Molecular Ecology (2008) **17**, 1745–1762 doi: 10.1111/j.1365-294X.2008.03703.x

Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?

M. SCANDURA,* L. IACOLINA,* B. CRESTANELLO,†‡ E. PECCHIOLI,‡§ M. F. DI BENEDETTO,* V. RUSSO,¶ R. DAVOLI,¶ M. APOLLONIO* and G. BERTORELLE†

Dipartimento di Zoologia e Genetica Evoluzionistica, Università di Sassari, via Muroni 25, 07100 Sassari, Italy, †Dipartimento di Biologia, Università di Ferrara, via Borsari 46, 44100 Ferrara, Italy, ‡Centro di Ecologia Alpina, Viote del Monte Bondone, 38100 Trento, Italy, §Biosfera, Biologia della Conservazione, Ricerca e Didattica, 50122 Firenze, Italy, ¶DIPROVAL, Università di Bologna, Via Rosselli, 42100 Reggio Emilia, Italy

Abstract

The European wild boar is an important game species, subjected to local extinctions and translocations in the past, and currently enormously and worryingly expanding in some areas where management is urgently required. Understanding the relative roles of ancient and recent events in shaping the genetic structure of this species is therefore not only an interesting scientific issue, but it represents also the basis for addressing future management strategies. In addition, several pig breeds descend from the European wild boar, but the geographical location of the domestication area(s) and the possible introgression of pig genomes into wild populations are still open questions. Here, we analysed the genetic variation in different wild boar populations in Europe. Ten polymorphic microsatellites were typed in 252 wild boars and the mtDNA control region was sequenced in a subset of 145 individuals. Some samples from different pig breeds were also analysed. Our results, which were obtained considering also 612 published mtDNA sequences, suggest that (i) most populations are similarly differentiated, but the major discontinuity is found along the Alps; (ii) except for the Italian populations, European wild boars show the signature of a postglacial demographic expansion; (iii) Italian populations seem to preserve a high proportion of preglaciation diversity; (iv) the demographic decline which occurred in some areas in the last few centuries did not produce a noticeable reduction of genetic variation; (v) signs of human-mediated gene flow among populations are weak, although in some regions the effects of translocations are detectable and a low degree of pig introgression can be identified; (vi) the hypothesis of an independent domestication centre in Italy is not supported by our data, which in turn confirm that Central European wild boar might have represented an important source for domestic breeds. We can therefore conclude that recent human activities had a limited effect on the wild boar genetic structure. It follows that areas with high variation and differentiation represent natural reservoirs of genetic diversity to be protected avoiding translocations. In this context controlling some populations by hunting is not expected to affect significantly genetic variation in this species.

Keywords: genetic diversity, microsatellites, mtDNA, phylogeography, population expansion, *Sus scrofa*

Received 17 September 2007; revision received 27 November 2007; accepted 11 January 2008

Introduction

Present-day levels and patterns of genetic variation are strongly affected by the characteristics of a species (e.g. dispersal rates), the landscape (e.g. geographical barriers),

Correspondence: Massimo Scandura, Fax: +39 079 228665; E-mail: scandura@uniss.it

but also by different specific events that occurred in the past (Avise 2004). For example, the genetic impact of the last glaciation was relevant to several species (Hewitt 2000), but many other processes with a potential effect on the genetic variation certainly occurred since then. This is especially true for game and domestic species because of extensive human manipulation. In this study, we focused on several European populations of wild boar (*Sus scrofa*), whose genetic variation patterns have been possibly affected by both domestication and hunting/management activities. We used a large set of molecular data in order to distinguish and estimate the effects of different processes that occurred at different timescales. In addition, this data set enabled us to test a specific hypothesis regarding the domestication of the species in Europe.

The Eurasian wild boar is one of the most widely distributed terrestrial mammals. Its geographical range, excluding recent introductions, extends from Western Europe and Northern Africa to Japan. Possibly originated in Southeastern Asia, where the highest numbers of wild pig taxa are observed (Lucchini *et al*. 2005), *S. scrofa* can be now subdivided into at least two major genetic clades roughly corresponding in the domestic form to European and Asian pigs (Giuffra *et al*. 2000; Okumura *et al*. 2001; Alves *et al*. 2003; Larson *et al*. 2005). The separation of these two clades can be dated back to between 0.5 and 0.9 million years ago, suggesting at least two independent domestication events (Giuffra *et al*. 2000; Kijas & Anderson 2001; Alves *et al*. 2003). Subspecies are usually classified into four major groups, with the European wild boar corresponding to the so-called 'Western races' group (Oliver *et al*. 1993).

The quaternary climatic oscillations, and in particular the last glaciation and the subsequent warm period, produced remarkable consequences on the levels and patterns of genetic variation in several species (Taberlet *et al*. 1998; Hewitt 2000; Petit *et al*. 2003; Hofreiter *et al*. 2004). As regards the European wild boar, however, at least four additional processes related to human activities have occurred ever since: the domestication in the Neolithic; a severe bottleneck in different areas in the last few centuries; a demographic expansion in the last 50 years; several more or less uncontrolled introductions of individuals, which also occurred in the last decades, to restock areas where wild boar was extinct or present at low density.

