

## Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data

E. Alves\*, A. I. Fernández, C. Barragán, C. Ovilo, C. Rodríguez and L. Silió

*Dpt. of Animal Breeding and Genetics. INIA. Ctra. de La Coruña km 7. 28040 Madrid. Spain*

---

### Abstract

The census and structure of Iberian pig breed have experienced important changes along the last decades. Bayesian methods based on multilocus genotypes have been applied for ascertaining the actual breed structure and for identifying genetically distinctive populations. DNA samples from 170 Iberian pigs previously assigned to the strains or varieties Torbiscal, Guadyerbas, Retinto, Entrepelado and Lampiño and 64 Duroc pigs were genotyped for 36 microsatellites. A best partition of only five clusters was estimated in the clustering analysis at group level, when the previous assignation to populations was taken into account. But the individual-based assessment of population structure, ignoring the previous assignation of animals to populations, showed a more complex partition of ten clusters. Results of admixture analyses for partitioning individuals into the inferred clusters showed an important proportion of admixed individuals pre-assigned to the Retinto, Entrepelado and Lampiño varieties. The frequencies of private alleles of the *MC1R* gene also evidenced an important genetic flow between these varieties. The future definition of conservation units in the Iberian breed should consider these results.

**Additional key words:** clustering, *MC1R* gene, mixture and admixture analysis, within-breed variation.

### Resumen

#### Inferencia de la subestructura de la raza porcina Ibérica a partir de genotipos multigénicos de microsatélites

El censo y la estructura de la raza porcina Ibérica han experimentado cambios importantes durante las últimas décadas. Se han utilizado métodos bayesianos basados en genotipos multigénicos para discernir la actual estructura de la raza e identificar en ella poblaciones genéticamente singulares. Con este objetivo, se genotiparon 36 microsatélites en muestras de ADN procedentes de 170 cerdos Ibéricos asignados previamente a las estirpes o variedades Torbiscal, Guadyerbas, Retinto, Entrepelado y Lampiño además de 64 cerdos Duroc. Cuando se tuvo en cuenta esta asignación previa en el análisis, se obtuvo una partición óptima de sólo cinco clases, una de las cuales agrupaba las variedades Retinto y Entrepelado. El análisis individual, ignorando la asignación previa de los animales a las razas o variedades, permitió inferir una partición más compleja de diez clases. Los resultados de los análisis con modelos de mestizaje mostraron una importante proporción de animales mestizos preasignados a las variedades Retinto, Entrepelado y Lampiño. Las frecuencias de los alelos específicos del gen *MC1R* confirmaron el importante flujo genético producido entre estas variedades. Una futura definición de unidades de conservación en la raza Ibérica debería considerar estos resultados.

**Palabras clave adicionales:** análisis de mezcla y mestizaje de poblaciones, clusters, gen *MC1R*, variación intrarracial.

---

### Introduction

Diverse local varieties (Negros Lampiños, Retinto and Rubio or Dourado) were differentiated within the ancient population of the Iberian pig breed showing

important phenotypic and productive differences (Odriozola, 1976; Laguna, 1998; Benito *et al.*, 2000). Some of these varieties were exported to American countries during the colonization, being the direct ancestors of Creole pig breeds (Lemus-Flores *et al.*, 2001) and also contributed in the United States to the origin of the Duroc-Jersey breed (Vaughan, 1950).

The large breed census was drastically reduced since 1960 due to the outbreak of the African swine fever and

---

\* Corresponding author: esalves@inia.es

Received: 16-09-05; Accepted: 27-01-06.

the depreciation of animal fats. Along the last years, the production of pigs of Iberian type has largely increased to satisfy the new demand of top quality meat and dry-cured products, and the population bottleneck has been reversed. However, as a consequence of the past critical period, some ancestral varieties have disappeared and other ones could be endangered or blended.

Phylogenetic techniques based on genetic distances estimated from polymorphic microsatellite markers have been the method of choice to assess the genetic diversity of livestock breeds. This approach rely on the *a priori* definition of populations, and its usefulness will be greatly reduced if these populations do not accurately describe the present-day biological reality (Pearse and Crandall, 2004). Genetically similar groups can be labelled differently due to distinct phenotypes, but conversely, phenotypic similarity may mask underlying genetic variation (Rosenberg *et al.*, 2001). Martínez *et al.* (2000) used this classical approach for analysing the genetic structure of the Iberian breed, and their results mainly supported the division of the breed in the pre-defined varieties, although the traditional classification was not compatible with some singular cases.

Other methods that construct genetic clusters from a set of individual multilocus genotypes have been proposed as a more flexible alternative to those based on genetic distances. Both genetic distances and clustering methods have been used by Fabuel *et al.* (2004) to analyze genetic diversity and conservation priorities in Iberian pigs. The results of this study and another one based on the analysis of mitochondrial DNA sequences (Alves *et al.*, 2003) supported new evidence for the introgression among the traditional Iberian pig varieties.

