View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repositório Institucional da Universidade de Aveiro

Mammalian Biology
Zeitschrift für Säugetierkunde

www.elsevier.de/mambio

ORIGINAL INVESTIGATION

Genetic structure of the wild boar population in Portugal: Evidence of a recent bottleneck

Eduardo Ferreira*, Luís Souto, Amadeu M.V.M. Soares, Carlos Fonseca

CESAM & Departamento de Biologia, Universidade de Aveiro, Aveiro, Portugal

Received 9 February 2008; accepted 29 May 2008

Abstract

The present study assesses the degree of genetic structure and the presence of recent genetic bottlenecks in the wild boar population in Portugal. One hundred and ten individuals were sampled after capture during organised legal drive hunts, conducted in 58 municipalities across the continental territory, during the game seasons of 2002/2003 and 2003/2004. Individuals were genetically typed at six microsatellite loci using multiplex PCR amplification. Significant deviations from Hardy–Weinberg equilibrium were found for the total population of wild boar in Portugal. Wild boar population genetic structure was assessed using Bayesian methods, suggesting the existence of three subpopulations (**North**, **Centre** and **South**). Tests were conducted to detect the presence of potential migrants and hybrids between subpopulations. After exclusion of these individuals, three sets of wild boars representative of respective subpopulations were distinguished and tested for the effects of recent bottlenecks. Genetic distances between pairs of subpopulations were quantified using F_{ST} and R_{ST} estimators, revealing a variation of 0.138–0.178 and 0.107–0.198, respectively. On the basis of genetic and distribution data for Portuguese wild boar from the beginning of the 20th century, a model of strong demographic decline and contraction to isolated refuge areas at the national level, followed by a recovery and expansion towards former distribution limits is suggested. Some evidence points to present admixture among subpopulations in contact areas.
© 2008 Deutsche Gesellschaft für Säugetierkunde. Published by Elsevier GmbH. All rights reserved.

Keywords: *Sus scrofa* L.; Genetic structure; Bottleneck; Microsatellites; Bayesian statistics

Introduction

Wild boar (*Sus scrofa*, Linnaeus, 1758) is a widely distributed ungulate whose success can be attributed to a variety of ecological features such as opportunistic omnivorous behaviour (Massei et al. 1996; Fonseca 2004), high prolificness and adaptive plasticity (Fonseca et al. 2004). It is described as a philopatric species depending on age class (Spitz and Janeau 1990; Lemel et al. 2003), and presenting differences on annual and seasonal space use (Keuling et al. 2007). However, evidence indicates low genetic flow among local populations (Okumura et al. 1996).

According to Fonseca (2004), wild boars were once very abundant in Portugal. However, at the beginning of the 20th century, the species was confined to mountain areas near the national border with Spain and to some royal hunting areas, due to strong hunting pressure (Fonseca 2004). By the 1960s, wild boar populations were extremely reduced due to both continued over-hunting and an outbreak of classic swine fever. In 1967, wild boar hunting was banned in Portugal (Serôdio,

*Corresponding author. Tel.: +351 234 370350; fax: +351 234 372587.
E-mail address: elferreira@ua.pt (E. Ferreira).

1616-5047/$ - see front matter © 2008 Deutsche Gesellschaft für Säugetierkunde. Published by Elsevier GmbH. All rights reserved.
doi:10.1016/j.mambio.2008.05.009

1985 **in** Fonseca 2004), and by the end of the 1970s the species began to recover. However, at that time the distribution of wild boar remained fragmented. Nowadays, wild boars are found throughout the country, with the exception of major urban areas and some parts of the coastline (Fonseca 2004), and they are of "low concern" in terms of their national conservation status (Cabral et al. 2005). Similarly in Spain, the wild boar is distributed throughout most of the continental territory (Rosell and Herrero 2002).

Wild boar populations are widely exploited and managed throughout Europe. However, knowledge of the genetic, demographic and ecological impacts of these management practices remains incomplete (Randi 1995). The genetic impact of wild boar relocations was considered by Vernesi et al. (2003). Animal health can be impacted by relocating wild boars (Fernandez-de- Mera et al. 2003), particularly when these animals come in contact with farmed domestic pig (Vidal et al. 2006; Melzer et al. 2006).

Wild boar and domestic pig belong to the same species, *S. scrofa*, Linnaeus, 1758. There is some evidence that pig domestication occurred independently in Europe and Asia (Giuffra et al. 2000; Larson et al. 2005). The divergence between ancestral forms of wild boar has been estimated to have occurred much earlier than that estimated for its domestication (Giuffra et al. 2000).

The species genome has been widely deciphered. Large contributions were made by the "PigMap" (Archibald et al. 1995) and "USDA Pig Genome" (Rothschild 2003) programs in Europe and the USA, respectively. Full sequencing of *S. scrofa* mitochondrial DNA was achieved by Ursing and Arnason (1998). Microsatellite mutation rates were estimated for this species by Yue et al. (2002) with 7.52×10^{-5} per **locus** and generation. Lowden et al. (2002) tested several microsatellite markers developed for domestic pig in wild suiforms, finding a high level of conservation in the studied loci. Several studies on domestic pig breeds' biodiversity have been performed using these kinds of markers (e.g. Harcet et al. 2006; San Cristobal et al. 2006). Microsatellites were also applied in studies of feral pig populations in Australia (Hampton et al. 2004; Spencer et al. 2006). However, only a few studies have applied pig microsatellites to wild boar populations in Eurasia (Vernesi et al. 2003; Lorenzini 2005; Fickel and Hohmann 2006).

The main goal of this study was to assess the genetic variability and the degree of genetic structure of the wild boar (*S. scrofa*) population in Portugal, using a set of microsatellite markers. We also aimed to identify the main factors determining the degree of genetic structure and variability. Finally, we compared the results of this study with the available documented data on the demographic history of wild boars in Portugal.

Material and methods

Sampling

Wild boar samples were collected during the hunting seasons in 2002/03 and 2003/04, in 58 municipalities across continental Portugal (Fig. 1). Blood was collected in K_3EDTA tubes, from 110 wild boars (shot during drive hunts) and stored at 4 °C. Each animal was described according to weight, sex and estimated age. In the lab, a portion of each sample was transferred to FTA® individual cards (Whatman) and stored at room temperature. The remaining blood sample was stored at −20 °C for subsequent use.

DNA extraction

DNA was extracted from blood dried on FTA® cards, using the standard *Chelex*® procedure described by Walsh et al. (1991). Extractions were performed in a total volume of 200 µl, using small areas of the blood card (1–2 mm^2). Samples were used immediately for amplification or stored at −20 °C for later use. In the latter case, samples were stirred and centrifuged prior to amplification.

