

## Genetic diversity and allelic richness in Spanish wild and domestic pig population estimated from microsatellite markers

J. Rodrigáñez<sup>1</sup>, C. Barragán<sup>1</sup>, E. Alves<sup>1</sup>, C. Gortázar<sup>2</sup>,  
M. A. Toro<sup>1</sup> and L. Silió<sup>1\*</sup>

<sup>1</sup> *Departamento de Mejora Genética Animal. INIA. Ctra. A Coruña, km 7,5. 28040 Madrid. Spain*

<sup>2</sup> *Instituto de Investigación en Recursos Cinegéticos. CSIC-UCLM.  
Ronda de Toledo, s/n. 13071 Ciudad Real. Spain*

---

### Abstract

Genetic diversity measures support the conservation decisions aiming to maintain the genetic flexibility of animal populations and breeds. Many of these measures utilize neutral genetic markers and are based on classical concepts as coancestry and expected heterozygosity. As a component of genetic diversity, allelic richness is also important in conservation genetics. Its measurement requires that variations of sample size be taken into account using either the rarefaction and extrapolation techniques. Methods to estimate genetic diversity and allelic richness were compared in this study. DNA samples from 68 wild boars and 234 domestic pigs of the Duroc and Iberian breeds, including 63 animals of the Torbiscal and Guadyerbas Iberian lines, were genotyped for 18 microsatellites (one per autosome). As the results outline, the rank of these pig populations according to their contributions to the diversity will be different depending on the criteria utilized, because genetic diversity and private allelic richness do not exactly address the same type of diversity. Rarefaction and extrapolation-based techniques also produce partially discrepant results. The desirable integration of allelic richness into the diversity theory poses at the moment some unsolved difficulties.

**Additional key words:** Duroc, extrapolation, Iberian pig, microsatellites, private alleles, rarefaction, wild boar.

### Resumen

#### Diversidad genética y riqueza alélica en poblaciones españolas de cerdos y jabalíes estimadas a partir de marcadores microsatélite

Diversas medidas de diversidad genética sustentan las tácticas conservacionistas que procuran mantener la flexibilidad genética de poblaciones animales y razas domésticas. Muchas de estas medidas utilizan marcadores genéticos neutrales y se basan en conceptos clásicos como el parentesco y la heterocigosidad esperada. Como componente de la diversidad genética, la riqueza alélica tiene asimismo importancia en la genética de la conservación. Su medida requiere tener en cuenta las variaciones de tamaño muestral mediante técnicas de rarefacción o extrapolación. En este estudio se han comparado métodos de análisis de diversidad genética y riqueza alélica. Se genotiparon para 18 microsatélites (uno por autosoma) muestras de ADN de 68 jabalíes y 234 cerdos domésticos de las razas Duroc e Ibérico, incluyendo 63 animales de las líneas de cerdo Ibérico Torbiscal y Guadyerbas. Los resultados indican que la ordenación de estas poblaciones de acuerdo a sus contribuciones a la diversidad total se modifica en función de los criterios empleados, ya que diversidad genética y riqueza de alelos privados no contemplan el mismo tipo de diversidad. Asimismo, las técnicas empleadas de rarefacción o extrapolación dan lugar a resultados parcialmente discrepantes. La deseable integración de la riqueza alélica en la teoría de la diversidad genética plantea algunas dificultades no resueltas por el momento.

**Palabras clave adicionales:** alelos privados, cerdo Ibérico, Duroc, extrapolación, jabalí, microsatélites, rarefacción.

---

\* Corresponding author: [silio@inia.es](mailto:silio@inia.es)

Received: 16-07-07; Accepted: 05-12-07.

## Introduction<sup>1</sup>

Molecular markers can be used to explore the diversity present at the level of the individual, population, breed and species, in both wild and domestic animals (Frankham *et al.*, 2002). Most of the analyses of genetic diversity at the breed and population level have used microsatellite markers, whose genomic locations have been defined by genetic mapping, and which are highly variable and available to define diversity across the entire genome. The application of technologies such as multi-locus microsatellite genotyping, and its potential impact on the management of genetic diversity are extremely important challenges both for geneticists and breed managers (Bruford, 2004).

