\tilde{f}$	0.443	0.424	0.349	0.326	0.332
$\overline{D}$	0.103	0.101	0.056	0.027	0.036
$\bar{f}$	0.340	0.323	0.293	0.298	0.297

Ollivier and Foulley (2002) proposed an aggregate diversity (linear combination of within and between population diversity weighted appropriately),

$$F_{ST}V + (1 - F_{ST})(1 - H(S/k) / H(S)),$$

where

V = Weitzman measure of loss of diversity,

H(S) = average within heterozygosity,

H(S/k) = average heterozygosity deleting breed k.

Although this expression is intuitively appealing, it has not an interpretation in terms of classical measures of genetic diversity.

Piyasatian and Kinghorn (1999) argued that the weights to be given to within *vs* between population genetic diversity will depend on the scenario imagined for the

medium term use of the genetic diversity. They suggest giving five times more weight to the variation between breeds than to that within breeds. The reason is that variation between breeds is more desirable because genetic effects are 'packed' in a more accessible way. It is easier access to known genes of quantitative traits loci that are at extreme frequencies if we are looking towards a greater adaptation to a changing or novel environment, and the five value reflects the speed with which genetic change can be made across breeds compared with selection within one large mixed population.

Reist-Marti *et al*. (2003) consider between breed variation as much more important because the most valuable characteristics are likely to be those for which genes are fixed or at high frequencies within the breed the breed displaying these characteristics. Between breed diversity will also be more important if the plan is to use them as part of crossbreeding programmes, the diversity between populations should be priorized because both heterosis and complementarity are functions of this type of genetic variation. The same will apply if the plan is introgression programmes of some specific trait. On the other hand, if we are thinking of the future creation of a new population able to cope with a challenging environment or with diversified production conditions, within-population diversity will be important (Notter, 1999).

Finally, in the last years there has been several attempts to include different sources of information besides of the analysis of genetic diversity of the breeds. Piyasatian and Kinghorn (2003) suggest a method to balance genetic diversity, population viability and genetic merit of the breed as objective function in breed conservation. Simianer *et al*. (2003) extended the Weitzman approach to include extinction probabilities over a chosen time period. This allows to estimate the expected diversity at the end of assumed period as $E(D) = \sum P(K)D_K$ where K is a vector of size n (the number of subpopulations) containing the indicator variables k_i, where $k_i = 1$ if the breed is still existing and $k_i = 0$ if it is extinct. They also defined the marginal diversity m_i as $\frac{\delta E(D)}{\delta z_i}$, that reflects the change of diversity when the extinction probability of breed i is increased by one unit. A simple way of setting the extinction probabilities is to assume that they are directly proportional to $\Delta F = 1/2N_e$, but it could be done 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 indiscriminated crossing), four related with the environment (organisation among the 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). Simianer (2002) also proposed to use a utility function to include information for specific properties of certain breeds (trypanotolerance, fertility, etc.) which make one breed more valuable that others.

New directions in future studies

Besides the advances of molecular approaches, such as the availability of high-density genotyping, via SNPs or sequencing, there are two main topics that will probably be important in future diversity studies:

1) The relationship between the degree of divergence in neutral markers and the degree of divergence in genes coding for quantitative traits, and the related question of the relative importance of random genetic drift and directional selection as causes of population differentiation in quantitative traits. This will perhaps be carried out by the comparison between the fixation index F_{ST} and its analogous for quantitative traits, termed Q_{ST} by Spitze (1993), as has been done in natural populations (Merilä and Crnokrak, 2001).

2) The a Assessment of differences between breeds that are not neutral but functional, either based on individual loci or on genome regions. This could provide new criteria and measurements to back-up conservation decisions. There are two ways of approaching the problem. The first is to use the existing type I markers (markers associated to known functional genes) to characterise the breeds, as it is planned in recent biodiversity projects (Blott, 2003). The second is trying to identify loci that have been subject to selection showing that they present deviations from neutral expectations or, in other words, identifying signatures of selection among molecular markers (Vitalis *et al*., 2001).

A final word

Here, we have been mainly dealing with the use of genetic information but we must recall that this is only one of the criteria to consider in the final decision of setting priorities in livestock conservation. There seems to be a consensus about such criteria (Oldenbroek, 1999; Ruane, 1999): 1) The species to which the breed belongs; 2) The adaptation to specific environment or the disease resistance; 3) The possession of

specific traits of present or future economic or scientific value and ; 4) The historical or cultural value of the breed.

References

Aranguren-Méndez J, Jordana J, Gómez M (2002) Genetic conservation of five endangered Spanish donkey breeds. *J Anim Breed Genet* **119**: 256-263.