Amplification and genotyping

Six markers were chosen (from an original set of 91 pairs of primers) based on their known polymorphism, chromosome location, annealing temperature, size range, fluorescence dye applicability and performance under constant amplification conditions. Marker selection was performed with the aim of multiplex amplification of several markers, and a level of compromise among primer specific criteria was established. The chosen markers (chromosome location between brackets) were: *S0008* (I), *SW986* (V), *SW1129* (VI), *SW1701* (VII), *SW1517* (II) and *SW828* (III)—(GenBank access numbers: M97235, AF235422, AF235199, AF235485, AF253650, AF253852). Relevant information about the markers can be accessed through the websites of the NAGRP Pig Genome Coordination Program (http://www.animalgenome.org/resources/fprimerset9.html) and the US Meat Animal Research Centre (http://www.marc.usda.gov/genome/genome.html). The markers were divided in two triplex amplification sets (*S0008*, *SW986*, *SW1129* and *SW1701*, *SW1517*, *SW828*). Both sets were amplified in an Eppendorf Mastercycler® device, using a Qiagen Multiplex PCR Kit®, adding 2 µl of Qsolution® to the reaction mix, following manufacturer's instructions. An annealing temperature of 58 °C was applied to all markers. The final concentration of each primer was 0.2 µM, and 2.5–5 µl of *Chelex* extract were used, in a final reaction volume of 25 µl. For both triplex sets, the amplification program consisted of 15 min at 95 °C; 30 cycles of 30 s at 94 °C, 3 min at 58 °C, 60 s at 72 °C, and a final extension of 30 min at 60 °C. Amplified products were analysed by capillary electrophoresis in an automated sequencer ABI PRISM™ 310 and allele sizing was performed with GENESCAN® (v.3.1.2, Applied Biosystems).

Data analysis

After genotyping all individuals, allele frequencies for the six loci and potential deviation from Hardy–Weinberg

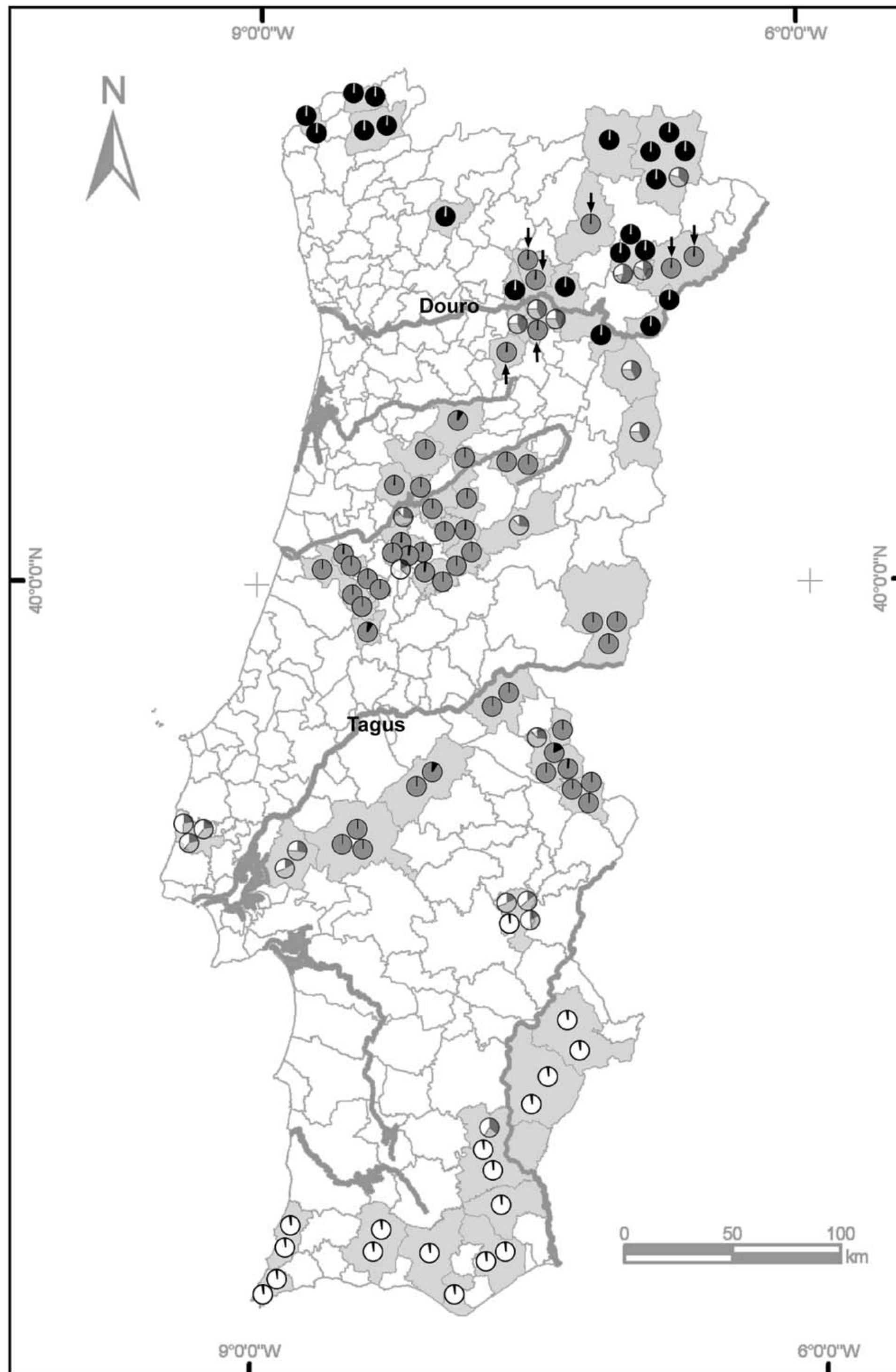

Fig. 1. Map of Portugal showing municipal boundaries. Surveyed municipalities are shaded in grey. Individual likelihoods of assignment to each cluster (Q), estimated using STRUCTRURE, are plotted for each individual (represented by a pie chart) within the study area. Pie areas of different colours represent likelihoods of assignment to each of the three inferred clusters. Unassigned individuals are plotted as semi-transparent pies. Portuguese main rivers are depicted by bold grey lines. See text for the meaning of black arrows.

equilibrium (HWE) as well as linkage equilibrium (LE) were estimated for the microsatellite allele distribution of the wild boar population, using ARLEQUIN version 2.000 (Schneider et al. 2000), applying a modified Fisher's exact test (Guo and Thompson 1992). Results of multiple comparisons were Bonferroni-corrected. Population structure was assessed using

STRUCTURE 2.1 (Pritchard et al. 2000), as were individual assignment probabilities to each one of the inferred clusters. Several scenarios were simulated, varying the parameter set values (including the number of clusters to be simulated, the presence of admixture and allele frequencies correlation), to assess the robustness of results (Table 1). Three to five replicate runs were performed for each simulated scenario.