Phylogenetic techniques based on genetic distances estimated from microsatellites have been the method of choice to assess the genetic diversity of livestock breeds (Barker, 1999). Thaon d'Arnoldi *et al.* (1998) emphasized 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, and based on these results, conservation priorities can be determined. But, the methods employed for calculating genetic distances to infer relationships within and among breeds are poorly suited for managing within-breed genetic diversity and for setting between breeds conservation priorities (MacHugh *et al.*, 1998). Genetic distances were developed with the concept of species in mind, and ignore important features of livestock populations: migration, short divergence period, and small role of mutation in creating genetic differences, which are mainly assignable to selection and genetic drift (Laval *et al.*, 2002). The high polymorphism of microsatellites makes them extremely sensitive to changes in effective population size and they can therefore accumulate major changes in allele frequencies very rapidly. Caballero and Toro (2002) consider these methods inappropriate for within-species breed conservation, because genetic variation within groups is ignored in this approach, although it may be of great importance for the management of livestock breeds. Other statistical tools, focused to the partition of genetic diversity within and between breeds, and based on the classical concepts of coancestry and expected heterozygosity, have been proposed to analyse genetic diversity in populations of livestock species (Toro *et al.*, 2007).

As a component of genetic diversity, allelic richness is an alternative criterion to measure genetic diversity, and some authors consider that this parameter is of key relevance in conservation programmes (Petit *et al.*, 1998; Simianer, 2005; Foulley and Ollivier, 2006a). Allelic diversity is particularly important from a long-term perspective, because the limit of selection response is mainly determined by the initial number of alleles regardless of the allelic frequencies (Hill and Rasbash, 1986) and, because it reflects better past fluctuations in population size. Observed allelic richness needs correction for the different sample size of populations using the rarefaction and/or extrapolation techniques (El Mousadik and Petit, 1996; Foulley and Ollivier, 2006a). Allelic diversity and gene diversity can behave rather differently, for example, a higher differentiation between populations can be observed for allelic diversity than for gene diversity (Foulley and Ollivier, 2006a). The objective of this study was to apply these new statistical tools for assessing the genetic and allelic diversity in a set of wild boar and domestic pig populations, and for exemplifying the difficulties of their possible application to the definition of conservation priorities.

## Material and Methods

### Genotypes

We genotyped DNA samples from 68 wild boars collected in diverse regions of the North, Center and South of Spain, and from 234 domestic pigs of the Iberian and Duroc breeds. The 170 sampled Iberian pigs were inscribed in the breed herdbook, and represent the closed lines Torbiscal (31) and Guadyrbas (32), preserved from the years 1944/45 in «El Dehesón del Encinar» (Oropesa, Toledo), and 15 private or public breeding nuclei of the Iberian breed (107). The 64 sampled Duroc pigs represent seven Spanish breeding nuclei, founded with animals of diverse origins (Canada, USA, Hungary, Denmark) within this cosmopolitan breed.

All the individuals were genotyped for 18 microsatellites (one on each autosome) chosen according to their reproducibility, high polymorphism and absence of null alleles. Details of DNA extraction, PCR-reactions and microsatellite analysis are giving in Alves *et al.* (2006). Genomic DNA was extracted using standard

<sup>1</sup> Abbreviations used: DRE (allelic richness dissimilarities estimated by extrapolation), DRR (allelic richness dissimilarities estimated by rarefaction).

protocols from blood samples (live domestic pigs) or from lymph node or ear samples (hunted wild boars).

### Analysis of genetic diversity

According to Caballero and Toro (2002), in a meta-population subdivided in  $n$  breeds or populations, the total genetic diversity or expected heterozygosity ( $H_T$ ) is partitioned into a component within breeds ( $H_S$ ) and another ( $H_T - H_S$ ) between breeds:

$$H_S = 1 - \frac{1}{n} \sum_i \left( \sum_k p_{i,k}^2 \right), \quad H_T = 1 - \sum_k \left( \sum_i \frac{p_{i,k}}{n} \right)^2,$$

where  $p_{i,k}$  is the frequency of allele  $k$  in breed  $i$ . Both results are averaged across loci. The between breed component of genetic diversity ( $H_T - H_S$ ) is also the average Nei's minimum distance between populations ( $\bar{D}$ ). Wright's (1969) fixation index is simply the proportion of diversity between breeds relative to the total diversity:  $F_{ST} = (H_T - H_S) / H_T = \bar{D} / H_T$ . Recently, Hedrick (2005) has proposed to standardise it by the maximum level that can be obtained,  $F_{ST(max)} = (1 - H_S) / (1 + H_S)$ , given the observed heterozygosity within breeds. Thus,  $F_{ST}^2 = F_{ST} / F_{ST(max)}$  is a measure of population differentiation relative to the maximum possible.

The contributions of each one of these populations to the whole population diversity can be also specified (see Caballero and Toro, 2002). Another way of studying the relevance of the different breeds or populations to the overall diversity is, following Petit *et al.* (1998), to calculate the loss or gain of diversity if one population is removed from the whole set.