Clustering methods allow to separate a set of individuals in several populations if their genetic origin is unknown beforehand or to study the correspondence between inferred genetic clusters and known pre-defined population categorizations (Pritchard *et al.*, 2000). Recently, fully Bayesian methods have been proposed for estimating hidden population substructure, which treat both the allele frequencies of the molecular markers and the number of genetically diverged populations as random variables (Corander *et al.*, 2003). These methods allow to cluster data (genetic mixture analysis) either at group level or at individual level, and also to perform admixture analysis, in which the genome of an individual represents a mixture of alleles of different ancestries (Anderson and

Thompson, 2002). The objective of this study was to apply these new statistical tools for detecting whether the actual breed structure preserves the traditional differentiated varieties or consist of mixed or admixed populations, and for defining more accurately genetic units useful to design a rational management of the Iberian breed genetic resources.

## Material and Methods

### Animals

Two out of the five groups of Iberian pigs considered (Guadyervas and Torbiscal) belong to an early conservation programme. Guadyervas is a black hairless strain and Torbiscal is a composite strain obtained from ancient black and red varieties (Rodrigáñez *et al.*, 2000; Toro *et al.*, 2000). The complete genealogy of all the animals of these strains is available since 1945, with 18.9 (Guadyervas) and 21.0 (Torbiscal) generations from the population founders until the animals genotyped here. The remaining pre-defined Iberian pig groups represent the main three extant varieties: black hairless (Negro Lampiño), red (Retinto) pigs and a black hairy variety (Entrepelado), whose piglets show a chestnut colour at birth. Genomic DNA was extracted from blood using standard protocols. Samples were collected from 170 individuals inscribed in the breed herdbook, being their distribution by strains and varieties as follows: 31 Torbiscal, 32 Guadyervas, 50 Retinto (seven breeding nucleus), 30 Lampiño (three breeding nucleus) and 27 Entrepelado (five breeding nucleus). Due to their historical and current relations with the Iberian pigs, a total number of 64 Duroc pigs from seven breeding nucleus was also sampled and analyzed.

### Microsatellites

All the animals were genotyped for 36 microsatellite markers, two on each autosome (Table 1). They were chosen according to a good reproducibility, high polymorphism and absence of null alleles. In seven out of the 18 chromosomes, both microsatellites map on the same chromosome arm with an average genetic distance of 34.5 cM, although the distances were only 13.4 and 14.2 cM for the chromosomes 1 and 18, respectively.

**Table 1.** Number of alleles ( $N_A$ ) and observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities of microsatellite markers in Iberian and Duroc breeds. For each marker the chromosome (Chr) localization is indicated

Chr	Marker	Iberian			Duroc			Chr	Marker	Iberian			Duroc		
		$N_A$	$H_o$	$H_e$	$N_A$	$H_o$	$H_e$			$N_A$	$H_o$	$H_e$	$N_A$	$H_o$	$H_e$
1q	S0113	5	0.671	0.728	3	0.375	0.403	1q	S0155	5	0.447	0.483	4	0.672	0.593
2p	SW240	8	0.594	0.724	7	0.797	0.774	2q	SW395	6	0.565	0.713	3	0.266	0.300
3p	SW72	6	0.512	0.687	6	0.656	0.708	3q	S0002	8	0.606	0.677	5	0.672	0.752
4p	S0301	5	0.365	0.524	5	0.641	0.730	4q	S0097	12	0.718	0.826	7	0.531	0.576
5p	SW413	4	0.576	0.720	5	0.484	0.765	5q	S0005	14	0.647	0.783	15	0.625	0.803
6p	SW1057	8	0.653	0.722	5	0.453	0.723	6q	SW2419	8	0.706	0.758	5	0.609	0.674
7p	S0025	8	0.594	0.672	5	0.425	0.608	7q	SW632	10	0.541	0.678	7	0.469	0.520
8q	S0178	9	0.635	0.758	7	0.656	0.712	8q	S0225	7	0.571	0.661	4	0.594	0.721
9p	SW911	7	0.547	0.699	6	0.656	0.678	9q	SW1349	5	0.482	0.553	5	0.171	0.626
10p	S0038	8	0.465	0.765	6	0.484	0.447	10q	S0070	11	0.712	0.819	10	0.609	0.625
11p	S0385	5	0.529	0.665	6	0.734	0.770	11q	SW703	5	0.418	0.566	4	0.453	0.579
12p	SW874	7	0.582	0.774	9	0.578	0.783	12q	S0106	13	0.694	0.816	6	0.500	0.578
13p	S0219	4	0.476	0.650	4	0.359	0.472	13q	S0291	7	0.523	0.719	6	0.719	0.751
14q	SW210	5	0.535	0.676	4	0.453	0.555	14q	SW857	5	0.624	0.711	5	0.453	0.668
15q	SW1111	6	0.653	0.785	6	0.651	0.656	15q	SW936	9	0.723	0.810	5	0.641	0.723
16q	S0026	7	0.341	0.460	5	0.562	0.667	16q	S0061	5	0.659	0.657	5	0.672	0.713
17q	SW1920	8	0.612	0.759	7	0.453	0.760	17q	SW24	6	0.582	0.692	7	0.719	0.763
18q	S0120	7	0.582	0.723	4	0.406	0.544	18q	SW787	6	0.612	0.680	7	0.797	0.812
									Average	7.2	0.576	0.697	5.8	0.556	0.653