A preliminary search for migrants and hybrids between major regional populations (henceforth called subpopulations) was based on the results of the first STRUCTURE analysis, without any *a priori* information about the geographic origin of individuals. On the basis of these results, an *ad hoc* assignment was performed for all individuals to one of the three inferred clusters. These sets of individuals were tested for HWE; however, their allele distribution departed strongly from equilibrium conditions.

Detection of migrants and hybrids between subpopulations was performed based on the results of STRUCTURE 2.1 and NEWHYBRIDS 1.0 (Anderson and Thompson 2002). Analysis with NEWHYBRIDS included all individuals and consisted of a pairwise analysis of the groups of individuals initially assigned to each of the inferred subpopulations. Sets of representative individuals were pairwise analysed for hybrid or parental origin of each individual. Only parental classes and *F1*-hybrids were considered in this analysis. Three replicate runs were performed for each pair of subpopulations, with *burnin* lengths of 5×10^4 and run lengths of 10^5 iterations. There was general agreement among likelihood distributions from different replicate runs.

Information gathered from the two approaches was used to define groups of individuals that were representative of the previously identified clusters. All individuals with potential migrant or hybrid origin as well as those that we were unable to assign were excluded from the analysis. As an objective criterion, we also excluded all individuals that were assigned to the putative parental population (defined by the initial results from STRUCTURE and data on their geographic origin) with a likelihood of <0.60. Exceptions were made for seven individuals (marked with black arrows in Figs. 1 and 2) that were captured close to the border between North and Centre subpopulations, and were considered to have been miss-assigned. These were re-assigned to the Centre subpopulation. On the basis of the defined sets of individuals, ARLEQUIN was used to assess potential HWE and LE deviations as well as to estimate allele frequencies and gene diversity for the inferred subpopulations and the total population.

Bottleneck analysis was also carried out with these new sets of individuals in order to reduce any possible bias due to structure, migration and admixture. This analysis was carried out using the program *M ratio* (Garza and Williamson 2001), where M is defined as the ratio between the number k of observed alleles of a given locus and the range r of the distribution of allele sizes for that microsatellite locus. The software calculates an average M value for stable theoretical populations as well as a critical M, above which 95% of the ratios for equilibrium populations are placed. Both average and critical M were calculated considering the same sample size of the studied subpopulations and given the parameters of the model: *ps*—proportion of mutations involving just one repeat unit; Δg—average size of mutations evolving more than one repeat unit; θ—parameter based on effective population size previous to the bottleneck and mutation rate. Two sets of parameter values were simulated: a theoretical, conservative one (Garza and Williamson 2001), with $\Delta g = 3.5$ (Δg: mean size of larger mutations) and $ps = 0.9$ (*ps*: mean % of mutations that add or delete only one repeat), and another one based on published data with $\Delta g = 2.8$ and $ps = 0.88$ (based on empirical mutation data from literature Garza and Williamson 2001). The parameter θ was allowed to vary over several orders of magnitude to account for a range of mutation rates between 10^{-4} and 10^{-5} (Jarne and Lagoda 1996; Yue et al. 2002). By varying θ over several orders of magnitude, we also accounted for the differences in effective population size. For example, considering a mutation rate of 7.52×10^{-5} and with θ set at 0.01, effective population size would be 33. With θ equal to 10, effective population size would be greater than 30,000. We found this approach to be conservative, excluding the need to estimate effective population size (Vernesi et al. 2003).

Pairwise genetic distances among the inferred populations were estimated using the parameters F_{ST} (Wright 1951) and its analogue R_{ST} (Slatkin 1995), the latter designed for microsatellites. These parameters were estimated with ARLEQUIN, using the estimators introduced by Weir and Cockerham (1984) and Rousset (1996), respectively. Global F_{IT}, F_{ST} and F_{IS} were estimated for the total population according to Weir and Cockerham (1984), using FSTAT (v. 2.9.3; Goudet 1995). Inbreeding coefficients for each subpopulation were also calculated using FSTAT.

Results

Preliminary structure analysis

Allele number, allele size range and private alleles for each marker and subpopulation are given in Table 2.

Table 1. Results from STRUCTURE simulations, without **a priori** origin information

Simulation	K tested	Replicates	*Burnin* length	Run length ($\times 10^6$)	Best K	$P(K/X)$
a	1–10	5	5×10^4	1	3	1.00
b	1–5	3	5×10^5	2.5	3	1.00
c	2–4	3	5×10^5	1	3	1.00
d	2–4	3	5×10^5	1	3	1.00

a, *b*: simulation with setting ⟨correlated frequencies model⟩; *c*: the same, with a predefined at 0.0714; *d*: the same, with different a for each population; *K*: number of clusters being considered; replicates: number of replicated runs; *burnin* and run lengths: length of each period in number of iterations; and best*K*: *K* number of clusters with higher likelihood, and associated likelihood $P(K|X)$