### Analysis of allelic diversity

Let  $N_i$  being the total number of genes sampled from the breed  $i$  at one locus, i.e. twice the number of animals sampled from that breed for that locus, and  $N_{ik}$  the number of copies found of the  $k$ th allele in this sample of size  $N_i$ . The allelic richness at one locus estimated by rarefaction is denoted as the expected number of different alleles that a sample had if the sample size had been  $g$  genes instead of  $N_i$  ( $g \leq N_i$ ). The estimation of allelic richness  $R_i$  through rarefaction method is then,

$$R_i = \sum_k (1 - P_{ik}),$$

$$\text{where } P_{ik} = \frac{\binom{N_i - N_{ik}}{g}}{\binom{N_i}{g}}$$

is the probability that allele  $k$  does not occur in a sample of  $g$  genes chosen as reference, being the numerator the number of combinations of  $g$  genes that do not include allele  $k$  and the denominator the number of possible combinations of  $g$  genes taken from the sample of  $N_i$  genes. An extension of the rarefaction method to count private alleles (alleles present in a given breed but absent from all the others) was defined

by Kalinowski (2004) as  $PR_i = \sum_k (1 - P_{ik}) \prod_{j \neq i} P_{jk}$ .

Foulley and Ollivier (2006a) proposed an alternative method, based on extrapolation, that consists of adding to the number of alleles actually observed in a sampled population the expected number of alleles missing, given the number of genes examined in the sample and the allelic frequencies observed in the whole set of populations. If  $K_i$  is the number of alleles in the sample of population  $i$ ,  $K$  is the total number of alleles sampled in the whole set of populations and  $\pi_k$  the frequency of the  $k$  allele in the whole population, the allelic richness of breed  $i$  with a sample size  $N_i$  estimated by extrapolation

would be:  $R_i = K_i + \sum_k (1 - \pi_k)^{N_i}$ , where the summation

is over the subset of alleles actually missing in the sample. Therefore, the allelic richness for each population obtained with this method will be a value between the actual number of different alleles sampled in the population,  $K_i$ , and the total number of alleles in the whole set of populations,  $K$ .

A partition of diversity within and between populations can also be made considering allelic diversity instead of gene diversity. Foulley and Ollivier (2006a) proposed a partition analogous to that advocated by Petit *et al.* (1998) for gene diversity. Irrespective of the correction for sampling size in the allelic richness of populations, the contribution of each population  $i$  to the total allelic richness would be  $CT_i = TR - R(S/i)$ , where  $TR$  is the allelic richness of the whole set of populations considered as a single one, and  $R(S/i)$  is the allelic richness of this whole set excluding population  $i$ ; notice that  $TR = K$  when the extrapolation method is applied.  $R(S/i)$  is estimated by considering as a single population the subset  $S$  of  $n-1$  populations where breed  $i$  is excluded, and then calculating for  $S$ , either by rarefaction or by extrapolation, the allelic richness as explained before. Both, the extrapolation-based contribution ( $CTE_i$ ) and

the rarefaction-based contribution ( $CTR_i$ ) may be considered as estimations of the number of private alleles in the population  $i$ , corrected for the sample size. The extrapolation-based contribution of breed  $i$  to the within-breed allelic richness would be defined as  $CWE_i = (R_i - \bar{R}) / (n - 1)$ , where  $R_i$  and  $\bar{R}$  are respectively the allelic richness of population  $i$  and the average allelic richness of the whole set of populations, both calculated by the extrapolation technique. Similarly for  $CWR_i$  if  $R_i$  and  $\bar{R}$  are calculated by the rarefaction technique.

A measure of allelic richness dissimilarity (or distance) may be defined as the expected number of alleles present in a population  $i$  and absent in population  $j$  calculated by the rarefaction technique

$$d_R(i, j) = \sum_k (1 - P_{ik}) P_{jk}, \quad \text{where the summation is over}$$

all observed alleles of  $i$  and  $j$ . A similar distance can be derived from the extrapolation-based allelic richness (Foulley and Ollivier, 2006b). Both rarefaction and extrapolation-based dissimilarities are not symmetric, since in general  $d(i, j) \neq d(j, i)$ . Irrespective of the correction technique, symmetric dissimilarities may be obtained by expressing  $d(i, j)$  as the difference between allelic richness of a population ( $R_i$ ) minus

joint allelic richness of  $i$  and  $j$  ( $R_{ij}$ ), and similarly for  $d(j, i)$ .  $R_{ij}$  is calculated either through rarefaction or extrapolation method when considering  $i$  and  $j$  together as a single population. Dissimilarity is obtained as  $1 - S = (R_i + R_j - 2R_{ij}) / (R_i + R_j - R_{ij})$ .

