

Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers

E Fabuel, C Barragán, L Silió, MC Rodríguez and MA Toro

Departamento de Mejora Genética Animal, INIA, Carretera La Coruña km. 7, Madrid 28040, Spain

The Iberian breed is the most important pig population of the Mediterranean type. The genetic structure of two strains (Torbiscal and Guadyerbas) and three varieties (Retinto, Entrepelado and Lampiño) of this breed was studied using 173 pigs genotyped for 36 microsatellites. In addition, 40 pigs of the related Duroc breed were also analysed. In the 1960s, the Iberian breed's numbers were severely reduced by disease, due to economic change and to crossbreeding. Varieties are in danger of disappearance or blending. A new conservation strategy is required. An analysis was performed that allows us to ascertain the loss or gain of genetic diversity

if one or several subpopulations are removed. The results are compared with those using the Weitzman method. The two methodologies produce conservation priorities that are completely different, the reason being that the Weitzman method does not take into account the within-population genetic diversity. We apply optimal contribution theory and a new procedure for cluster analysis, and discuss their value in the general framework of the problems of setting of priorities and tactics for the conservation of genetic resources.

Heredity (2004) **93**, 104–113, advance online publication, 19 May 2004; doi:10.1038/sj.hdy.6800488

Keywords: Iberian pig; global coancestry; genetic distances; conservation priorities; microsatellites

Introduction

The Iberian breed constitutes the largest of the surviving populations of pigs of Mediterranean type. The population of Iberian pigs was maintained for centuries in large herds grazing the sparse oak and cork oak woodlands of the SW of the Iberian Peninsula. The morphology of the Iberian pig (dark skin and hair colour, pointed snout and strong legs) makes it resistant to sunstroke and high summer temperatures and adapts it for pasturing. From November to February, it can be quickly fattened by the consumption of available acorns, grass, small roots and bulbs (López-Bote, 1998). Although the Iberian pigs share the above features, several local varieties arose in the traditional population, showing phenotypic differences (black hairless, red and blond pigs) and with only rare genetic interchange. Some of these varieties were exported during the colonisation of the Americas, being the direct ancestors of several Creole pig breeds (Lemus-Flores *et al.*, 2001). Red Iberian pigs imported from Portugal and Spain in the XIX century also contributed to the origin of the Duroc–Jersey breed in the United States (Vaughan, 1950).

The numbers of the Iberian breed have been drastically reduced since 1960 due to the outbreak of the African swine fever, the lowered value of animal fats and the massive introduction of more efficient foreign breeds. In the last few years, this population bottleneck has been

reversed and the production of pigs of the Iberian type has increased to satisfy the new demand of top quality meat and cured products. However, the old breed structure, with differentiated varieties locally distributed, has been replaced by a pyramidal structure based on crossbreeding with Duroc, and a strong dependence on a small number of breeding nuclei supplying purebred Iberian animals to all the production tier (Silió, 2000). In these circumstances, some ancestral varieties have disappeared, others are endangered or blended, necessitating a new design for programmes of conservation of these genetic resources.

In the establishment of conservation programmes for a domestic species (or breed) in which a number of different breeds (or varieties) exist, a question to be asked is which populations should contribute to the population to be conserved and in what proportions. In other words, by what criteria should the conservation priorities be set. This is a controversial issue, although the final decision should, together with genetic information, take into account several factors such as the adaptation to specific environments or diseases, and the possession of specific traits of cultural, scientific or future economic value (Ponzoni, 1997; Oldenbroek, 1999; Ruane, 1999). A second question, especially relevant in implementing germplasm banks, is how many individuals from each population should contribute to the bank of gametes.

Phylogenetic techniques based on genetic distances estimated from polymorphic microsatellite markers are the method of choice to assess the genetic diversity of livestock breeds (Hall and Bradley, 1995). Thaon d'Arnoldi *et al.* (1998) emphasised the analysis of genetic distances by the Weitzman (1992) approach, to measure

Correspondence: M Toro, Departamento de Mejora Genética Animal, INIA, Carretera La Coruña km. 7, Madrid 28040, Spain.

E-mail: toro@inia.es

Received 31 December 2002; accepted 26 March 2004; published online 19 May 2004

the global diversity and the marginal loss of diversity attached to each breed. Conservation priorities can be based on these results. 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 and Aranguren-Méndez *et al.* (2002) to five endangered Spanish donkey breeds, supporting the value of the preservation of local endangered breeds in the maintenance of species diversity. Caballero and Toro (2002), however, consider these phylogenetic methods inappropriate for within-species breed conservation, because of the genetic variation within groups is ignored, although it may be of great importance for the management of livestock breeds. They have proposed other basic tools, based on molecular coancestry measured from markers, to analyse genetic diversity in subdivided populations as breeds or varieties of livestock species.

Recently, alternative clustering methods, have been proposed, which allow the inference of population structure and the assignment of individuals to populations (Pritchard *et al.*, 2000; Dawson and Belkhir, 2001). These methods are also useful for the definition of conservation units, and have been applied to 20 chicken breeds by Rosenberg *et al.* (2001).

The objective of this study was to assess the genetic diversity in the main varieties and preserved strains of Iberian pigs and their relationship to the Duroc breed, comparing the approaches mentioned above in their utility in establishing conservation strategies for this breed.