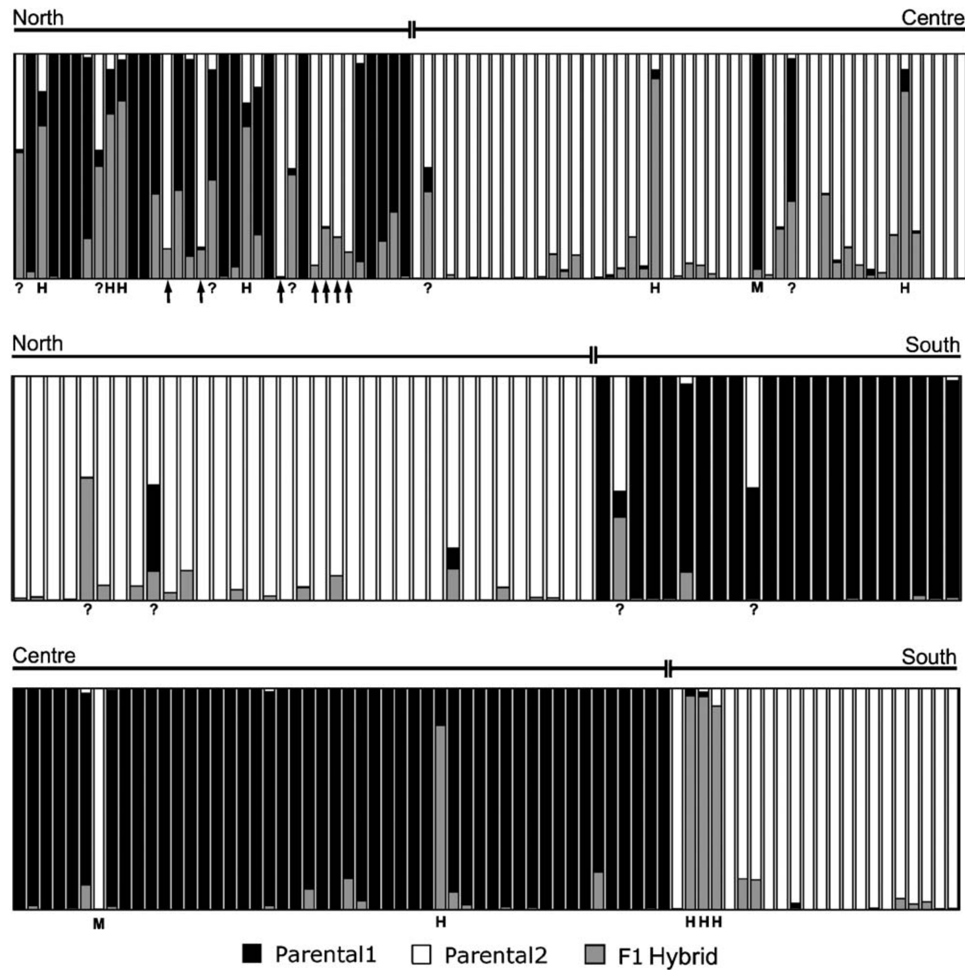

Fig. 2. Results from the simulation considering two parental and one hybrid classes, for each pair of subpopulations, using NEWHYBRIDS. Each individual is represented by a vertical bar. Likelihoods of assignment to parental and *F1* classes are plotted in different colours. Horizontal bars represent *a priori* geographic origin. H represents a potential hybrid, M represents migrant and ? represents unassigned individuals. See text for the meaning of black arrows.

Table 2. Allele number, size range and private alleles for each marker and (sub) population

Locus	North ($n = 20$)			Centre ($n = 51$)			South ($n = 18$)			Portugal ($n = 110$)	
	N_a	Size range	A_{priv}	N_a	Size range	A_{priv}	N_a	Size range	A_{priv}	N_a	Size range
S0008	8	175–193	193	5	175–195	195	8	165–191	165, 179	11	165–195
SW986	3	147–151		5	135–159	135, 137, 159	3	147–151		6	135–159
SW1129	5	139–159		8	139–159	143, 149	7	139–177	167, 169, 177	11	139–177
SW1701	7	90–128	128	14	90–132	92, 110, 112, 130, 132	5	90–122		15	90–132
SW1517	8	132–148		11	118–154	126, 154	8	118–158	152, 156, 158	15	118–158
SW828	3	211–221		3	211–221		2	211–221		3	211–221

N_a: number of alleles; size range: smallest and largest allele for each group of samples; and A_{priv}: private alleles.

Allelic distribution at four out of six loci showed statistically significant deviations from HWE, after Bonferroni correction for multiple comparisons (Table 3). Individual allele distributions are presented in Table 4.

For all scenarios simulated with STRUCTURE, there was excellent agreement among replicate runs and also among different simulations. The estimated number of clusters (subpopulations) in all simulations was $K = 3$. There was a strong overall agreement between the individual assignments resulting from the three clusters model, based on individual genotypes, and the geographic origin of the individuals analyzed. On the basis of that, the clusters were subsequently labelled **North**, **Centre** and **South**.

Table 3. Observed and expected heterozygosity *per locus*, for the total population and for each subpopulation (sample size)

	Portugal (110)		North (20)		Centre (51)		South (18)	
	H_0	H_E	H_0	H_E	H_0	H_E	H_0	H_E
S0008	0.709	**0.783**	0.600	0.756	0.745	0.640	0.778	0.819
SW986	0.500	0.562	0.500	0.576	0.412	0.424	0.722	0.637
SW1129	0.618	**0.759**	0.250	0.397	0.725	0.806	0.667	0.767
SW1701	0.773	**0.886**	0.850	0.860	0.824	0.864	0.500	0.717
SW1517	0.718	**0.832**	0.750	0.783	0.627	0.734	0.667	0.800
SW828	0.445	0.523	0.400	0.559	0.431	0.405	0.444	0.494
Average	0.6273	0.7242	0.5583	0.6552	0.6275	0.6452	0.6296	0.7056
S.D.	0.1193	0.1350	0.2029	0.1584	0.1565	0.1772	0.1186	0.1121
Pair loci LD				0/15		1/15		0/15

H_0: observed; H_E: expected under HWE; bold figures: significant deviation from HWE proportions; average heterozygosity across all *loci*, and respective standard deviation; and pairs loci LD: number of pairs of loci in linkage disequilibrium, out of 15 possible combinations.

Migrants and hybrids among subpopulations

For most individuals, the genotype-based assignment matched their previous geographic clustering criterion (Fig. 2). However, according to their genotype, some individuals were either clearly miss-assigned (marked with black arrows in Figs. 1 and 2), appeared to be *F1* hybrids or migrants or their origin could not be determined (marked respectively with capital H, M or a question mark in Fig. 2).

Definition of subpopulation representative sets

For the definition of a representative set of individuals for each subpopulation, we excluded hybrids, migrants and individuals whose origin could not be determined. One exception was made to a set of individuals (marked with dark arrows in Figs. 1 and 2), which were initially assigned to the **North** subpopulation. These individuals came from locations near the suture zone between **North** and **Centre** subpopulations, and after the pairwise analysis with NEWHYBRIDS were shown to have genotypes more closely related to the **Centre** population. In contrast to the total population, the sets of individuals representative of each subpopulation showed no significant deviations from either HWE or LE, after Bonferroni correction, with the exception of **Centre**, where there was significant deviation from LE conditions for one pair of loci out of 15 possible combinations. On the basis of the resulting sets of individuals: **North** ($n = 20$), **Centre** ($n = 51$) and **South** ($n = 18$), a new simulation was performed using STRUCTURE. All 110 individuals were included in the analysis, but **a priori** origin information was only included for individuals from the referred sets, representative of each subpopulation. Three replicates were run.

Allele frequency correlation was considered and admixture was allowed among subpopulations. Average likelihood of assignment to each of the three inferred clusters, weighted across all individuals of each subpopulation, was estimated and is presented in Table 5.

For each individual assigned to a subpopulation, the estimated membership coefficients for each individual in each cluster were plotted as individual pie charts in a map of the study area (Fig. 1). Each pie chart represents one individual and is placed inside the borders of the municipality where it was captured.

Detection of recent bottlenecks

Critical M values, as well as the average M ratios for the three subpopulations were estimated based on the number of individuals and three variable parameters *ps*, Δg and θ (Fig. 3). Sample M ratios were calculated taking into account the allele distribution of the sample and not the parameters used to estimate average and critical M values (by modulation). Calculated M ratios for North, Centre and South subpopulations were 0.666, 0.547 and 0.488, respectively. Estimated critical M ratios varied (depending on the parameters) between 0.617 and 0.833 (North), 0.678 and 0.833 (Centre), and 0.612 and 0.828 (South). Average M ratios varied between 0.745 and 0.949 (North), 0.795 and 0.950 (Centre), and 0.741 and 0.949 (South). The average M value represents the M ratio from an average equilibrium population (given the set of referred parameters). The critical M is the value above which 95% of ratios from equilibrium populations should lie. If the calculated sample M ratio was smaller than the critical M ratio (with a 0.05 probability of error), we assumed that the population had passed trough a recent bottleneck.