## Results and Discussion

Table 1 shows the number of detected alleles for each population and genotyped microsatellite arranged by pig autosomes. This number varies from 1 to 12 (average 5.3 alleles) for the total set of populations. The number of alleles is higher in the wild boar population and Duroc and Iberian breeds. It is lower in the Guadyerbias and Torbiscal closed lines of the Iberian breed. Note that these two related lines have been preserved in a conservation programme established at the years 1944-1945, from four Spanish and Portuguese ancient Iberian strains. Torbiscal is the result of blending the four ancient founder populations (Fernández *et al.*, 2002), and Guadyerbias is one of these founder populations (Toro *et al.*, 2000). Table 1 also gives, for each locus and population, the observed heterozygosity by direct count ( $H_O$ ) and the expected heterozygosity ( $H_E$ ) under Hardy-Weinberg equilibrium. In wild boar,

**Table 1.** Number of alleles ( $N_A$ ) and observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosities of microsatellite markers in the wild boar population ( $n = 68$ ), Duroc ( $n = 64$ ) and Iberian ( $n = 107$ ) pig breeds and Guadyerbias ( $n = 32$ ) and Torbiscal ( $n = 31$ ) Iberian lines

Chr	Marker	Wild boar			Duroc			Iberian			Guadyerbias			Torbiscal		
		$N_A$	$H_O$	$H_E$	$N_A$	$H_O$	$H_E$	$N_A$	$H_O$	$H_E$	$N_A$	$H_O$	$H_E$	$N_A$	$H_O$	$H_E$
1	S0155	4	0.571	0.659	4	0.672	0.593	5	0.458	0.518	3	0.594	0.533	2	0.258	0.275
2	SW240	6	0.647	0.749	7	0.797	0.774	7	0.673	0.768	3	0.625	0.595	3	0.290	0.256
3	S0002	9	0.701	0.848	5	0.672	0.752	8	0.589	0.682	2	0.562	0.500	4	0.710	0.718
4	S0097	6	0.500	0.770	7	0.531	0.576	12	0.729	0.838	4	0.625	0.549	7	0.774	0.722
5	SW413	6	0.448	0.562	5	0.484	0.765	4	0.598	0.726	2	0.500	0.490	3	0.581	0.574
6	SW1057	10	0.647	0.597	5	0.453	0.723	8	0.692	0.709	2	0.625	0.476	5	0.548	0.636
7	SW632	9	0.382	0.483	7	0.469	0.520	10	0.654	0.750	1	0.000	0.000	3	0.710	0.654
8	S0178	7	0.515	0.686	7	0.656	0.712	9	0.673	0.744	4	0.500	0.441	4	0.645	0.668
9	SW911	5	0.333	0.487	6	0.656	0.678	7	0.617	0.750	2	0.187	0.173	4	0.677	0.690
10	S0070	12	0.578	0.857	10	0.609	0.625	11	0.729	0.842	4	0.719	0.741	3	0.645	0.593
11	S0385	5	0.573	0.740	6	0.734	0.770	5	0.682	0.719	1	0.000	0.000	4	0.548	0.563
12	SW874	9	0.758	0.732	9	0.578	0.783	7	0.607	0.792	4	0.750	0.651	4	0.323	0.548
13	S0219	4	0.493	0.505	4	0.359	0.472	4	0.579	0.690	1	0.000	0.000	3	0.613	0.674
14	SW210	10	0.797	0.855	4	0.453	0.555	5	0.673	0.692	2	0.031	0.031	3	0.581	0.600
15	SW936	6	0.492	0.651	5	0.641	0.723	8	0.673	0.775	4	0.844	0.744	6	0.744	0.730
16	S0026	7	0.588	0.757	5	0.562	0.667	6	0.308	0.418	2	0.500	0.490	2	0.290	0.297
17	SW24	7	0.618	0.715	7	0.719	0.763	6	0.598	0.711	2	0.344	0.424	3	0.774	0.629
18	SW787	5	0.463	0.647	7	0.797	0.812	6	0.626	0.685	3	0.500	0.506	4	0.677	0.701
	Average		0.561	0.683		0.602	0.681		0.620	0.712		0.439	0.408		0.579	0.585

**Table 2.** Contribution of each pig population to the within-breed ( $H_S$ ), between-breed ( $\bar{D}$ ) and total genetic diversity ( $H_T$ ). Relative contributions of each breed (in %) between brackets