Material and methods

Populations

Two of the groups of Iberian pigs considered here, Guadyerbas and Torbiscal, belong to an early conservation programme established in 1945 at 'El Dehesón del Encinar' (Oropesa, Toledo, Spain). The herd was formed from four founder populations, representative of the most important varieties of Iberian pigs, that were kept genetically isolated until 1963. Then, the four groups were slowly blended, resulting in the composite strain named Torbiscal (Rodrigáñez *et al.*, 2000). One of the founder populations (Guadyerbas) was also maintained as a separate closed strain, representing the endangered black hairless variety from the Guadiana Valley (Toro *et al.*, 2000). The complete genealogy of all the animals is available since 1945, there having been 18.9 (Guadyerbas) and 21.0 (Torbiscal) generations since the population founders until the animals were genotyped in this study. The remaining Iberian pig groups represent the main extant varieties: black hairless (Negro Lampiño) and red (Retinto) pigs and a black hairy variety (Entrepelado), whose piglets show a chestnut colour at birth. Blood samples were collected from 173 individuals (77 boars and 96 sows) inscribed in the breed herdbook, their distribution being as follows: 31 Torbiscal, 32 Guadyerbas, 50 Retinto (seven herds), 30 Lampiño (three herds) and 30 Entrepelado (five herds). Owing to their historical and current relations with the Iberian pigs, a further 40 Duroc pigs from seven herds were also sampled and analysed.

Microsatellites

All the animals were genotyped for 36 microsatellite markers, two on each autosome (Table 1) They were chosen for their reproducibility, position on the chromosome, polymorphism and absence of null alleles. Most markers belong to the panel recommended by the ISAG-FAO Advisory Committee for genetic distance studies (FAO, 1998). However, some of those included in this panel, for example, CGA and S0068, were discarded because of the low reproducibility of their alleles. PCR-amplified microsatellite markers were analysed with Genescan software on capillary electrophoresis equipment with fluorescent detection (ABI PRISM 310 genetic analyzer). To increase the accuracy of allele size determination, four control animals were genotyped in all the gels. Genotypes were stored in the Gemma database (Iannuccelli *et al.*, 1996).

Statistical analysis

Analysis of genetic diversity in a subdivided population: There are several measures of genetic variability, but the most popular is the genetic diversity, defined by Nei (1973) as the heterozygosity expected under the Hardy-Weinberg (H-W) equilibrium conditions. It also corresponds to the complement of the average global molecular coancestry, defined as the probability that two alleles taken at random from the population are the same. In a subdivided population, the study of genetic diversity is focused on the partition on the components between and within subpopulations (strains and varieties in this case). Here, we follow closely the development of Caballero and Toro (2002) who expressed the average global coancestry as

$$\bar{f} = \sum_{i=1}^n \frac{1}{n} \left[s_i - D_{ii} - \frac{\sum_{j=1}^n D_{ij}}{n} \right] \quad (1)$$

where n is the number of populations, \bar{f} is the average global coancestry, s_i the average of within-individual coancestry of the i th population, D_{ii} the average distance between individuals of the i th population and D_{ij} is Nei's minimum distance between subpopulations i and j . The definitions and the way of calculating the above parameters are given in Appendix A.

The right-hand side of Eq. (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

$$\begin{aligned} (1 - \bar{f}) &= (1 - \bar{f}) + \bar{D} = 1 - \bar{s} + \bar{D} + \bar{D} \\ &= (1 - \bar{s}) + (\bar{s} - \bar{f}) + (\bar{f} - \bar{f}) \end{aligned} \quad (2)$$

The last expression represents the partition of the total genetic diversity (heterozygosity), $GD_T = 1 - \bar{f}$, into two components: the genetic diversity within subpopulations $GD_{WS} = 1 - \bar{f}$ and the genetic diversity between subpopulations $GD_{BS} = (\bar{s} - \bar{f})$. The first also has two components, the genetic diversity within individuals, $GD_{WI} = 1 - \bar{s}$ and the genetic diversity between individuals, $GD_{BI} = \bar{s} - \bar{f}$. If G is the proportion of within-subpopulation genetic diversity corresponding to within-individual diversity $G = GD_{BI}/GD_{WS}$, then Wright's

Table 1 Number of alleles and observed and expected heterozygosities of microsatellite markers in Iberian and Duroc breeds

Chromosome	Markers	Iberian			Duroc		
		No. of alleles	Obs. het.	Exp. het.	No. of alleles	Obs. het.	Exp. het.
1q	S0113	6	0.676	0.727	3	0.375	0.377
1q	S0155	5	0.457	0.492	4	0.650	0.570
2p	SW240	8	0.601	0.726	6	0.800	0.748
2q	SW395	6	0.555	0.709	3	0.300	0.322
3p	SW72	6	0.503	0.683	6	0.600	0.729
3q	S0002	8	0.607	0.672	5	0.725	0.733
4p	S0301	5	0.364	0.533	5	0.750	0.743
4q	S0097	12	0.711	0.823	7	0.475	0.592
5p	SW413	4	0.584	0.720	5	0.500	0.756
5q	S0005	14	0.636	0.782	10	0.625	0.782
6p	SW1057	8	0.659	0.724	5	0.475	0.748
6q	SW2419	8	0.711	0.756	4	0.500	0.646
7p	S0025	8	0.595	0.675	5	0.425	0.608
7q	SW632	9	0.549	0.682	7	0.525	0.543
8q	S0178	9	0.636	0.754	5	0.650	0.703
8q	S0225	7	0.561	0.660	4	0.625	0.723
9p	SW911	7	0.549	0.701	6	0.625	0.664
9q	SW1349	5	0.491	0.560	6	0.175	0.637
10p	S0038	8	0.462	0.770	6	0.425	0.429
10q	S0070	11	0.711	0.815	9	0.625	0.658
11p	S0385	5	0.538	0.668	5	0.750	0.769
11q	SW703	5	0.410	0.558	2	0.475	0.485
12p	SW874	7	0.578	0.771	7	0.500	0.725
12q	S0106	13	0.682	0.813	6	0.625	0.600
13q	S0219	4	0.480	0.650	4	0.400	0.482
13q	S0291	7	0.532	0.723	6	0.675	0.733
14q	SW210	5	0.532	0.672	4	0.450	0.575
14q	SW857	5	0.624	0.709	5	0.400	0.662
15q	SW1111	6	0.659	0.783	6	0.600	0.607
15q	SW936	9	0.711	0.808	5	0.625	0.710
16q	S0026	7	0.341	0.458	5	0.525	0.658
16q	S0061	5	0.665	0.657	5	0.575	0.707
17q	SW1920	8	0.601	0.761	7	0.450	0.773
17q	SW24	6	0.572	0.688	6	0.750	0.754
18q	S0120	7	0.584	0.729	5	0.400	0.554
18q	SW787	6	0.607	0.681	7	0.725	0.810
	Average	7.2	0.576	0.697	5.4	0.549	0.648