Amplified microsatellite markers were analyzed with Genescan software on capillary electrophoresis equipment with fluorescent detection (ABI PRISM 3100 genetic analyzer). To increase the accuracy of allele size determination, four control animals were genotyped in all the runs. The genetic variability within each one of the sampled populations was measured by the number of alleles ( $N_A$ ) and the expected heterozygosity ( $H_e$ ) for each genotyped microsatellite.

### MC1R genotyping

The *MC1R* intragenic haplotypes (Chr. 6p) present in the analyzed Iberian and Duroc populations were determined by the procedures described by Fernández *et al.* (2004) to obtain additional genetic information.

### Clustering analysis

Mixture analysis of microsatellite data were performed according to Corander *et al.* (2003) to

provide posterior distributions of partitions  $S = (s_1, \dots, s_k)$  among the  $N_P$  sampling units into  $k$  non-empty classes (clusters), which have non-identical allele frequency parameters over the  $N_L$  genotyped loci. Independently for each particular partition  $S$ , the joint distribution of the data and parameters is proportional to the Multinomial-Dirichlet expression

$$\prod_{i=1}^k \prod_{j=1}^{N_L} \prod_{l=1}^{N_{A(j)}} p_{ijl}^{n_{ijl} + \alpha_j}$$

where  $p_{ijl}$  is the unknown allele frequency,  $n_{ijl}$  is the observed number of copies of allele  $l$  at locus  $j$  among sampling units into the cluster  $s_i$  and  $\alpha_j$  is the Dirichlet prior hyperparameter, chosen as  $\alpha_j = 1 / N_{A(j)}$ , being  $N_{A(j)}$  the number of alleles observed at locus  $j$ . Both Hardy-Weinberg and linkage equilibrium within each class  $s_i$  were assumed in the above. The prior distribution for the partition parameters is chosen to be uniform in the space of all the possible partitions that are considered *a priori* equally likely. For small values of  $N_P$ , it is possible to use complete enumeration to obtain exactly the posterior distribution of the

parameters  $S$  and  $k$  over all the possible partitions. When  $N_P > 10$ , the number of possible partitions is too large for exhaustive enumeration, and values from the posterior distribution may be obtained using the Metropolis-Hastings algorithm. Additional details of the method can be found in Corander *et al.* (2003, 2004), and the calculations were performed using the BAPS2.0 software. In the mixture analysis at group level, the previous assignation of the pigs to six groups was used to inform the analysis and then  $N_P = 6$  is the maximum number of possible clusters. These groups are the sampling units to be clustered to study the correspondence between inferred genetic clusters and the pre-defined categorization of breeds or varieties. After inferring the structure, the estimates of the coefficient of genetic differentiation  $F_{ST}$  (Nei, 1977) and the pairwise genetic distances  $D_m$  (Nei, 1972) and  $D_R$  (Reynolds *et al.*, 1983) were calculated in a Bayesian model averaged sense, since the distance measures between genetic clusters are obtained by averaging over the posterior distribution of partition parameters ( $s_1, \dots, s_k$ ). A second mixture analysis was carried out at individual level ( $N_P = 234$  individuals) to identify the optimal allocation of individuals to genetically divergent clusters, without any pre-defined categorization.

The admixture clustering analysis provides an additional parameter  $q_i^{(m)}$ , the proportion of the genome of the individual  $m$  proceeding from the cluster  $s_i$ , for partitioning individuals into clusters based on multilocus genotypes. Two admixture analyses were performed, based on previous inferences about clusters obtained by mixture clustering analysis either at group level and at individual level. Inferences about admixture were obtained using the software BAPS version 3.1, that calculates for each individual the posterior mode of  $q_i^{(m)}$  conditional to the structure parameters  $S$  and  $k$ , and the posterior probability ratio for the model with admixture and the model where the individual is forced to have pure ancestry. Both models are considered *a priori* equally likely and the posterior probability ratio coincides with the Bayes Factor (BF). As Kass and Raftery suggest, it was considered as no evidence against the no admixture hypothesis when the BF was lower than 3.2, substantial evidence when the BF was within 3.2 and 10, and strong evidence when the BF was greater than 10 (Kass and Raftery, 1995).