Population	Within-breed diversity	Between-breed diversity	Total genetic diversity
Wild boar	0.136 (23)	0.021 (17)	0.156 (21)
Duroc	0.135 (22)	0.030 (25)	0.165 (23)
Iberian	0.142 (23)	0.017 (14)	0.159 (22)
Guadyrbas	0.081 (13)	0.031 (25)	0.112 (15)
Torbiscal	0.115 (19)	0.023 (19)	0.138 (19)
	$H_S = 0.609$	$\bar{D} = 0.122$	$H_T = 0.730$

Duroc and Iberian, the  $H_O$  values were generally lower than the  $H_E$  values, indicating deviations from the random mating, due to the isolation among the different breeding nuclei of each pig breed or among the diverse territories of wild boar sampling.

### Analysis of genetic diversity

The total heterozygosity of the whole set of populations is  $H_T = 0.730$ , with components within breeds  $H_S = 0.609$  and between breeds  $\bar{D} = 0.122$ . The amount of differentiation is  $F_{ST} = 0.167$ , the maximum possible being  $F_{ST(max)} = 0.243$ , so that the standardized value is in this case  $F'_{ST} = 0.687$ .

The contributions of each one of these populations or breeds to the whole population diversity are shown in Table 2. For example, the pigs representing the Iberian breed contribute the most (23%) to the within-population diversity component ( $H_S$ ), but relatively little (14%) to the between-population component ( $\bar{D}$ ). In contrast, the Guadyrbas line contributes the least to the within-population variation (13%), but substantially (25%) to the averaged between-population variation.

The loss or gain of genetic diversity when one population is removed from the whole set, expressed as percentages of the recalculated genetic diversity, are presented in Table 3. For example, the removal of Iberian

breed causes a decrease of 4.1% in the within-population variation, but an increase of 11.6% in the between-population diversity. In contrast, the removal of the genetically isolated Guadyrbas line involves a gain in the total (3.0%) and in the within-population variation (8.5%), but a substantial loss (24.0%) in the between-population component. Although it may seem paradoxical that the removal of one population could increase total genetic diversity, we must realize that, if owing to the removal of that population gene frequencies become more equalized, then this will increase the expected heterozygosity. The population with a higher rank for genetic diversity in this particular scenario turns out to be the Duroc breed, which removal implies a loss of both within (2.9%) and between-population (19.3%) genetic diversity. The removal of wild boar, who has not been submitted to artificial selection, implies also a loss of 2.9% of within population genetic diversity despite of being in the second position in the rank.

### Analysis of allelic diversity

The respective observed values of allelic richness and private allelic richness for the different populations are shown in Table 4. The observed numbers of alleles reflect events of the history of each studied population

**Table 3.** Loss (-) and gain (+) in percentage (%) of genetic diversity when each population is removed from the whole set

Population removed	Within-breed diversity	Between-breed diversity	Total genetic diversity	Rank of genetic diversity
Wild boar	-2.9	+3.5	-1.8	2
Duroc	-2.9	-19.3	-5.6	1
Iberian	-4.1	+11.6	-1.5	3
Guadyrbas	+8.5	-24.0	+3.0	5
Torbiscal	+1.3	-3.0	+0.6	4

**Table 4.** Allelic richness ( $R$ ), and private allelic richness ( $PR$ ) of 18 microsatellite loci in the wild boar population, Duroc and Iberian breeds and Guadyerbias and Torbiscal Iberian lines: total number of observed alleles ( $R_T$ ) and private alleles ( $PR_T$ ), and mean number of alleles and private alleles observed per locus ( $R_O$ ,  $PR_O$ ), estimated by rarefaction ( $R_R$ ,  $PR_R$ ) with uniform sample sizes of  $g=62$ , and estimated by extrapolation ( $R_E$ ,  $PR_E$ )

Population	Allelic richness ( $R$ )				Private allelic richness ( $PR$ )			
	$R_T$	$R_O$	$R_R$	$R_E$	$PR_T$	$PR_O$	$PR_R$	$PR_E$
Wild boar	127	7.055	6.293	7.720	24	1.333	1.316	1.149
Duroc	110	6.111	5.608	6.972	16	0.889	1.107	0.746
Iberian	128	7.111	6.060	7.810	6	0.333	0.466	0.285
Guadyerbias	46	2.555	2.547	4.668	1	0.056	0.083	0.056
Torbiscal	67	3.722	3.722	5.804	1	0.056	0.095	0.033

or breed, which largely diverge in their genetic origin, effective census, and the impact of selection, introgression and genetic isolation. The most noticeable result is the high number of private alleles actually observed in the wild boar and Duroc populations: 24 and 16, respectively. In spite of its similar observed allelic richness and greater sampling size, the Iberian breed present only six private alleles. It may be explained by the presence among the considered populations of other two Iberian lines, which share alleles present in the genetic pool of the Iberian breed and absent in the wild boar and Duroc populations.