(1969) F-statistics can be written as

$$\frac{GD_{WI}}{GD_T} = (1 - G)(1 - F_{ST}), \quad \frac{GD_{BI}}{GD_T} = G(1 - F_{ST}), \quad (3)$$

$$\frac{GD_{WS}}{GD_T} = (1 - F_{ST}) \quad \text{and} \quad \frac{GD_{BS}}{GD_T} = F_{ST}$$

Finally, let us consider a different question that may be useful for the conservation of genetic diversity either in a live conservation or in a cryoconservation programme. 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 (see below), the genetic diversity could be obtained as

$$GD_T = 1 - \bar{f} = 1 - \sum_{i,j=1}^n f_{ij}c_i c_j$$

$$= 1 - \sum_{i=1}^n c_i \left[f_{ii} - \sum_{j=1}^n \mathcal{D}_{ij}c_j \right] \quad (4)$$

This question can be answered by obtaining the values of c_i in Eq. (4) that maximise genetic diversity with the

restrictions $c_i \geq 0$ and $\sum_{i=1}^n c_i = 1$. The appropriate values of c_i can be obtained by integer quadratic programming. Similar equations were constructed to give different weighting to genetic diversity and genetic distance between populations (see Results).

Phylogenetic trees and the Weitzman method to measure genetic diversity: Although there are many measures of genetic distances, Reynolds's distance is considered to be the best for studying populations that have recently diverged, with neighbour joining (NJ) being the preferred technique to obtain a graphical representation of the distance matrix, when the populations differ in effective population size, as is usual in domestic breeds (Eding and Laval, 1999). It was constructed from the Reynolds distances using the PHYLIP package (Felsenstein, 2001).

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.

Clustering analysis: Recently, a clustering method has been proposed (Pritchard *et al.*, 2000; Dawson and Belkhir, 2001) 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. Here we applied, for this purpose, the Structure algorithm of Pritchard *et al.* (2000).

Results

Number of alleles, heterozygosity and deviations from H–W

Table 1 shows the number of alleles for each genotyped microsatellite that ranged from 4 to 14 (average 7.2) summed across the Iberian breed as a whole and from 2 to 10 in the Duroc population (average 5.4). In the Iberian breed, the average number of alleles per population ranged from 2.8 (microsatellite S0219) to 7.8 (microsatellite S0005). Table 1 also gives, for each locus, the observed heterozygosity by direct count and the expected heterozygosity under H–W equilibrium. The observed values were generally lower than the expected values, indicating heterogeneity between populations within the breed.

In Table 2 we show for each population i , the self-coancestry f_{ii} , the average of individual self-coancestries s_i , the average inbreeding F_i , together with the para-

meters α_i and G_i (see Appendix A). For both coancestry and inbreeding, the Guadyervas population shows the highest values. In a randomly mating population, the values of f_{ii} and F_i should coincide and α_i measures the magnitude of the possible discrepancy. The negative values of α_i for the Torbiscal and Guadyervas strains indicate that there exists an excess of heterozygous genotypes with respect to the expected value under H–W equilibrium. The parameter G_i is related to α_i and measures the proportion of genetic variability between individuals. With random mating $G = \frac{1}{2}$, lower values indicating that variability is preferentially distributed as variability within individuals, as it happens in the aforementioned strains.

Genetic distances and clusters: The matrix of coancestry values between populations and the values of Reynolds distance are included in Table 3. An unrooted NJ tree showing the relationships among all the populations, evaluated by 1000 bootstrap replicates, was constructed using the Reynolds genetic distances (Figure 1).

For all the Iberian subpopulations, the genetic distance to the Duroc breed is greater than that to any of the other subpopulations of the breed. Another way of visualising this differentiation is to apply the structure algorithm to classify the genome of the 213 genotyped individuals in two clusters. The results appear in Table 4 and 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 Guadyervas strains are the subpopulations whose genomes are differentiated the most unambiguously from Duroc. However, about 4.7% of the genotypes of the Entrepelado variety were classified in the Duroc cluster.