## Results

Table 1 shows the number of detected alleles for each genotyped microsatellite arranged by autosome. This number varies from 4 to 14 (average 7.2 alleles) summed across the Iberian breed as a whole, and from 3 to 15 in the Duroc breed (average 5.8 alleles). In the Iberian breed, the average number of alleles per population ranged from 2.08 (microsatellite *S0219*) to 7.8 (microsatellite *S0005*). Table 1 also gives, for each locus and breed, the observed heterozygosity by direct count and the expected heterozygosity under Hardy-Weinberg equilibrium. The observed values were generally lower than the expected values, indicating heterogeneity between populations within each breed.

### Clustering analysis at group level

The results of group level mixture analysis showed just one partition of five clusters with differing population frequencies. Four of them corresponded to the Duroc breed ( $s_1$ ), the Guadyerbas ( $s_3$ ) and Torbiscal ( $s_5$ ) strains and the Lampiño variety ( $s_4$ ). The last cluster ( $s_2$ ) combines the Entrepelado and Retinto varieties of Iberian pigs. The posterior mean value of the coefficient of genetic differentiation among these five clusters  $F_{ST}$  was 0.170 and the posterior standard deviation PSD = 0.003. The means of the posterior distributions of  $D_m$  and  $D_R$  distances indicate a maximum distance between Duroc and Guadyerbas, and a minimum distance between the Lampiño variety and the mixed cluster  $s_2$  grouping the Retinto and Entrepelado varieties. The values of both genetic distances are dependent on the number of generations since divergence and the effective sizes of populations. But the  $D_m$  distance also depends on the founder frequencies, and it could explain its higher values between the Duroc and Iberian clusters (Table 2).

The admixture analysis based on this clustering revealed no evidence of admixture from different clusters found for pigs pre-assigned to the Duroc breed and the Guadyerbas and Torbiscal strains. The BF values for the comparison of models with admixture and with pure ancestry were lower than 3.2 for all the individuals grouped into the clusters  $s_1$ ,  $s_3$  and  $s_5$ . However, the admixture analysis allowed to identify 38

**Table 2.** Results of mixture analysis at group level: posterior means and standard deviations of genetic distances among the five inferred clusters<sup>1, 2</sup>. Nei's distance  $D_m$ , above the diagonal and Reynolds's distance  $D_R$ , below the diagonal

Inferred cluster	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
$s_1$		0.561 (0.018)	0.905 (0.026)	0.524 (0.020)	0.769 (0.025)
$s_2$	0.183 (0.005)		0.482 (0.019)	0.194 (0.013)	0.301 (0.015)
$s_3$	0.337 (0.006)	0.243 (0.007)		0.412 (0.022)	0.416 (0.022)
$s_4$	0.175 (0.005)	0.080 (0.005)	0.222 (0.008)		0.324 (0.020)
$s_5$	0.253 (0.006)	0.137 (0.006)	0.248 (0.009)	0.144 (0.007)	

<sup>1</sup> Partition of pig populations  $S$  ( $s_1, s_2, s_3, s_4, s_5$ ) with maximum posterior probability [ $p(S|\text{data}) = 1.00$ ].

<sup>2</sup> Populations grouped into each cluster:  $s_1$ =Duroc,  $s_2$ =Entrepelado and Retinto,  $s_3$ =Guadyervas,  $s_4$ =Lampião and  $s_5$ =Torbiscal.

individuals grouped into the remaining clusters ( $s_2$  and  $s_4$ ) with at least substantial evidence against the no admixture hypothesis ( $\text{BF} > 3.2$ ). The respective proportions of these admixed pigs were 12/27, 13/50 and 13/30, for each one of the pre-defined varieties Entrepelado, Retinto and Lampião. Several details of these results should be outlined: *i*) some Entrepelado pigs showed a remarkable proportion of alleles from the clusters  $s_3$  and  $s_4$ , corresponding to the black hairless populations Guadyervas and Lampião, and other Entrepelado individuals showed admixture with the clusters  $s_1$  (Duroc) and  $s_4$  (Lampião); *ii*) the cluster  $s_5$  (Torbiscal) represented one important proportion of the genome of six Retinto pigs; *iii*) one important proportion of the genome of six pigs pre-assigned to the Lampião variety and grouped into the cluster  $s_4$  is represented by the cluster  $s_3$  (Guadyervas); *iv*) one important proportion of the genome of seven Lampião individuals is represented by the cluster  $s_2$  (Entrepelado and Retinto), and one additional proportion of the genome of two of them is represented by the cluster  $s_5$  (Torbiscal), and finally *v*) seven out of the 38 admixed pigs showed a non trivial proportion ( $q_1 \geq 0.10$ ) of Duroc alleles. Strong evidence against the no admixture hypothesis ( $\text{BF} > 10$ ) was found for 20 out of the 38 admixed pigs, which are presented in Table 3.