But allelic richness also depends on sample size, and their observed values need to be corrected by rarefaction or extrapolation techniques. Note that the rarefaction method estimates the allelic richness for a uniform number of sampled genes ( $g$ ), and can be applied only when there are inequalities in the sample sizes. In the present analysis,  $g$  coincides with the smallest sample size (Torbiscal line,  $g=62$ ), and the application of the rarefaction method clearly reduces the allelic richness of the other populations: the estimated  $R_R$  values are always lower than the observed  $R_O$  values. This general reduction of the number of alleles is associated to an increase of the mean number of private alleles per locus ( $PR_R$ ) in four of the five studied populations (Table 4).

As expected, the use of the extrapolation method provides estimates clearly exceeding the observed richness: the  $R_E$  values are always greater than  $R_O$  values. This increase of the number of alleles is related to a diminution of the number of private alleles per locus ( $PR_E$ ) in each population (Table 4). An assumption of this approach is that all the populations or breeds are drawn at random from the same founder population (Foulley and Ollivier, 2006a). Thus, if one allele is missing in a given population, the unique reason is its low sampling size, ignoring the possibility that it could be truly lost by genetic drift or other genetic causes. This can be a questionable assumption for analyzing livestock breeds.

Irrespective of the correction for sampling size, the definition of the contribution of each population  $i$  to the allelic richness [ $CT_i = TR - R(S/i)$ ] implies that only those populations with private alleles contribute to the allelic diversity. The results of the partition of these contributions within and between populations are presented in Table 5, showing that wild boar, Duroc and Iberian populations contribute positively to the total private allelic richness, whereas the contributions of the populations with lower number of private alleles (Torbiscal and Guadyerbias) are close to zero ( $CT_O$  and  $CT_E$ ) or even negative ( $CT_R$ ). This strong discrepancy between the results of rarefaction and extrapolation methods may be also appreciated for the  $CW_R$  and  $CW_E$

**Table 5.** Population contributions to total, within-population and between-population private allelic richness: observed ( $CT_O$ ,  $CW_O$ ,  $CB_O$ ), estimated by rarefaction ( $CT_R$ ,  $CW_R$ ,  $CB_R$ ) and estimated by extrapolation ( $CT_E$ ,  $CW_E$ ,  $CB_E$ )

Population	$CT_O$	$CW_O$	$CB_O$	$CT_R$	$CW_R$	$CB_R$	$CT_E$	$CW_E$	$CB_E$
Wild boar	1.333	0.436	0.897	0.690	0.362	0.328	1.149	0.281	0.868
Duroc	0.889	0.200	0.689	0.730	0.191	0.539	0.746	0.094	0.652
Iberian	0.333	0.450	-0.117	0.165	0.304	-0.139	0.285	0.304	-0.019
Guadyerbias	0.055	-0.689	0.744	-0.079	-0.575	0.496	0.055	-0.482	0.537
Torbiscal	0.056	-0.397	0.453	-0.039	-0.281	0.242	0.033	-0.198	0.231

**Table 6.** Correlations between Reynolds ( $D_{\text{REY}}$ ) and Nei ( $D_{\text{NEI}}$ ) genetic distances and allelic richness dissimilarities<sup>1</sup> calculated by rarefaction (DRR) and extrapolation (DRE) techniques among the five pig populations

	$D_{\text{NEI}}$	DRR	DRE
$D_{\text{REY}}$	0.976	0.170	0.798
$D_{\text{NEI}}$		0.316	0.855
DRR			0.577

<sup>1</sup> Transformed to absolute values of the estimated DRR and DRE.

values shown in Table 5. Moreover, wild boar, Duroc and Iberian populations contribute positively to the within-breed allelic richness, whereas Torbiscal and Guadyerbas contribute negatively. Analogously, the between-breed contribution can be obtained as  $CB_i = CT_i - CW_i$ , but the meaning of this parameter is unclear (Toro *et al.*, 2007). The  $CB_O$ ,  $CB_R$  and  $CB_E$  values are also presented in Table 5, but these results seem to lack an intuitive justification.