Finally, 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

Table 2 Self-coancestries (f_{ii} , s_i), inbreeding (F_i), H–W deviations (α_i) and proportion of diversity between individuals (G_i) for each population i

	f_{ii}	s_i	D_{ii}	F_i	α_i	G_i
Torbiscal	0.415	0.700	0.285	0.400	-0.026	0.487
Guadyervas	0.556	0.762	0.205	0.523	-0.075	0.463
Retinto	0.343	0.722	0.379	0.443	0.153	0.577
Entrepelado	0.331	0.687	0.351	0.373	0.064	0.532
Lampião	0.322	0.681	0.359	0.362	0.059	0.530
All Iberian	0.394	0.710	0.316	0.420	0.035	0.517
Duroc	0.352	0.725	0.373	0.451	0.152	0.576

Table 3 Values of coancestry (f_{ij} , above the diagonal) and Reynolds' genetic distance (DR_{ij} , below the diagonal) among the populations analysed

	Torbiscal	Guadyervas	Retinto	Entrepelado	Lampião	Duroc
Torbiscal		0.322	0.279	0.272	0.271	0.182
Guadyervas	0.242		0.267	0.262	0.286	0.178
Retinto	0.139	0.249		0.310	0.275	0.193
Entrepelado	0.138	0.246	0.039		0.271	0.203
Lampião	0.133	0.214	0.080	0.076		0.205
Duroc	0.246	0.336	0.191	0.174	0.166	

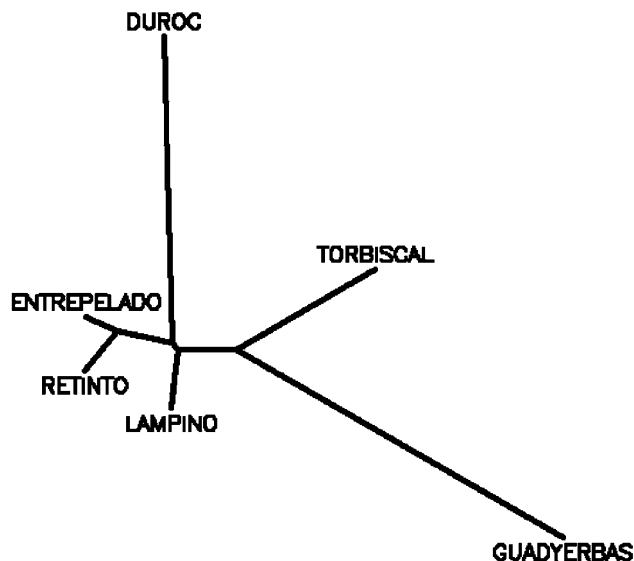


Figure 1 Unrooted NJ dendrogram showing the relationships among five Iberian pig strains and varieties (Torbiscal, Guadyerbás, Negro Lampiño, Entrepelado and Retinto) and the Duroc breed inferred from 36 microsatellite data. This tree is based on Reynolds' genetic distances. Numbers on the nodes are the percentage values for 1000 bootstrap resamplings.

Table 4 Proportion of membership of each predefined population in each of the two possible clusters

Population	Inferred cluster	
	1	2
Torbiscal	0.001	0.999
Guadyerbás	0.001	0.999
Retinto	0.011	0.989
Entrepelado	0.050	0.950
Lampiño	0.010	0.990
Duroc	0.997	0.003

Table 5 Proportion of memberships of each Iberian pig strain or variety assigned to each one of the five possible clusters

Population	Inferred clusters				
	1	2	3	4	5
Torbiscal	0.004	0.003	0.002	0.985	0.006
Guadyerbás	0.001	0.001	0.001	0.002	0.995
Retinto	0.449	0.451	0.009	0.084	0.007
Entrepelado	0.527	0.419	0.008	0.030	0.016
Lampiño	0.321	0.223	0.351	0.024	0.081

the Table 5. They indicate that on average 98.6% of the Torbiscal genomes and 99.5% of Guadyerbás 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 emphasised that the first two strains constitute more defined populations than the others.

Partition of genetic diversity in the Iberian breed: The overall values for the Iberian breed of the different statistics that summarise the genetic diversity were

$$\tilde{f} = 0.393, \tilde{F} = 0.420, \tilde{s} = 0.710, \tilde{f} = 0.304$$

which indicates that the total genetic diversity $GD_T = 1 - \tilde{f} = 0.607$ can be partitioned in the three components: the genetic diversity within individuals, $GD_{WI} = (1 - \tilde{s}) = 0.290$, the genetic diversity between individuals, $GD_{BI} = (\tilde{s} - \tilde{f}) = 0.317$. The sum of these two elements gives the genetic diversity within subpopulations, $GD_{WS} = (1 - \tilde{f}) = 0.607$, and the genetic diversity between subpopulations, $GD_{BS} = (\tilde{f} - \tilde{f}) = 0.089$. The corresponding Wright statistics are $F_{IS} = 0.045$, $F_{ST} = 0.129$, $F_{IT} = 0.167$.

The proportional contribution of each subpopulation to the global coancestry of the Iberian breed is given in Table 6. The Guadyerbás strain contributes most due its own coancestry but, because it shows the highest genetic distance from the other strains, its total contribution is lower than the Retinto strain.

Another question is to ascertain the loss or gain of diversity if one or several subpopulations are removed from the pool. This can be calculated by disregarding one (or more populations) and recalculating the global average coancestry from the remaining pool as shown in Table 7. The removal of the Torbiscal and Guadyerbás strains increases the within-subpopulation genetic diversity, but decreases the average genetic distance. The global net effect would be negative for Torbiscal and positive for Guadyerbás. That means that the removal of the Guadyerbás strain actually will improve the total genetic diversity of the breed, measured as its expected heterozygosity (Nei, 1973). The removal of each one of the three varieties would decrease the within-subpopulation genetic diversity, but would increase the genetic distance, the global balance being negative.