### Clustering analysis at individual level

The results of the individual-based assessment of population structure showed a best partition of ten

genetically divergent clusters. The composition of these clusters is described in the Table 4. Four results can be outlined: *i*) two of the inferred clusters ( $s_E$  and  $s_F$ ) included separately animals from the closed strains Guadyervas and Torbiscal, corresponding to clusters  $s_3$  and  $s_5$  obtained in the analysis at group level; *ii*) the cluster of greatest size ( $s_D$ ) was a pool of Iberian genotypes merging all the animals pre-assigned to the Entrepelado variety (27 pigs), 33 out of the 50 Retinto pigs, and 9 out of the 30 Lampião pigs; *iii*) most of the Duroc pigs (47 animals) were grouped into the same cluster  $s_A$ ; *iv*) the remaining six small clusters corresponded to individuals sampled from isolated breeding nuclei of Duroc and Iberian pigs.

The results of the admixture analysis based on clustering at individual level should be cautiously examined because a minimum cluster size (30 individuals) was imposed. As a consequence the possible genetic origins of individuals were reduced to four clusters, and 17 Duroc and 38 Iberian pigs, grouped into the clusters of lower size ( $s_B, s_C, s_G, s_H, s_I$  and  $s_J$ ), were removed from the analysis. As in the previous analysis at group level, admixed individuals were not identified into the clusters  $s_A$  (Duroc),  $s_E$  (Guadyervas) and  $s_F$  (Torbiscal). According to the correspondent BF values, 17 out of the 69 pigs grouped into the cluster  $s_D$  showed substantial evidence against the hypothesis of pure ancestry ( $\text{BF} > 3.2$ ), being 3/27, 9/33 and 5/9 the respective proportions of admixed animals for pigs pre-assigned to the varieties Entrepelado, Retinto and Lampião. Strong evidence against the no admixture hypothesis ( $\text{BF} > 10$ ) was

**Table 3.** Results of admixture analysis conditional on clusters inferred by mixture clustering analysis at group level<sup>1</sup>: Bayesian posterior mode estimates of the proportion of the genome [ $q_i^{(m)}$ ] that belongs to the cluster  $s_i$  in pigs with strong evidence against the hypothesis of no admixture (Bayes Factor<sup>2</sup> > 10)

Animal code ( $m$ )	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	Bayes Factor
Entrepelado						
70	0.00	0.78	0.11	0.11	0.00	15.63
72	0.13	0.81	0.00	0.06	0.00	14.76
73	0.12	0.69	0.00	0.19	0.00	15.75
91	0.00	0.78	0.22	0.00	0.00	40.48
Retinto						
166	0.00	0.65	0.00	0.35	0.00	46.23
184	0.13	0.80	0.07	0.00	0.00	20.76
190	0.00	0.52	0.00	0.00	0.48	2463.90
191	0.00	0.62	0.00	0.13	0.25	80.54
192	0.00	0.67	0.17	0.00	0.16	23.54
193	0.00	0.56	0.10	0.02	0.32	114.60
196	0.00	0.55	0.01	0.12	0.32	83.18
199	0.00	0.66	0.00	0.00	0.34	102.30
Lampião						
129	0.00	0.00	0.27	0.72	0.01	32.31
132	0.00	0.00	0.26	0.74	0.00	15.23
135	0.00	0.00	0.24	0.76	0.00	18.96
136	0.00	0.58	0.00	0.42	0.00	283.20
144	0.00	0.17	0.00	0.59	0.24	26.12
146	0.00	0.20	0.23	0.44	0.13	109.80
147	0.00	0.35	0.04	0.52	0.09	26.57
149	0.00	0.41	0.01	0.58	0.00	10.06

<sup>1</sup> Populations grouped into each cluster:  $s_1$ =Duroc,  $s_2$ =Entrepelado and Retinto,  $s_3$ =Guadyerbás,  $s_4$ =Lampião and  $s_5$ =Torbiscal. <sup>2</sup> Bayes Factor, posterior probability ratio for the models with admixture and forced to have pure ancestry.

**Table 4.** Results of mixture analysis at individual level: optimal partition of individual pigs  $S$  ( $s_A, \dots, s_J$ ) and individuals grouped into each cluster

Inferred cluster	Number and pre-defined population of pigs clustered
$s_A$	47 Duroc
$s_B$	9 Duroc
$s_C$	8 Duroc
$s_D$	27 Entrepelado; 33 Retinto; 9 Lampião
$s_E$	32 Guadyerbás
$s_F$	31 Torbiscal
$s_G$	12 Lampião
$s_H$	9 Lampião
$s_I$	7 Retinto
$s_J$	10 Retinto

found for eight pigs particularly atypical of the Retinto and Lampião varieties (Table 5).