In general, the calculated M ratios for the three subpopulations were smaller than the corresponding

Table 4. Allele frequencies distributions at the six loci analyzed, given separately for each subpopulation and for the total population

Local	Alelle	North (20)	Centre (51)	South (18)	Portugal (110)
S0008	165			0.056	0.009
	175	0.105	0.548	0.194	0.364
	179			0.056	0.014
	181	0.053		0.333	0.073
	183	0.026	0.038		0.032
	185	0.342	0.125	0.083	0.159
	187	0.368	0.240	0.056	0.223
	189	0.026		0.028	0.009
	191	0.026		0.194	0.073
	193	0.053		0.194	0.073
	195		0.048		0.036
SW986	135		0.019		0.009
	137		0.029		0.014
	147	0.421	0.202	0.194	0.273
	149	0.526	0.740	0.306	0.595
	151	0.053		0.500	0.105
	159		0.010		0.005
SW1129	139	0.026	0.029	0.028	0.045
	141	0.842	0.327	0.444	0.432
	143		0.019		0.009
	149		0.019		0.014
	153		0.135	0.083	0.086
	155	0.026	0.163	0.222	0.177
	157	0.053	0.183		0.118
	159	0.053	0.125		0.073
	167			0.139	0.032
	169			0.056	0.009
	177			0.028	0.005
SW1701	90	0.184	0.029	0.361	0.123
	92		0.010		0.009
	106	0.158	0.010		0.036
	108	0.263	0.192		0.177
	110		0.144		0.086
	112		0.019		0.018
	114		0.173	0.389	0.191
	118		0.010	0.139	0.027
	120	0.105	0.077	0.083	0.086
	122		0.202	0.028	0.109
	124	0.079	0.010		0.032
	126	0.053	0.096		0.059
	128	0.158			0.027
	130		0.019		0.009
	132		0.010		0.009
SW1517	118		0.010	0.111	0.023
	126		0.010		0.005
	132	0.026	0.010		0.018
	134		0.481	0.250	0.327
	136	0.026	0.077		0.055
	138	0.079	0.058	0.028	0.059
	140	0.421	0.144		0.164
	142	0.079	0.087		0.077
	144	0.105	0.058	0.417	0.141
	146	0.053		0.056	0.027
	148	0.211	0.058		0.073
	152			0.028	0.05
	154		0.010		0.005
	156			0.083	0.018
	158			0.028	0.005
SW828	211	0.421	0.144	0.667	0.332
	217	0.053	0.087		0.064
	221	0.526	0.769	0.333	0.605

Table 5. Average proportions of genotype assignments from each **a priori** origin to each of the inferred clusters (subpopulations)

Subpopulations	Inferred clusters			
	n	1	2	3
North	20	0.991	0.003	0.005
Centre	51	0.011	0.986	0.003
South	18	0.002	0.003	0.995
Undetermined	21	0.316	0.379	0.305

critical and average *M* ratio values. The only exception was for the **North** subpopulation and for the most extreme simulated scenario (with a mutation rate of 7.52×10^{-5} and an effective population size of more than 30,000), when the sample ratio was equal or greater than the critical ratio. This result clearly suggests a recent bottleneck in the inferred subpopulations, and nationally in the wild boar population of Portugal.

Genetic distances

Genetic distances between pairs of subpopulations, calculated using R_{ST} and F_{ST} estimators, were always highly significant ($P<0.001$) with exception of the R_{ST} estimate for the **North**/**South** pair of subpopulations, which was also significant but at a lower probability level ($P<0.01$). Respective estimated F_{ST} and R_{ST} values per pair of subpopulations were 0.138 and 0.198 (**North**/**Centre**), 0.155 and 0.107 (**North**/**South**) and 0.178 and 0.153 (**Centre**/**South**). All estimated values corresponded to moderate to high values of genetic differentiation (Wright 1978). Estimated levels of autozygosity, for total population, resulting from random genetic drift (F_{ST}) and inbreeding (F_{IS}) and both processes (F_{IT}), were 0.155, 0.056 and 0.202, respectively. Estimated inbreeding coefficients for each subpopulation were: 0.133 (**North**), 0.014 (**Centre**) and 0.087 (**South**).