Finally, the allelic richness symmetric dissimilarities ( $1 - S$ ) among the pig populations were calculated by rarefaction (DRR) and extrapolation (DRE) techniques. Genetic distances were also calculated on the same populations, using either the standard Reynolds distance (Reynolds *et al.*, 1983) or the Nei's minimum distance (Nei, 1972). Table 6 shows the correlations between these measures of genetic distance and allelic richness dissimilarity, which can be summarized as follows: *a*) high correlation between the two genetic distances, *b*) correlation values between distances and dissimilarities close to 0.80 for DRE, and markedly lower for DRR, and *c*) medium correlation between DRR and DRE dissimilarities. These results confirm that genetic diversity and private allelic richness do not exactly address the same type of diversity. Whereas the first one is based on the expected heterozygosity, the second one provides a measure of the singularity of a population in terms of allele numbers.

### Number of alleles as a measure of maximal diversity

From the above, it seems that the formal treatment of the allelic diversity is not clear, and in the case of the partition in components within and between populations it lacks an intuitive interpretation. Furthermore, only private alleles contribute to the allelic diversity of the system, and it would be interesting to have a procedure that takes into account allelic diversity even if they do not have private alleles. Finally, it would also be desirable a procedure implemented in a similar way as the classical partition of heterozygosity in population genetics.

Toro *et al.* (2008) have proposed an alternative based in the idea that the larger is the number of alleles, the larger is the potential diversity of a breed because the maximal diversity occurs when alleles are at equal frequencies. Therefore, a way of highlighting the importance of alleles is to assume that all alleles present in a population have identical frequencies and afterwards calculate the standard partition of expected heterozygosity. The estimate of gene diversity under this situation may take into account of the allelic diversity, by considering the potentiality of each breed according to the number and type of alleles that it carries.

Applying the same calculations as above, the total heterozygosity of the whole set of breeds is  $H_T = 0.831$ , with components within breeds  $H_S = 0.746$  and between breeds  $\bar{D} = 0.085$ , being the amount of differentiation  $F_{ST} = 0.102$ . The proportional contributions of each population to the whole population diversity are shown in Table 7. The wild boar and the Iberian breed contribute almost the same to the within-breed diversity because both populations have almost the same number of alleles. However the wild boar contributes more to the between-breed diversity because it possesses more private alleles. The overall balance indicates that the wild boar contributes more than the Iberian to the total

**Table 7.** Contribution of each pig population to the within-breed ( $H_S$ ), between-breed ( $\bar{D}$ ) and total genetic diversity ( $H_T$ ), assuming identical allelic frequencies. Relative contributions of each breed (in %) between brackets

Population	Within-breed diversity	Between-breed diversity	Total genetic diversity
Wild boar	0.169 (23)	0.015 (18)	0.184 (22)
Duroc	0.165 (22)	0.015 (18)	0.180 (22)
Iberian	0.169 (23)	0.013 (15)	0.182 (22)
Guadyerbas	0.103 (14)	0.026 (31)	0.129 (15)
Torbiscal	0.141 (19)	0.016 (18)	0.156 (19)
	$H_S = 0.746$	$\bar{D} = 0.085$	$H_T = 0.831$

**Table 8.** Loss (–) and gain (+) in percentage (%) of genetic diversity, assuming identical allelic frequencies, when each population is removed from the whole set

Population removed	Within-breed diversity	Between-breed diversity	Total genetic diversity	Rank of genetic diversity
Wild boar	–3.3	+2.0	–2.7	1
Duroc	–2.6	–0.4	–2.4	2
Iberian	–3.3	+8.7	–2.1	3
Guadyerbás	+7.8	–40.0	+2.9	5
Torbiscal	+1.4	–1.7	+1.1	4

diversity. The contribution of the Duroc is lower than the previous ones because although it has more private alleles than Iberian it harbours less number of alleles. Of the last two populations Guadyerbás and Torbiscal, the second is the one that contributes more to the diversity because, although both have just one private allele, it has more number of alleles.

The losses or gains of diversity when one population is removed assuming identical allelic frequencies are also shown in Table 8. Note that, according to this criterion, the most relevant population is again the wild boar. However, the second one is the Duroc breed instead of the Iberian one, reflecting the fact that it contains more private alleles in this particular scenario.

All the previous paragraphs have been mainly focused to the partition of the total diversity of a metapopulation in two components: between and within populations. This partition when applied to genetic diversity or expected heterozygosity can be easily done following the standard population genetics approach. However, the allelic richness as a relevant measure of diversity is more difficult to deal with. First, because it should be standardized to a uniform sample size using either the rarefaction or extrapolation techniques, which respective results may present some discrepancy. Second, in spite of recent theoretical developments (Ollivier and Foulley, 2005) it is far from being obvious how to partition this allelic richness in between and within population components. As our results outline, the rank of the populations according to their contributions to the diversity will be different depending on the criteria utilized. Although in our case this ambiguity has an academic interest given the performed joint analysis of wild and domestic animals, in other settings such as the prioritization of domestic breeds for conservation could be more worrying (Toro *et al.*, 2007).