### MC1R genotypes

The complementary analysis, based on the private alleles of the coat color *MC1R* gene, revealed some results unexpected under the hypothesis of strict isolation between the traditional varieties of Iberian pigs (Table 6): *i*) the presence in some Lampião pigs of *MC1R\*6* allele, jointly with the *MC1R\*3* allele, characteristic of black populations (Kijas *et al.*, 1998) and *ii*) the presence of *MC1R\*3* alleles in some Retinto pigs, jointly with the *MC1R\*6* or *MC1R\*7* alleles,

**Table 5.** Results of admixture analysis conditional on clusters inferred by mixture clustering analysis at individual level<sup>1</sup>: Bayesian posterior mode estimates of the proportion of the genome [ $q_i^{(m)}$ ] that belongs to the cluster  $s_i$  in pigs with strong evidence against the hypothesis of no admixture (Bayes Factor<sup>2</sup> > 10)

Animal code ( $m$ )	$s_A$	$s_D$	$s_E$	$s_F$	Bayes Factor
Retinto					
190	0.00	0.51	0.00	0.49	1203.60
191	0.00	0.73	0.00	0.27	11.28
192	0.00	0.68	0.15	0.17	10.50
198	0.00	0.70	0.00	0.30	12.67
199	0.00	0.70	0.00	0.30	19.57
Lampião					
146	0.00	0.67	0.27	0.06	169.60
151	0.03	0.54	0.25	0.18	577.20
152	0.09	0.72	0.14	0.05	10.59

<sup>1</sup> Based on inferences about clusters grouping at least 30 pigs:  $s_A$ ,  $s_D$ ,  $s_E$  and  $s_F$  (see Table 4). <sup>2</sup> Bayes Factor, posterior probability ratio for the models with admixture and forced to have pure ancestry.

**Table 6.** Frequencies of MC1R alleles in the six Iberian and Duroc pig populations studied. N = number of individuals

	Guadyerbas (N = 32)	Torbiscal (N = 31)	Lampião (N = 30)	Retinto (N = 50)	Entrepelado (N = 27)	Duroc (N = 64)
Black allele						
<i>MC1R*3</i>	1.00	0.10	0.92	0.04	0.07	0.00
Red alleles						
<i>MC1R*4</i>	0.00	0.00	0.00	0.00	0.00	1.00
<i>MC1R*6</i>	0.00	0.73	0.08	0.65	0.54	0.00
<i>MC1R*7</i>	0.00	0.17	0.00	0.31	0.39	0.00

characteristic of red populations (Fernández *et al.*, 2004). However, the joint segregation of *MC1R\*3*, *MC1R\*6* or *MC1R\*7* alleles in Torbiscal pigs could be expected according to their genetic origin. The presence of the black *MC1R\*3* allele in the red Retinto, Torbiscal and Entrepelado populations refutes the assumed dominance of this allele.

## Discussion

The Iberian breed had its origin long before the period of development of European breeds from the end of the 1700s to the beginning of the 1900s, mainly based on racial standards and herdbooks controlled by breed societies. For centuries, this pig population was

extensively farmed in the sparse woodlands of the Southwest of Iberian peninsula to satisfy the high demand for animal fats. Without selective preponderance of any group of breeders and scarce genetic flow between herds, its genetic singularity was developed through a process of adaptation to hard environmental conditions derived from seasonal availability of feeding resources and semiarid continental climate. Besides of the empirical selective breeding, demographic fluctuations and population isolation have been other moulding influences on the within breed differentiation of locally diffused varieties, with rare genetic flow between herds. This heterogeneity was acknowledged in the breed standard type, which was lately proposed during the past century. As a consequence of the census reduction, the old reticular

structure was replaced by a pyramidal breed structure. In this new structure, a small number of breeding nuclei supplies purebred Iberian reproducers to the production tiers, based on unrestricted crossbreeding of Iberian with Duroc and other colour-coated breeds (Silió, 2000). According to mtDNA studies at least the maternal contribution of Asian pigs to the Iberian breed seems unlikely, although East Asian pigs have contributed to the development of most of the European breeds (Alves *et al.*, 2003). The first goals of this paper were to analyse the actual genetic structure of the Iberian breed and the prevalence of the old varieties potentially submitted to the damaging effects of the genetic erosion. In this sense, the clustering analysis at group level revealed a partition of only four clusters for the Iberian breed, with the extant Entrepelado and Retinto varieties grouped in the same genetic cluster. But the range of plausible values for the number of clusters ( $k$ ) has the number of sampled populations ( $N_p$ ) as upper boundary, and therefore this analysis cannot detect underlying substructure if it occurs within each pre-defined population (Manel *et al.*, 2005). The clustering analysis at individual level overcome this restriction, and its results concerning the Iberian pigs delineated a more detailed partition of seven genetic clusters. The cluster of greater size grouped 69 pigs pre-assigned to the Entrepelado, Retinto and Lampiño varieties, suggesting shared ancestral origins for these pigs, regardless of their phenotypical dissimilarities. Finally, the results of admixture analysis showed a noticeable number of pigs pre-assigned to these three varieties with important proportions of their genome proceeding from recent admixture events. The most clear result of the analyses was the large blending detected among the ancient varieties.