According to the results of the Weitzman approach, also shown in Table 7, the removal of any subpopulation will always decrease the total genetic diversity. The removal of the Guadyerbás strain would have the greatest impact on the decrease of variability followed by the Torbiscal strain. On the other hand, when considering the removal of two subpopulations together, Torbiscal and Guadyerbás seem to be the most important and the Entrepelado and Lampiño varieties the least.

The optimal contributions of each strain to a possible synthetic or germplasm bank can be calculated solving Eq. (4) through quadratic programming to minimise $\lambda f - \bar{D}$. Different λ values weighting the global genetic diversity and the genetic distances have been considered and the results appear in Table 8. The results agree with those of the analysis of genetic diversity. For maximising the global genetic diversity ($\lambda = 1$), the strains that contributed most are Entrepelado and Lampiño, whereas if the objective was to maximise the genetic distance ($\lambda = 0$), Guadyerbás and Torbiscal strains should be prioritised. For $\lambda = 2$, two of the subpopulations would have a null contribution. It is also possible to specify a minimum for the contribution of any strain or variety using the quadratic programming solver. By way of illustration Table 8 includes results with a minimum of $c = 0.02$ for $\lambda = 2$.

Table 6 Proportional contribution of each strain or variety to the global coancestry of the Iberian breed

Strain	Distance within individuals	Distance between individuals	Distance within subpopulation	Distance to other subpopulations	Total
Torbiscal	0.125	0.051	0.074	0.017	0.058
Guadyerbass	0.141	0.038	0.103	0.026	0.077
Retinto	0.208	0.110	0.099	0.019	0.080
Entrepelado	0.119	0.062	0.057	0.012	0.045
Lampiño	0.118	0.062	0.056	0.012	0.043
	$\bar{s} = 0.710$	$\bar{D} = 0.317$	$\bar{f} = 0.393$	$\bar{D} = 0.090$	$\bar{f} = 0.304$

Table 7 Total genetic diversity and loss (–) or gain (+) of diversity when one or two Iberian pig subpopulations are removed

Strain	Within individuals	Between individuals	Within-strain genetic diversity	Average genetic distance	Total genetic diversity	Weitzman method
All strains	0.290	0.317	0.607	0.089	0.696	0.507
Torbiscal (T)	–0.002	+0.008	+0.005	–0.007	–0.002	–0.139
Guadyerbass (G)	+0.013	+0.028	+0.041	–0.035	+0.006	–0.249
Retinto (R)	+0.003	–0.016	–0.013	+0.004	–0.008	–0.046
Entrepelado (E)	–0.006	–0.010	–0.016	+0.005	–0.011	–0.039
Lampiño (L)	–0.007	–0.011	–0.018	+0.005	–0.013	–0.080
T+G	+0.014	+0.048	+0.062	–0.058	+0.003	–0.388
E+L	–0.017	–0.027	–0.045	+0.010	–0.035	–0.119
R+E	–0.004	–0.034	–0.038	+0.003	–0.131	–0.132
G+R	+0.021	+0.016	+0.038	–0.033	+0.004	–0.293

Table 8 Optimal contributions to a synthetic line or to a germplasm bank for different weights of the within- and between-population variability ($\lambda f - \bar{D}$)

Population	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 1$	$\lambda = 2$	$\lambda^* = 2$
Torbiscal	0.228	0.208	0.128	0	0.020
Guadyerbass	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
\bar{f}	0.443	0.424	0.349	0.326	0.332
\bar{D}	0.103	0.101	0.056	0.027	0.036
\bar{f}	0.340	0.323	0.293	0.298	0.297

Discussion

The maintenance of genetic variability is the main objective of conservation programmes. In the present paper, we have been dealing the expected heterozygosity as an adequate measure of genetic diversity, as proposed by Nei (1973). Although this measure is usually applied to molecular markers and refers to identity in state, it must be emphasised that it is equivalent to the classical Malécot (1948) coefficient of coancestry in a model where all the alleles in the base or reference population are assumed to be different. Therefore, it can be applied either to molecular coancestry measured from markers or to genealogical coancestry coming from pedigree information. The basic tools for analysing genetic diversity in subdivided populations have been recently reviewed by Caballero and Toro (2002).

In Conservation Biology there is much discussion about the place of molecular markers in establishing evolutionary significant units (ESU) defined as the population units that merit separate management and have a high priority for conservation. Some authors believe that the increasing availability of molecular

markers permit us to define ESUs as a function of these markers alone and therefore emphasise the role of genetic distances (Moritz, 1994). Other authors consider that ecological and genetic variation of adaptive significance are more relevant for conservation, the key concept being exchangeability: two populations are different ESUs if they cannot be exchanged because they do not share ecological and functional adaptations (Crandall *et al.*, 2000). A parallel controversy exists for the conservation of livestock breeds. As a result of the FAO's recommendation and their availability through genomic mapping, microsatellites have been widely used to characterise genetic diversity. Genetic distances calculated from microsatellite data have been advocated as an objective criterion for making conservation decisions (Barker, 1999; Wimmers *et al.*, 2000). However, other authors have severely criticised this approach and emphasised that aspects such as adaptive features and traits of economic or scientific importance are infrequently available for genetic distance measures. At most, genetic distances could show that populations that perform similarly are actually different but not the other way round (Ponzoni, 1997; Ruane, 1999). The Iberian breed is an interesting case in which to analyse these questions. Its suitability for the sustainable use of the Mediterranean ecosystem, the presence of several varieties and strains and the high quality of its meat and cured products are unusual characteristics, which largely prevent the exchangeability of this pig breed with others.