Barker JSF (1999) Conservation of livestock breed diversity. *AGRI* **25**: 33-43.

Blott SS (2003) Characterisation of genetic variation in the pig breeds of China and Europe- The pigbiodiv2 project. Arch. Zootec. **52**: 207-217.

Caballero A, Toro MA (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* **3**: 289-299.

Cañón J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S *et al.* (2001) Genetic diversity of local European beef cattle breeds for conservation purposes. *Genet Sel Evol* **33**: 311-332.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics **163**: 367-374.

Crow J, Kimura M (1970) An Introduction to Population Genetics Theory. Harper & Row, New York.

Dawson KJ, Belkhir K (2001) A bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet Res* **78**: 59-73.

Eding H, Crooijmans PMA, Groenne MAM, Meuwissen THE (2002) Assessing the contribution of breeds to genetic diversity in conservation schemes. *Gen Sel Evol* **34**: 613-633.

Fabuel E, Barragán C, Silió L, Rodríguez MC, Toro MA (2004) Análisis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. Heredity **93**: 104-113.

Felsenstein J (1982) How can we infer geographically and history from gene frequencies. J theor Biol **96**:9-20.

Fernández J, Toro MA, Caballero, A (2004) Managing individuals' contribution to maximize the allelic diversity maintained in small conserved populations. Conservation Biology (in press)

Frankham R (1995) Conservation Genetics. *Ann Rev Genet* **29**: 305-327.

Hill W.G. , Rasbash, J. (1988) Models of long term artificial selection in finite populations. Genet. Res. **48**: 41-50.

Laval G, SanCristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genet. Sel.. Evol. **34**: 481-507.

Laval G, Iannuccelli N, Legault C, Milan D, Groenen MAM, Giuffra E *et al.* (2000) Genetic diversity of eleven European pig breeds. *Genet Sel Evol* **32**: 187-203.

Malécot G (1948) *Les Mathématiques de l'Hérédité*. Masson et Cie: Paris.

Merilä J., Crnokrak P (2001). Comparison of genetic differentiation at marker loci and quantitative traits. J Evol Bio **14**, 892-903.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad. Sci USA* **70**: 3321-3323.

Notter DR (1999) The importance of genetic diversity in livestock populations of the future. J Anim Sci **77**: 61-69.

Oldenbroek JK (1999) *Genebanks and the conservation of farm animal genetic resources*. DLO Institute for Animal Sciences and Health, Lelystad, The Netherlands.

Ollivier L, Foulley JJ (2002) Some suggestions on how to preserve both within- and between-breed genetic diversity. *53rd Annual Meeting of the European Association for Animal Production*, Cairo, Egypt.

Petit RJ, El Mousadik A, Pons, O (1998) Identifying populations for conservation on the basis of genetic markers. *Conser. Biol.* **12**: 844-845.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Reist-Marti SB, Simianer H, Gibson J, Hanotte O and Rege JEO (2003) Weitzman's approach and conservation of breed diversity: an application to African cattle breeds. Conser Biol **17**: 1299-1311.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groen MAM *et al.* (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**: 699-713.

Ruane J (1999) A critical review of the values of genetic distance studies in conservation of animal genetic resources. *J Anim Breed Genet* **116**: 317-323.

SanCristobal *et al*. (2002) Genetic diversity in pigs- Preliminary results on 58 European breeds and lines. Proc. 7th World Congress on Genetics Applied to Livestock Production, INRA Castanet-Tolosan, **33**: 523-528.

SanCristobal M, Chevalet C, Foulley J-L, Ollivier L (2003) Some methods for analysing genetic marker data in a biodiversity setting- example of the PIGBIODIV data. Arch. Zootec. **52**: 173-183.

Simianer H (2002) Noah's dilemma: which breeds to take aboard the ark? Proc. 7th World Congress on Genetics Applied to Livestock Production, INRA Castanet-Tolosan, **33**: 473 -480.

Simianer H, Marti SB, Gibson J, Hannotte O, Rege JEO (2003) an approach to the optimal allocation of conservation funds to minimize loss of genetic diversity between livestock breeds. Ecol Econ **45**: 377-392.

Spitze K (1993). Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. Genetics **135**: 367-374.

Thaon d'Arnoldi C, Foulley J-L, Ollivier L (1998) An overview of the Weitzman approach to diversity. *Genet Sel Evol* **30**: 149-161.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across loci as evidence of selection. Genetics **158**: 1811-1823.

Weitzman ML (1992) On diversity. *Quart J Econom* **107**: 363-405.

Wright S (1969) *Evolution and the Genetics of Populations*. Vol. 2, *The Theory of Gene Frequencies*. The University of Chicago Press: Chicago.