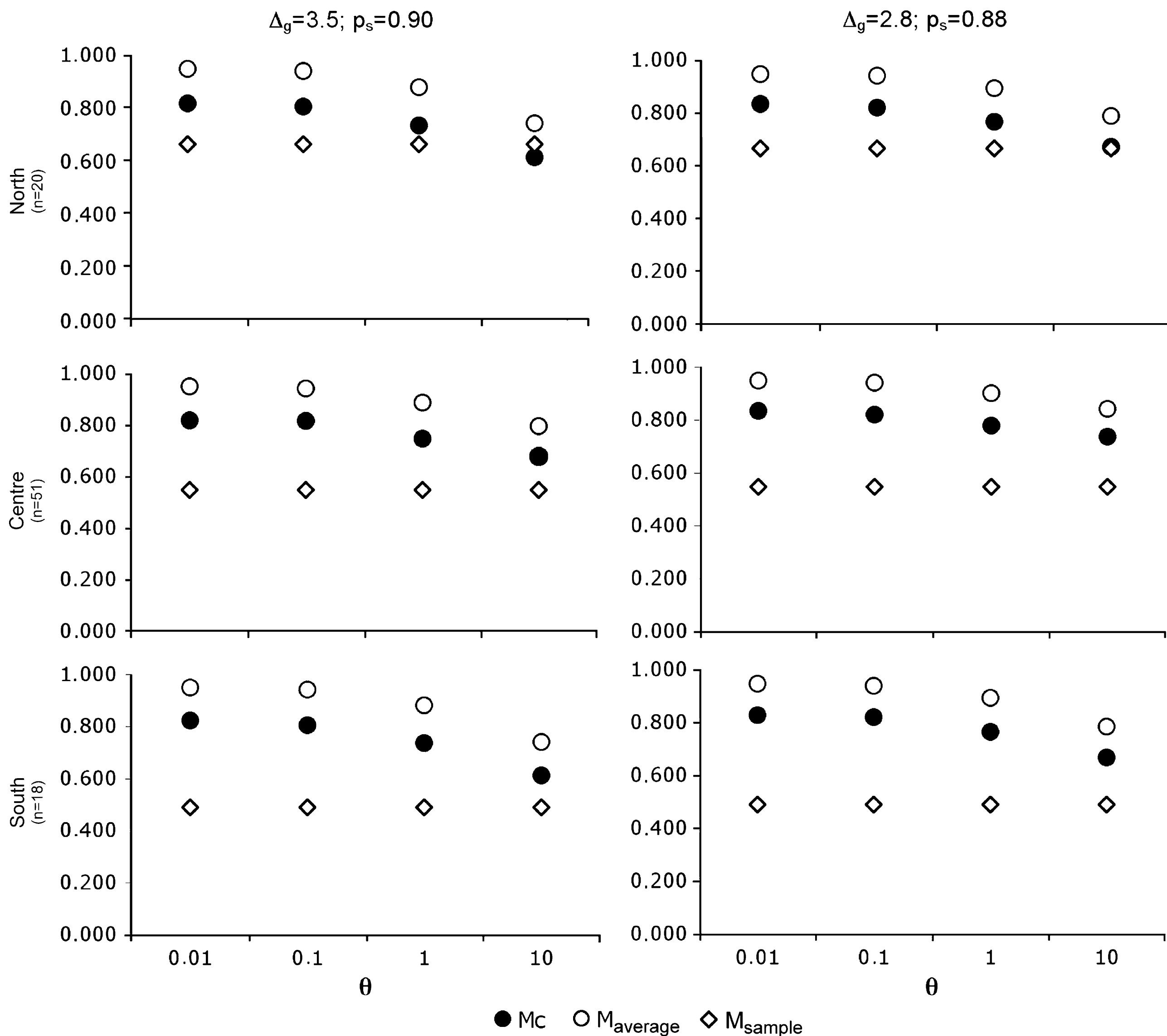

Fig. 3. *M* ratio graphics for each of the subpopulations and for both simulated sets of parameter values (Δg and ps). Sample *M* ratio, critical M_c and average *M* values are presented for each simulated scenario. Simulated values of θ are also indicated.

Discussion

Population structure

For many species, the geographic distribution is greater than the dispersal capacity of their individuals and most natural populations are structured into local populations (Nei and Li 1972; Balloux and Lugon-Moulin 2002). Wright (1951) noted that a certain degree of population structure could represent an evolutionary advantage that facilitates a higher genetic diversity than a panmitic population. Deviation from HWE conditions could represent evidence of population structure (Hartl and Clark 2007). In the same way, linkage disequilibrium (between unlinked *loci*) in a population might also indicate an admixture between subpopulations with different allele frequencies (Nei and Li 1972). For the population studied here, such a deviation from HWE conditions was found, indicating that the wild boar population in Portugal is genetically structured. Further support for this scenario was provided by the results obtained by STRUCTURE analysis. Without any *a priori* information about geographic location of the individuals, a model with $K = 3$ populations was suggested, always associated with a strong likelihood.

Similar results had been suggested in a previous study (Ferreira et al. 2006) that employed a topological approach, where the two main rivers (Tagus and Douro) where assumed to separate the three putative subpopulations. In the previous study, F_{ST} estimates for pairwise distances between subpopulations were always significant. However, in the study presented here, with an

enlarged sample set and based on a different approach, estimated pairwise distances were on average 2.5-fold greater than the previously estimated values by Ferreira et al. (2006). On the basis of our new approach, we were also able to identify individuals that belonged to the same subpopulation but were present on both sides of a river; sometimes they were sampled more than 50 km beyond the river (e.g. in Fig. 1, **Centre** subpopulation individuals were found south of the river Tagus). On the other hand, the northern limit of the **South** subpopulation lies well south of the river Tagus, suggesting that this river might not be responsible for limiting the dispersal of individuals from this subpopulation. However, we do not exclude the possibility that the two rivers systems may play a role in the structure of this population (Ferreira et al. 2006).

The global *F*-statistics indicate that the autozygosity component due to population subdivision ($F_{ST} = 0.155$) is about three times greater than that relative to inbreeding ($F_{IS} = 0.056$), pointing to a major role of population structure in accounting for the reduction in the expected heterozygosity. On the other hand, the inbreeding coefficient estimated for the **Centre** subpopulation is, respectively, about ten and six times lower than for the **North** and the **South** subpopulations. This may indicate that inbreeding plays a more important role in these subpopulations, which are represented in the overall boar sample by less individuals and are also confined to a smaller geographic area. Indeed, the highest inbreeding coefficient was found in the **North** subpopulation, which inhabits a smaller geographic range, while the central subpopulation presented the smallest inbreeding coefficient. Combining the results from STRUCTURE and NEWHYBRIDS simulations enabled the definition of sets of individuals that could be used as representatives of respective subpopulations, excluding all individuals with a potential hybrid or migrant origin. Both methods provided good assignment likelihoods without *a priori* geographic information. The NEWHYBRIDS approach, despite being limited to pairwise analyses, was useful for the detection of potential hybrids and migrants that were not detected by STRUCTURE. Assignment probabilities (Table 5) of individuals in the *a priori* attributed subpopulations were on average higher ($Q \approx 0.99$) than in previous studies, such as that of Vernesi et al. (2003) with wild boar populations ($Q > 0.86$), and that of Andersone et al. (2002) with wolf and dog populations ($Q > 0.93$). In both of these studies, the populations being compared were geographically much more distant than the wild boar subpopulations studied here.

Genetic distances

According to Wright (1978), F_{ST} estimators with values between 0.05 and 0.15 represent a moderate genetic differentiation, while values between 0.15 and 0.25 represent high, and values above 0.25 represent very high genetic differentiation. Theoretically, these estimators could assume values between 0 and 1. However, some authors (e.g. Cornuet et al. 1999; Hedrick 1999; Kalinowski 2002) noted that the expected homozygosity in the subject populations represents the maximum possible value for these estimators. For the subpopulations in our study, this value was equal to 0.33 (equal to $1 - H_{expected}$, the latter calculated as average of expected heterozygosities for subpopulations), which would represent total fixation for different alleles in the different subpopulations. For the wild boar population in Portugal, estimated distances among subpopulations were on average 0.157, half of this total fixation value. Estimated genetic distances between the identified subpopulations always indicated a moderate to high genetic differentiation and were both statistically and biologically significant. According to Hardy et al. (2003), a strong influence of mutation in differentiation would cause an increment of the ratio R_{ST}/F_{ST}, but when the main force of differentiation is genetic drift, these estimates tend to be similar. The latter seems to be the case for the wild boar population in Portugal, because the average F_{ST} value (0.157) is close to (and in fact greater than) the average R_{ST} (0.153). However, according to Balloux and Lugon-Moulin (2002), when populations are subjected to isolation by distance, geographically proximate populations tend to be genetically more similar than geographically distant populations. Although the F_{ST} estimate was greater for the **North/South** pair than for the **North/Centre** and **Centre/South** pairs, the opposite was apparent for the R_{ST} estimates. Taking the average of these values, the distance estimates for **North/South** (0.131) were smaller than those for the **North/Centre** (0.168) and **Centre/South** comparisons (0.165). Genetic distance estimates for these subpopulations appear not to conform to a simple model of isolation by distance.