The above results show the high level of genetic diversity in the Iberian breed, higher than the values reported for other European pig breeds (Laval *et al.*, 2000). The expected heterozygosity exceeds the observed heterozygosity, confirming the subdivision of the breed into strains and varieties. When the subpopulations are analysed separately, Guadyerbass shows less variability in both the expected and observed heterozygosity, as expected from its long maintenance as a closed herd with

a low effective population size (Toro *et al*, 2000). For both the Guadyerbas and the Torbiscal strains, there is an excess of observed *vs* expected heterozygosity ($\alpha_i < 0$) or in other words, more variability is stored within than between individuals ($G_i < 0.50$), which reflects the minimum coancestry mating tactics used in the management of these strains (Toro *et al*, 2000; Fernández *et al*, 2002). The positive values of the other subpopulations and especially the high values of the Retinto variety are more difficult to interpret, apart from their indicating a lack of control of matings.

The Iberian and the Duroc breeds show a clear differentiation according to the values of molecular coancestries and genetic distances (Table 3). The study of distinctiveness of populations has been monopolised by the methodology of genetic distances, but there are other approaches that are probably more efficient. When all the genotyped individuals are classified into two clusters, there are an almost perfect coincidence with the two breeds, the Guadyerbas and Torbiscal strains being those whose genomes are differentiated the most unambiguously from that of Duroc. In fact, any introgression from Duroc into these two strains can be absolutely ruled out, because they have complete purebred pedigree records since 1944 and it is only since 1962 that the first imported Duroc–Jersey pigs have arrived to Spain. The results of the cluster analysis of Iberian subpopulations indicate that almost 99% of individuals from Torbiscal and Guadyerbas strains are classified as two separate clusters, whereas a variable proportion of the other varieties could be attributed to diverse clusters. Guadyerbas and Retinto contribute least to the overall genetic diversity of the breed, due to their own high coancestry even though, especially in the case of Guadyerbas, there is a high genetic distance from the other subpopulations. On the other hand, Entrepelado and Torbiscal are the subpopulations that contribute most, for the opposite reasons: they show the lowest genetic distances to other subpopulations, but also the lowest within-subpopulation coancestry.

Another way of studying the relevance of the different Iberian strains and varieties to the breed diversity is to calculate the loss or gain of diversity if one or several groups are removed and recalculating the global average coancestry (Table 7). The effect on the total genetic diversity of the removal of some groups 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, the gene frequencies are 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. 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.

These results are in sharp contrast with those obtained using the Weitzman approach to calculate the contribution of each subpopulation and the marginal diversity of

each breed. Several authors have criticised the Weitzman approach (Caballero and Toro, 2002; Eding *et al*, 2002). This method does not have a clear genetic interpretation and therefore has properties, (eg that the removal of an element always decreases the variability) 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.

Ignoring 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 (monotonicity 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 noninbred populations with an even distribution of gene frequencies. 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.

The important point is that the results obtained either using the between-population diversity or the total diversity will produce different and sometimes opposite conservation priorities. Therefore, some compromise should be attempted. Thacon d'Arnoldi *et al* (1998) suggest the inclusion, together with the Weitzman method, of the probability of extinction of each population. However, as Eding *et al* (2002) pointed out, this will make the things worse because inbred populations will get an even higher weight. A more practical proposal could be to consider an aggregate diversity, expressed as a linear combination of within- and between-population diversity weighted in an appropriate manner, as suggested by Ollivier and Foulley (2002).

The weights will depend on the scenario imagined for the medium term use of the genetic diversity. If the plan is to use them as part of crossbreeding plans, the diversity between populations should be prioritised 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, the within-population diversity will be important (Notter, 1999). For example, Chaiwong and Kinghorn (1999) suggest giving five times more weight to the variation between breeds than to that within breeds. However arbitrary this figure is, it indicates that some kind of weighting is perceived by several authors as a necessary exercise.

The method for calculating the optimal contribution of populations to a possible synthetic line or to a germplasm bank is also a valuable methodology. Eding *et al* (2002) have used such criterion to establish a core set for 45 poultry breeds. Here, we show the flexibility of this technique (Table 8), which can be utilised by giving different weights to the total diversity and to the genetic distances or by imposing desirable restrictions, as a minimum contribution of each population to the genebank. Similar restrictions could preserve nonautosomal genetic variation that is useful to trace relevant events of

the population history (mitochondrial DNA or Y chromosome lineages).

Throughout the paper we have only considered genetic diversity measured as expected heterozygosity calculated from neutral molecular markers. However, there are other types of genetic variability that can be described and measured in a population. For some loci, diversity is of known importance, such as the histocompatibility system in animals and the self-incompatibility system in plants (Petit *et al*, 1998). The importance of diversity for neutral markers is more questionable.

An additional restriction on the number of alleles could be integrated in the optimisation process. The objective in this case would be to maximise genetic diversity subject to the constraint of maintaining all the distinct alleles already present. Complications could arise if we want to preserve linkage disequilibrium, which is to maintain specific combinations of coadapted alleles that only occur in one breed. More work needs to be carried out on existence of such situations and the strategies to face them.

Finally, other aspects such as including data on particular traits of economic importance, specific adaptive features and local or regional importance of the breed (Oldenbroek, 1999) should be considered in the final conservation decision. In our situation, although the analysis based on the partition of the global coancestry indicates that Torbiscal and Guadyerbas are the sub-populations that contribute least to the total genetic variability, they are probably the most interesting ones according to other criteria, such as the traceability of their genetic origin, the superior productive performance of Torbiscal (Benito *et al*, 2000) and the large scientific knowledge available from these strains (Fernández *et al*, 2002; Ovilo *et al*, 2002).