## Acknowledgements

*Nuestra gratitud a Pepe Malpica, con quien compartimos nuestras primeras ilusiones profesionales. Entre otras muchas cosas por contagiarnos su entusiasmo por la materia experimental, tantas veces transformada en el propio objetivo de nuestro quehacer (JR, LS, MAT)<sup>2</sup>. We thank to the CIA «Dehesón del Encinar» (Oropesa, Toledo), AECERIBER and M. Delibes (Estación Biológica de Doñana, CSIC) for providing biological samples from pigs and wild boars. This research was funded by INIA RZ03-010 grant.*

## References

- ALVES E., FERNÁNDEZ A.I., BARRAGÁN C., OVILO C., RODRÍGUEZ C., SILIÓ L., 2006. Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data. *Span J Agric Res* 4, 37-46.
- BARKER J.S.F., 1999. Conservation of livestock breed diversity. *AGRI* 25, 33-43.
- BRUFORD M.W., 2004. Conservation genetics of UK livestock: from molecules to management. In: *Farm animal genetic resources* (Simm G., Villanueva B., Sinclair K.D., Townsend S., eds). Nottingham Univ Press, Nottingham, UK. pp. 151-169.
- CABALLERO A., TORO M.A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* 3, 289-299.
- EL MOUSADIK A., PETIT R.J., 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol Ecol* 5, 547-555.
- FERNÁNDEZ A., RODRIGÁÑEZ J., TORO M.A., RODRÍGUEZ C., SILIÓ L., 2002. Inbreeding effects on the parameters of the growth function in three strains of Iberian pigs. *J Anim Sci* 80, 2267-2275.

<sup>2</sup> Our gratitude to Pepe Malpica, with whom we shared those fascinating first professional steps. Among many other things for his contagious enthusiasm for the experimental material, so many times becoming the very object of our work (JR, LS, MAT).



- FOULLEY J.L., OLLIVIER L., 2006a. Estimating allelic richness and its diversity. *Liv Sci* 101, 150-158.
- FOULLEY J.L., OLLIVIER L., 2006b. Allele diversity. 8<sup>th</sup> World Congress on Genetics Applied to Livestock Production, August 13-16, Belo Horizonte, Brazil.
- FRANKHAM R., BALLOU J.D., BRISCOE D.A., 2002. Introduction to conservation genetics. Cambridge Univ Press, Cambridge, UK.
- HEDRICK P.W., 2005. A standardized genetic differentiation measure. *Evolution* 59, 1633-1638.
- HILL W.G., RASBASH J., 1986. Models of long term artificial selection in finite populations. *Genet Res* 48, 41-50.
- KALINOWSKI S.T., 2004. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conserv Genet* 5, 539-543.
- LAVAL G., SANCRISTÓBAL 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.
- MACHUGH D.E., LOFTUS R.T., CUNNINGHAM P., BRADLEY D.G., 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim Genet* 29, 333-340.
- NEI M., 1972. Genetic distance between populations. *Amer Nat* 106, 283-292.
- OLLIVIER L., FOULLEY J.L., 2005. Aggregate diversity: new approach combining within- and between-breed genetic diversity. *Liv Prod Sci* 95, 247-254.
- PETIT R.J., EL MOUSADIK A., PONS O., 1998. Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12, 844-855.
- REYNOLDS J., WEIR B.S., COCKERHAM C.C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105, 767-779.
- SIMIANER H., 2005. Using expected allele number as objective function to design between and within breed conservation of farm animal biodiversity. *J Anim Breed Genet* 122, 177-187.
- 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 M.A., RODRIGÁÑEZ J., SILIÓ L., RODRÍGUEZ C., 2000. Genealogical analysis of a closed herd of black hairless Iberian pigs. *Conserv Biol* 14, 1843-1851.
- TORO M.A., FERNÁNDEZ J., CABALLERO A., 2008. Molecular characterization of breeds and its use in conservation. *Livest Prod Sci* (in press).
- WEITZMAN M.L., 1992. On diversity. *Quarterly J Econ* 107, 363-405.
- WRIGHT S., 1969. Evolution and the genetics of populations. The theory of gene frequencies, vol. 2. Univ Chicago Press, Chicago, USA.