Bottlenecks

The data presented in Tables 2 and 4 indicate the presence of several private alleles and fragmented allele distributions. According to England et al. (2003), lack of rare alleles in some populations and fragmented distribution of allele frequencies can be indicative of bottlenecks. In addition, the presence of private alleles may also be explained by population bottlenecks: severe population contractions led to lack of gene flow (Hellborg et al. 2002). The interpretation of recent demographic events, particularly bottlenecks, is generally based on the assumption that an isolated and non-structured population is being analysed. Population

structure, as well as migration or admixture among subpopulations, can mask the effects of a bottleneck (Cornuet and Luikart 1996; Garza and Williamson 2001). To reduce these effects, bottleneck detection was performed on each of the defined subpopulations rather than the overall population. Significantly low sample M ratio, when compared with critical M values, also constitutes evidence of a recent bottleneck (Garza and Williamson 2001). Significantly, low M ratios were found for all three subpopulations (Fig. 3). The only exception was for the **North** subpopulation and for the most extreme θ value (considering a mutation rate of 7.52×10^{-5} and an effective population size of more than 30,000). In this case, despite the estimated sample M ratio being below the average M value for an equilibrium population, it was above the critical value, where 95% of equilibrium populations are found. However, for this simulation scenario, the M ratio was equal to 0.666, lower than the theoretical minimum of 0.68 reported by Garza and Williamson (2001) for equilibrium populations. Estimated M values for the subpopulations in this study (average 0.567, maximum 0.666) were also smaller than those obtained for wild boar populations (minimum 0.70, average 0.805) by Vernesi et al. (2003), which were considered to represent stable populations. However, in their study, the lowest value of M (0.70), for the Hungarian population, was also significantly lower than the critical M value for stable populations. According to Garza and Williamson (2001), the recovery of M could be slower than for measurements based on the deficit of rare alleles, because not all mutations will increase its value, and also because this ratio will preserve information about bottlenecks for a longer period. The evidence for the occurrence of a bottleneck is consistent with the indicated prevalence of stochastic processes such as random genetic drift and corresponds with the relationship between F_{ST} and R_{ST} discussed above. We therefore conclude that the detected bottleneck in the wild boar population in Portugal had an important role in the development of the present structure of this population. Despite past fragmentation, the distribution of the three subpopulations illustrated in Fig. 1 suggests that they are now more or less contiguous, with some evidence of hybridisation and migration. In this respect it is important to note that wild boar in Portugal is extensively managed and hunted and it is therefore uncertain to what extent the detected levels of migration and hybridisation are natural phenomena. Whenever the authors contacted local hunting associations, regional or national federations, or forestry services, it was frequently reported that illegal practices such as the introduction of wild boars from different areas was regularly performed without any control or direct knowledge of competent authorities.

Conclusions

In this study, we report the detection of a genetic structure in the wild boar population in Portugal that may have resulted from a recent bottleneck from which the population has not yet fully recovered. Such an event is historically documented for the first half of the 20th century. We found supporting evidence of this previous bottleneck in Portuguese boar populations, followed by an expansion from refugial areas towards historical distributions. Nowadays, wild boar is widespread throughout the national territory and divided in three subpopulations, with overlapping distribution ranges and admixture in the contact areas. According to the historical data presented in the introduction, the recovery of the wild boar population in Portugal after the bottleneck appears to have occurred within a few decades following the ban on wild boar hunting in the 1960s.

Acknowledgements

The authors thank Dr. Max F. Rothschild (The USDA Supported US Pig Genome Coordination Project) for providing the primers used in this study; all those who helped to collect samples and the Confederação Nacional dos Caçadores Portugueses, for financial support. We also acknowledge Professor Ettore Randi for useful suggestions on data analysis, Carlos Ferreira for valuable help with the maps and images and Kieran Monaghan for help with English editing.

This study was partially supported by FCT (Science and Technology Portuguese Foundation) through an MSc grant to Eduardo Ferreira (Ref. SFRH/BM/13034/2003).

References

Anderson, E.C., Thompson, E.A., 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160, 1217–1229.

Andersone, Z., Lucchini, V., Randi, E., Ozolins, J., 2002. Hybridisation between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. Mamm. Biol. 67, 79–90.

Archibald, A.L., Haley, C.S., Brown, J.F., Couperwhite, S., McQueen, H.A., Nicholson, D., Coppieters, W., Vandeweghe, A., Stratil, A., Wintero, A.K., Fredholm, M., Larsen, N.J., Nielsen, V.H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J.C., Burgaud, G., Ollivier, L., Bidanel, J.P., Vaiman, M., Renard, C., Gelderman, H., Davoli, R., Ruyter, D., Verstege, E.J.M., Groenen, M.A.M., Davies, W., Hoyheim, B., Keiserud, A., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J.R., Dear, D.V.A., Signer, E.,

Jeffreys, A.J., Moran, C., Letissier, P., Muladno, Rothschild, M.F., Tuggle, C.K., Vaske, D., Helm, J., Liu, H.C., 1995. The PiGMaP consortium linkage map of the pig (*Sus scrofa*). Mamm. Genome 6, 157–175.

Balloux, F., Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11, 155–165.

Cabral, M.J., Almeida, J., Almeida, P.R., Delinger, T., Ferrand de Almeida, N., Oliveira, M.E., Palmeirim, J.M., Queiroz, A.L., Rogado, L., Santos-Reis, M., 2005. Livro Vermelho dos Vertebrados de Portugal. Instituto de Conservação da Natureza, Lisboa (in Portuguese).

Cornuet, J.M., Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144, 2001–2014.

Cornuet, J., Piry, S., Luikart, G., Estoup, A., Solignac, M., 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153, 1989–2000.

England, P.R., Osler, G.H.R., Woodworth, L.M., Montgomery, M.E., Briscoe, D.A., Frankham, R., 2003. Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conserv. Genet. 4, 595–604.

Fernandez-De-Mera, I., Gortazar, C., Vicente, J., Höfle, U., Fierro, Y., 2003. Wild boar helminths: risks in animal translocations. Vet. Parasitol. 115, 335–341.

Ferreira, E., Souto, L., Soares, A.M.V.M., Fonseca, C., 2006. Genetic structure of the wild boar (*Sus scrofa* L.) population in Portugal. Wildl. Biol. Pract. 2 (1), 17–25.

Fickel, J., Hohmann, U., 2006. A methodological approach for non-invasive sampling for population size estimates in wild boars (*Sus scrofa*). Eur. J. Wildl. Res. 52, 28–33.

Fonseca, C., 2004. Population dynamics and management of wild boar (*Sus scrofa* L.) in Central Portugal and Southeastern Poland. PhD Thesis, University of Aveiro.

Fonseca, C., Santos, P., Monzón, A., Bento, P., Alves da Silva, A., Alves, J., Silvério, A., Soares, A.M.V.M., Petrucci-Fonseca, F., 2004. Reproduction in the Wild Boar (*Sus scrofa* Linnaeus, 1758) populations of Portugal. In: C. Fonseca, J. Herrero, A. Luís, A.M.V.M. Soares (Eds.), Wild Boar Research 2002: A Selection and Edited Papers from the 4th International Wild Boar Symposium. Galemys 16, 53–65.