Acknowledgements

We are grateful to Jesús Fernández for the calculations of optimal contributions, Beatriz Villanueva for the English revision and Javier Cañón and David García for the software to calculate Weitzman distances. Financial support has been provided by the Grant RZ01-028-C2-1. Pig blood samples were kindly supplied by CIA 'Dehesón del Encinar' and AECERIBER.

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.

Benito J, Vázquez C, Menaya C, Ferrera JL, García-Casco JM, Silió L *et al* (2000). Evaluation of productive parameters in different strains of Iberian pig. In: Afonso JA, Tirapicos JL (eds) *Tradition and Innovation in Mediterranean Pig Production*. CIHEAM/ICAM-UE: Zaragoza, pp 113–121.

Caballero A, Toro MA (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Consero 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.

Chaiwong N, Kinghorn BP (1999). Use of genetic markers to aid conservation decisions for groups of rare domestic breeds. *Proc Assoc Adv Anim Breed Genet* **13**: 365–368.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000). Considering evolutionary processes in conservation biology. *TREE* **15**: 290–295.

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.

Eding H, Laval G (1999). Measuring genetic uniqueness in livestock. In: Oldenbroek JK (ed) *Genebanks and the Conservation of Farm Animal Genetic Resources*. DLO Institute for Animal Sciences and Health: Lelystad, The Netherlands, pp 33–58.

Excoffier L (2001). Analysis of population subdivision. In: Balding DJ, Bishop M, Cannings C (eds) *Handbook of Statistical Genetics*. John Wiley: New York, pp 271–307.

FAO (1998). Secondary guidelines for development of national farm animal genetic resources management plan. Measurement of domestic animal diversity (MoDAD); Working report. Available at: <http://dad.fao.org> Accessed Dec. 15, 2000.

Felsenstein J (2001). *PHYLIP Phylogeny inference Package, Version 3.6 (alpha2)*. Department of Genetics, University of Washington: Seattle.

Fernández A, Rodríguez J, Toro MA, Rodríguez MC, Silió L (2002). Inbreeding effects on the parameters of the growth function in three strains of Iberian pigs. *J Anim Sci* **80**: 2267–2275.

Hall SJG, Bradley DG (1995). Conserving livestock breed diversity. *TREE* **10**: 267–270.

Iannuccelli E, Wolosyn N, Arhains J, Gellin J, Milan D (1996). GEMMA: a database to manage and automate microsatellite genotyping. *Proceedings of the 25th International Conference on Animal Genetics, Tours*, p 88.

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.

Lemus-Flores C, Ulloa-Arvizu R, Ramos-Kuri M, Estrada FJ, Alonso RA (2001). Genetic analysis of Mexican hairless pig populations. *J Anim Sci* **79**: 3021–3026.

López-Bote CJ (1998). Sustained utilization of the Iberian pig breed. *Meat Science* **49**(Suppl 1): S17–S27.

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

Moritz C (1994). Defining 'evolutionary significant units' for conservation. *Trends Ecol Evol* **9**: 373–375.

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.

Ovilo C, Clop A, Noguera JL, Oliver MA, Barragán C, Rodríguez MC *et al* (2002). QTL mapping for meat quality traits in an Iberian × Landrace F₂ pig population. *J Anim Sci* **80**: 2801–2808.

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

Ponzoni RW (1997). Genetic resources and conservation. In: Piper LR, Ruvinsky A (eds) *The Genetics of Sheep*. CAB International: New York, pp 433–466.

- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rodríguez J, Toro MA, Rodríguez MC, Silió L (2000). Alleles survival from Portuguese and Spanish strains in a population of Iberian pig. In: Afonso JA, Tirapicos JL (eds) *Tradition and Innovation in Mediterranean Pig Production*. CIHEAM/ICAM-UE: Zaragoza, pp 57–61.
- 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.
- Silió L (2000). Iberian pig breeding programme. In: Galal S, Boyazouglu J, Hammond K (eds) *Developing Breeding Strategies for Lower Input Animal Production Environments*. International Committee for Animal Recording: Roma, pp 511–520.
- Thaon d'Arnoldi C, Foulley J-L, Ollivier L (1998). An overview of the Weitzman approach to diversity. *Genet Sel Evol* **30**: 149–161.
- Toro MA, Rodríguez J, Silió L, Rodríguez C (2000). Genealogical analysis of a closed herd of black hairless Iberian pigs. *Conserv Biol* **14**: 1843–1851.
- Vaughan HW (1950). *Breeds of Live Stock in America*. Long's College Book Co.: Columbus, OH.
- Weitzman ML (1992). On diversity. *Quart J Econom* **107**: 363–405.
- Wimmers K, Ponsuksili S, Hardge T, Valle-Zarate A, Mathur PK, Horst P (2000). Genetic distinctness of African, Asian and South American local chickens. *Anim Genet* **31**: 159–165.
- Wright S (1969). *Evolution and the Genetics of Populations. Vol. 2, The Theory of Gene Frequencies*. The University of Chicago Press: Chicago.

Appendix A

Let us consider three populations ($n=3$) with 3, 2 and 2 individuals genotyped for a codominant marker. The coefficient of molecular coancestry between all the individuals are given in Table A1. It is obtained by applying Malécot definition but referring to identity in state instead of identity by descent.