The Entrepelado variety is an emergent type of Iberian pigs, which origin is an intriguing topic. Martínez *et al.* (2000) reported dendrograms based on genetic distances between Iberian pig varieties, in which Retinto and Entrepelado samples appear mixed in the same cluster and the Lampiño samples were clustered into a different group. However, Diéguez (2001) hypothesized that the Entrepelado variety should proceed from the intercross among Iberian Retinto and Lampiño pigs. This genetic origin could explain the joint segregation of *MC1R*\*3, *MC1R*\*6 or *MC1R*\*7 alleles in Entrepelado pigs and the present clustering results. Moreover, Alves *et al.* (2003) found two mtDNA haplotypes, based on *Cyt B* and *D-loop* sequences, with simultaneous presence in Entrepelado

and either Lampiño or Retinto pigs suggesting maternal ancestries from both traditional varieties.

All methods based on cluster analysis involve some uncertainty unless the true populations are strongly divergent, as the Duroc pigs grouped in the clusters  $s_1$  or  $s_4$ , and the preserved Guadyerbas and Torbiscal strains grouped into unique clusters. Introgression of Duroc genes is the most important risk of genetic pollution for the Iberian breed, and breed specific markers based on polymorphisms found in two coat color genes (*MC1R* and *OCA2*) have been proposed to detect Duroc crossbred individuals (Fernández *et al.*, 2004). Although none Duroc *MC1R*\*4 allele was detected in the analyzed Iberian samples, the results of admixture analysis showed substantial or strong evidence of Duroc alleles in pigs from two breeding nuclei of the Entrepelado and Retinto varieties. The magnitude of this proportion (close to 0.125) could indicate the presence of one Duroc pig among the eight great-grandfathers of each one of these admixed animals. A more detailed investigation of possible recent crossbreeding in these two breeding nuclei may be advisable.

In the admixture analysis, the admixture proportions in the parental and hybrid populations was assumed to be measurable, but subsequent selective breeding and genetic drift from the admixture event until the sampled individuals may change the allele frequencies difficulting to get a precise inference of admixture proportions (Bruford, 2004). The results concerning the related strains Guadyerbas and Torbiscal illustrate this topic. According to the pedigree analysis, one third of the genome of the actual Torbiscal pigs proceed from the contribution of Guadyerbas ancestors to the foundation of this composite strain, kept isolated since 1963 (Rodrigáñez *et al.*, 2000). But 14 generations later, the admixture event become obscured by genetic drift and no evidence of admixture was observed in the present analyses of Torbiscal pigs.

The six small clusters grouping Duroc, Lampiño or Retinto pigs from isolated breeding nuclei require a more detailed description. The clusters  $s_B$  and  $s_C$  group descendants from ancient Duroc-Jersey pigs imported from United States forty years ago, and are respectively maintained by the *Centro Regional de Selección y Reproducción Animal* (CERSYRA) of Badajoz and one private breeder. More interesting are the clusters  $s_G$  and  $s_H$ , that join separately animals of two breeding nuclei of the Lampiño variety, one of them corresponding to

the ancient variety Lampiño de la Serena and the other one from Portuguese provenance. Finally the clusters  $s_I$  and  $s_J$  group separately pigs of two breeding nuclei of the Retinto variety, both showing singular morphological traits. One of them is characterized by pigs with narrow legs and steep angle pastern, and the other one presents a high frequency of animals showing wattles in their neck («mamellados»).

Diverse prioritisation approaches for livestock breed conservation are today largely debated by the specialists, and their application to the Iberian breed have been discussed by Fabuel *et al.* (2004). A rational management of the Iberian breed genetic resources should take account of these singular breeding nuclei with significant allele frequency differences at both nuclear and mitochondrial DNA, detected by these analyses and the previous one of Alves *et al.* (2003). Besides of their genetic uniqueness, they are vestiges of the old Iberian varieties. Both characteristics may justify their consideration as candidates to be preserved in well designed conservation programmes.

## Acknowledgements

The authors are grateful to Nines López for technical assistance, Jukka Corander for BAPS software and Miguel Toro for useful comments. Pig blood samples were kindly supplied by CIA «Dehesón del Encinar» and AECERIBER. This research was funded by INIA grant RZ03-010.