Garza, J.C., Williamson, E.G., 2001. Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. 10, 305–318.

Giuffra, E., Kijas, J.M.H., Amarger, V., Carlborg, Ö., Jeon, J.T., Andersson, L., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics 154, 1785–1791.

Goudet, J., 1995. FSTAT (vers. 1.2): a computer program to calculate F-statistics. J. Hered. 86, 485–486 (upgrade to vers. 2.9.3. in 2001).

Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 48, 361–372.

Hampton, J.O., Spencer, P.B.S., Alpers, D.L., Twigg, L.E., Woolnough, A.P., Doust, J., Higgs, T., Pluske, J., 2004. Molecular techniques, wildlife management and the importance of genetic population structure and dispersal: a case study with feral pigs. J. Appl. Ecol. 41, 735–743.

Harcet, M., Dikic, M., Gamulin, V., 2006. Low Genetic diversity of the Turopolje Pig Breed. Food Technol. Biotechnol. 44 (1), 105–109.

Hardy, O.J., Charbonnel, N., Fréville, H., Heuertz, M., 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163, 1467–1482.

Hartl, D.L., Clark, A.G., 2007. Principles on Population Genetics, fourth ed. Sinauer Associates Inc., Sunderland.

Hedrick, P.W., 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53 (2), 313–318.

Hellborg, L., Walker, C.W., Rueness, E.K., Stacy, J.E., Kojola, I., Valdmann, H., Vilà, C., Zimmermann, B., Jakobsen, K.S., Ellegren, H., 2002. Differentiation and levels of genetic variation in northern European lynx (*Lynx lynx*) populations revealed by microsatellites and mitochondrial DNA analysis. Conserv. Genet. 3, 97–111.

Jarne, P., Lagoda, P.J.L., 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11 (10), 424–429.

Kalinowski, S.T., 2002. Evolutionary and statistical properties of three genetic distances. Mol. Ecol. 11, 1263–1273.

Keuling, O., Stier, N., Roth, M., 2007. Annual and seasonal space use of different age classes of female wild boar *Sus scrofa* L. Eur. J. Wildl. Res., doi:10.1007/s10344-007-0157-4.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Anderson, L., Cooper, A., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307, 1618–1621.

Lemel, J., Truvé, J., Söderberg, B., 2003. Variation in ranging and activity behaviour of European wild boar *Sus scrofa* in Sweden. Wildl. Biol. 9 (1), 29–36.

Lorenzini, R., 2005. DNA forensics and the poaching of wildlife in Italy: a case study. Forensic Sci. Int. 153, 218–221.

Lowden, S., Finlayson, H.A., MacDonald, A.A., Downing, A.C., Goodman, S.J., Leus, K., Kaspe, L., Wahyuni, E., Archibald, A.L., 2002. Application of *Sus scrofa* microsatellite markers to wild suiforms. Conserv. Genet. 3, 347–350.

Massei, G., Genov, P.V., Staines, B.W., 1996. Diet, food availability and reproduction of wild boar in a Mediterranean coastal area. Acta Theriol. 41 (3), 307–320.

Melzer, F., Lohse, R., Nieper, H., Liebert, M., Sachse, K., 2006. A serological study on brucellosis in wild boars in Germany. Eur. J. Wildl. Res. 53, 153–157.

Nei, M., Li, W., 1972. Linkage disequilibrium in subdivided populations. Genetics 75, 213–219.

Okumura, N., Ishiguro, N., Nakano, M., Hirai, K., Matsui, A., Sahara, M., 1996. Geographic population structure and sequence divergence in the mitochondrial DNA control region of the Japanese wild boar (*Sus scrofa leucomystax*)

with reference to those of domestic pigs. Biochem. Genet. 34, 179–189.

Pritchard, J.K., Stephens, M., Donelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Randi, E., 1995. Conservation genetics of the genus *Sus*. Ibex J.M.E. 3, 6–12.

Rosell, C., Herrero, J., 2002. *Sus scrofa* Linnaeus, 1758, pp 306–309. In: Palomo, L.J., Gisbert, J. (Eds.), Atlas de los Mamíferos Terrestres de España. Direccíon General de Conservatíon de la Naturaleza—SECEM-SECEMU, Madrid.

Rothschild, M.F., 2003. Advances in pig genomics and functional gene discovery. Comp. Funct. Genom. 4, 266–270.

Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142, 1357–1362.

Sancristobal, M., Chevalet, C., Haley, C.S., Joosten, R., Rattink, A.P., Harlizius, B., Groenen, M.A.M., Amigues, Y., Boscher, M.-Y., Russell, G., Law, A., Davoli, R., Russo, V., Desautes, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J.V., Vega-Pla, J.L., Martinez, A.M., Glodek, P., Meyer, J.N., Gandini, G.C., Matassino, D., Plastow, G.S., Siggens, K.W., Laval, G., Archibald, A.L., Milan, D., Hammond, K., Cardellino, R., 2006. Genetic diversity within and between European pig breeds using microsatellite markers. Anim. Genet. 37, 189–198.

Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin Version 2.000: Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462.

Spencer, P.B.S., Hampton, J., Lapidge, S.J., Mitchell, J., Lee, J., Pluske, J.R., 2006. An assessment of the genetic diversity and structure within and among populations of wild pigs (*Sus scrofa*) from Australia and Papua New Guinea. J. Genet. 85 (1), 63–66.

Spitz, F., Janeau, G., 1990. Spatial strategies: an attempt to classify daily movements of wild boar. Acta Theriol. 35, 129–149.

Ursing, B.M., Arnason, U., 1998. The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). J. Mol. Evol. 47, 302–306.

Vernesi, C., Crestanello, B., Pecchioli, E., Tartari, D., Caramelli, D., Hauffe, H., Bertorelle, G., 2003. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. Mol. Ecol. 12, 585–595.

Vidal, D., Naranjo, V., Mateo, R., Gortazar, C., Fuente, J., 2006. Analysis of serum biochemical parameters in relation to *Mycobacterium bovis* infection of European wild boars (*Sus scrofa*) in Spain. Eur. J. Wildl. Res. 52, 301–304.

Walsh, P.S., Metzer, D.A., Higuchi, R., 1991. Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10, 506–513.

Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38 (6), 1358–1370.

Wright, S., 1951. The genetical structure of populations. Ann. Eugen. 15, 323–354.

Wright, S., 1978. Evolution and the Genetics of Populations. Vol. 4. Variability Within and Among Natural Populations. University of Chicago Press, Chicago.

Yue, G.H., Beeckmann, P., Geldermann, H., 2002. Mutation rates at swine microsatellite loci. Genetica 114, 113–119.