Adding the appropriate cells of the table, we can easily obtain the values of the average coancestry between subpopulations as

$$f_{ij} = \begin{bmatrix} 0.556 & 0.500 & 0.167 \\ 0.500 & 0.500 & 0.125 \\ 0.167 & 0.125 & 0.625 \end{bmatrix}$$

The average self-coancestries of each population, s_i , will add only the diagonal values and the average inbreeding

Table A1 Molecular coancestries for a hypothetical population of seven individuals

	A_1A_1	A_1A_2	A_1A_2	A_1A_1	A_2A_2	A_1A_2	A_3A_3
A_1A_1	1	1/2	1/2	1	0	1/2	0
A_1A_2	1/2	1/2	1/2	1/2	1/2	1/4	0
A_1A_2	1/2	1/2	1/2	1/2	1/2	1/4	0
A_1A_1	1	1/2	1/2	1	0	1/2	0
A_2A_2	0	1/2	1/2	0	1	0	0
A_1A_2	1/2	1/4	1/4	1/2	0	1/2	1/2
A_3A_3	0	0	0	0	0	1/2	1

will correspond to $F_i = 2s_i - 1$

$$s_i = \begin{bmatrix} 0.667 \\ 1 \\ 0.750 \end{bmatrix}, \quad F_i = \begin{bmatrix} 0.333 \\ 1 \\ 0.500 \end{bmatrix}$$

The average distance between individuals of subpopulations i and j is

$$D_{ij} = \left[\frac{s_i + s_j}{2} - f_{ij} \right] = \begin{bmatrix} 0.111 & 0.333 & 0.542 \\ 0.333 & 0.500 & 0.750 \\ 0.542 & 0.750 & 0.125 \end{bmatrix}$$

And the classical Nei's minimum and the Reynolds distances will be

$$D_{ij} = D_{ij} - \frac{D_{ii} + D_{jj}}{2} \\ = \frac{f_{ii} + f_{jj}}{2} - f_{ij} = \begin{bmatrix} 0 & 0.028 & 0.424 \\ 0.028 & 0 & 0.437 \\ 0.424 & 0.437 & 0 \end{bmatrix}$$

$$\mathcal{DR}_{ij} = \frac{D_{ij}}{1 - f_{ij}} = \begin{bmatrix} 0 & 0.056 & 0.508 \\ 0.056 & 0 & 0.500 \\ 0.508 & 0.500 & 0 \end{bmatrix}$$

For subpopulation I , the following relationships are met:

$$\alpha_i = \frac{F_i - f_{ii}}{1 - f_{ii}}, \quad G_i = \frac{s_i - f_{ii}}{1 - f_{ii}} = \frac{D_{ii}}{1 - f_{ii}}$$

Here, α_i is the deviation from H–W proportions and G_i is the proportion of diversity between individuals for subpopulation i , so that $(1 - \alpha_i) = 2(1 - G_i)$ and $(1 - f_{ii}) = (1 - s_i) + D_{ii}$.

For the entire population

$$\bar{f} = \frac{\sum_{i=1}^n \sum_{j=1}^n f_{ij}}{n} = 0.560 \\ \tilde{f} = \frac{\sum_{i=1}^n f_{ii}}{n} = 0.560$$

$$\tilde{s} = \frac{\sum_{i=1}^n s_i}{n} = 0.806 \\ \tilde{F} = \frac{\sum_{i=1}^n F_i}{n} = 0.611$$

$$\bar{D} = \frac{\sum_{i,j=1}^n D_{ij}}{n^2} = \tilde{f} - \bar{f} = 0.197$$

and finally the average global coancestry

$$\bar{f} = \frac{\sum_{i=1}^n \sum_{j=1}^n f_{ij}}{n^2} = \frac{\sum_{i=1}^n f_{ii}}{n} - \bar{D} \\ = \sum_{i=1}^n \frac{1}{n} \left[f_{ii} - \frac{\sum_{j=1}^n D_{ij}}{n} \right] = 0.363$$

The last expression allows to assign the contribution of each population to the global coancestry either to its own coancestry ($f_{ii} = s_i - D_{ii}$) or to the genetic distances to other populations.

It should also be noted that Wright's (1969) F-statistics are obtained as

$$\begin{aligned}
 F_{IS} &= \frac{\tilde{F} - \tilde{f}}{1 - \tilde{f}}, & F_{ST} &= \frac{\tilde{f} - \bar{f}}{1 - \bar{f}} \\
 &= \frac{\tilde{D}}{1 - \bar{f}} & \text{and } F_{IT} &= \frac{\tilde{F} - \bar{f}}{1 - \bar{f}}
 \end{aligned}
 \tag{3}$$

so that $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$

All the above measures are calculated for each marker locus and averaged over loci. There is an alternative way of calculating the above results that will lead to exactly the same figures (Caballero and Toro, 2002). It is also straightforward to modify the above formulae to assign different weights to subpopulations, although it is not clear as to which weights should be applied. From an

evolutionary perspective, Nei (1973) suggests equal weights, but other authors suggest that it is more reasonable for descriptive purposes to weigh each population proportionally either to the effective or to the actual population size. Finally, because the studied individuals of a subpopulation can be considered as a sample (or the subpopulations studied as a sample of a larger set of subpopulations), some corrections should be included to allow for the finite sample size. There is, however, some controversy about what is the appropriate correction (Excoffier, 2001). In the present work, we show the results ignoring these corrections in order to maintain the simplicity of the analysis and, since the numbers of sampled individuals per population are approximately equal, the results are essentially the same.