## References

- ALVES E., OVILO C., RODRÍGUEZ M.C., SILIÓ L., 2003. Mitochondrial DNA sequence variation and phylogenetic relationship among Iberian pigs and other domestic and wild pig populations. *Anim Genet* 34, 319-324.
- ANDERSON E.C., THOMPSON E.A., 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160, 1217-1229.
- BENITO J., VÁZQUEZ C., MENAYA C., FERRERA J.L., SILIÓ L., 2000. Evaluation of productive parameters in different strains of Iberian pig. In: Tradition and innovation in Mediterranean pig production (J.A. Afonso & J.L. Tirapicos, eds). CIHEAM/ICAM-UE, Zaragoza. pp. 113-121.
- BRUFORD M.W., 2004. Conservation genetics of UK livestock: from molecules to management. In: Farm animal genetic resources (G. Simms, B. Villanueva, K. D. Sinclair & S. Towsend, eds). Nottingham University Press. pp. 151-169.
- CORANDER J., WALDMANN P., SILLANPÄÄ M.J., 2003. Bayesian analysis of genetic differentiation between populations. *Genetics* 163, 367-374.
- CORANDER J., WALDMANN P., MAITTINEN P., SILLANPÄÄ M.J., 2004. BAPS2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 20(15), 2363-2369.
- DIEGUEZ E., 2001. Base animal: pasado, presente y futuro. In: Porcino Ibérico: aspectos claves (C. Buxadé & A. Daza, eds). Ediciones Mundi-Prensa, Madrid. pp. 49-82.
- FABUEL E., BARRAGÁN C., SILIÓ L., RODRÍGUEZ M.C., TORO M., 2004. Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity* 93, 104-113.
- FERNÁNDEZ A., FABUEL E., ALVES E., RODRÍGUEZ M.C., SILIÓ L., OVILO, C., 2004. DNA tests based on coat colour genes for authentication of the raw material of meat products from Iberian pigs. *J Sci Food Agric* 84, 1855-1860.
- KASS R., RAFTERY A.E., 1995. Bayes factors. *J Amer Stat Assoc* 90, 773-795.
- KIJAS J.H.M., WALES R., TÖRNSTEN A., CHARDON P., MOLLER M., ANDERSSON L., 1998. Melanocortin receptor 1 (MC1R) mutations and coat colour in pigs. *Genetics* 150, 1177-1185.
- LAGUNA E., 1998. El cerdo ibérico. Ediciones Mundi-Prensa, Madrid.
- LEMUS-FLORES C., ULLOA-ARVIZU R., RAMOS-KURIM M., ESTRADA F.J., ALONSO R.A., 2001. Genetic analysis of Mexican hairless pig populations. *J Anim Sci* 79, 3021-3026.
- MANEL S., GAGGIOTTI O.E., WAPLES R.S., 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol Evol* 20, 136-142.
- MARTÍNEZ A.M., DELGADO J.V., RODERO A., VEGA-PLA J.L., 2000. Genetic structure of the Iberian pig breed using microsatellites. *Anim Genet* 31, 295-301.
- NEI M., 1972. Genetic distance between populations. *Am Naturalist* 106, 283-292.
- NEI M., 1977. F-statistics and analysis of gene diversity in sub-divided populations. *Ann Hum Genet* 41, 225-233.
- ODRIOZOLA M., 1976. Investigaciones sobre datos acumulados en dos pjaras experimentales. IRYDA, Ministerio de Agricultura, Madrid. pp. 146.
- PEARSE D.E., CRANDALL K.E., 2004. Beyond  $F_{ST}$ : analysis of population genetic data for conservation. *Conserv Genet* 5, 585-602.
- PRITCHARD J.K., STEPHENS M., DONNELLY T., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- 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.
- RODRIGÁNEZ J., TORO M.A., RODRIGUEZ M.C., SILIÓ L., 2000. Alleles survival from Portuguese and Spanish strains in a population of Iberian pig. In: Tradition and

- innovation in Mediterranean pig production (J.A. Afonso & J.L. Tirapicos, eds). CIHEAM/ICAM-UE, Zaragoza. pp. 57-61.
- ROSENBERG N.A., BURKE T., ELO K., FELDMAN M.W., FREIDLIN P.J., GROENEN M.A.M., HILLEL J., MAKI-TANILA A., TIXIER-BOICHARD M., VIGNAL A., WIMMERS K., WEIGEND S., 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* 159, 699-713.
- SILIÓ L., 2000. Iberian pig breeding programme. In: *Developing breeding strategies for lower input animal production environments* (S. Galal, J. Boyazoglu & K. Hammond, eds). Rome: ICAR. pp. 511-519.
- TORO M.A., RODRIGÁÑEZ J., SILIÓ L., RODRÍGUEZ M.C., 2000. Genealogical analysis of a closed herd of black hairless Iberian pigs. *Conserv Biol* 14, 1843-1851.
- VAUGHAN H.W., 1950. *Breeds of live stock in America*. Long's College Book Co.: Columbus, Ohio.