Postglaciation dispersal

The model initially proposed for the population dynamics of several species during the last glaciation (i.e. one or more southern refugia and postglacial re-expansion towards northern areas) is probably too simplistic for some taxa (see for example Magri *et al*. 2006). However, the current distribution of the wild boar and its dispersal and reproductive potential suggest that the genetic variation in this species should be initially investigated with this model in mind. In fact, as the wild boar is being only sporadically observed in northern areas like central and northern taiga (Briedermann 1990; Danilkin 2001), its presence in Central and Northern Europe during the last Ice Age, when the permafrost almost isolated Iberia, Italy, and the Balkans (Hewitt 2000), seems unlikely. In addition, unlike for example other ungulate species, the wild boar shows the typical attributes of r-strategists: high ecological adaptability, opportunistic feeding, and very high reproductive potential (Boitani *et al*. 1995; Fernandez-Llario & Mateos-Quesada 1998; Geisser & Reyer 2005). We can therefore suppose that after the last glaciation, the wild boar easily recolonized Central and Northern European forests, thus reaching an almost continuous and stable distribution modified only by seasonal variations (Jedrzejewska *et al*. 1997; Bieber & Ruf 2005).

Domestication and hybridization with pigs

Direct consequences on the wild populations during the domestication process are expected to be limited, but recent effects related to the co-existence of domestic and wild forms should be considered. In some areas, pigs are reared in semi-wild conditions (e.g. in Bulgaria and Sardinia, Apollonio *et al*. 1988; Genov *et al*. 1991) and crossbreeding with the wild form is possible. Furthermore, in other regions (e.g. Central Italy), pigs were occasionally crossed with wild boars in captivity, and hybrids were released for hunting purposes (Randi *et al*. 1989; I. Boschi, unpublished report). Therefore, divergent pig genomes, which were subjected to strong selective and likely drift effects during and after the domestication, could have introgressed into the wild boar genetic variation.

Overhunting and demographic decline

In the last few centuries, loss of habitat and overhunting drove the wild boar to extinction in some European regions such as the British Isles, Scandinavia, and several Italian and Western Russian areas (Apollonio *et al*. 1988; Oliver *et al*. 1993). A demographic decline was documented in many other countries, and yet the genetic effects of this event, when analysed in a few geographically restricted areas, were surprisingly not evident (Vernesi *et al*. 2003).

Recent expansion

After the Second World War, the density and geographical distribution of the wild boar have increased almost everywhere in Europe (Sáez-Royuela & Tellería 1986; Feichtner 1998; Danilkin 2001) as a consequence of several factors whose relative weight is uncertain. These factors are: global warming, changes in agricultural practices,

setting up of artificial feeding sites, reduced numbers of predators, increase of mast seeding of beech, and restocking (Bieber & Ruf 2005; Geisser & Reyer 2005). Growth rates of wild boar populations have been so high in some areas that damages to agricultural cultivations and natural ecosystems are frequently reported (Singer *et al*. 1984; Welander 2000; Schley & Roper 2003; Geisser & Reyer 2004).

Translocations

The genetic variation in some wild boar populations has possibly been affected also by artificial long-distance migrations associated to uncontrolled and rarely documented restocking plans over the last 50 years. Restocking could have modified the genetic variation both by contributing to the recent demographic expansion and by mixing genetic pools belonging to different subspecies or differentiated populations. For example, wild boars from Central Europe were repeatedly introduced into Italy (Apollonio *et al*. 1988). The genetic impact of such events in some areas is controversial, with authors suggesting either massive (Randi 2005) or limited (Vernesi *et al*. 2003) introgression. It is noteworthy that two opposite wild boar management policies, none of which considering the conservation of genetic biodiversity as a priority, are often suggested by local authorities: eradication, which is meant to remove the problems from cultivated areas, and restocking, in order to preserve traditional forms of 'social' hunting practices.

All the five processes above have potentially left a signature in the present-day wild boar genetic variation. We used mitochondrial and nuclear markers to test whether they did or not, and which was the possible role of each of them. Considering the wild boar as a model, the results we obtained are also valuable for a better reconstruction of the historical events that affected other species. Therefore, our results have important and more general implications for the development of management and conservation plans of game species.

Finally, we also addressed a topic which is not directly related to the wild boar genetic structure, and yet it is relevant to the understanding of the origin and the number of independent domestication events in this species. Archaeological evidences suggest that, like many other domestic animals, European pigs were domesticated in the Near East and selected breeds were subsequently introduced into Europe by Neolithic farmers (Epstein & Bichard 1984). However, two *in loco* domestication processes, one in Central Europe and the other in Italy, were recently hypothesized on the basis of the analysis of mitochondrial DNA sequences (Larson *et al*. 2005). This hypothesis, which is relevant also for the management of local breeds and the conservation of pig diversity, was tested using the same large data set.

Materials and methods

Sampling and DNA isolation

Hair, skin, skeletal muscle, or ear tissue from 252 wild boars were collected in 15 different sampling areas across Europe (Fig. 1) and stored in 95% ethanol at –20 °C. Sample abbreviations used later in the text are specified in Fig. 1. Additionally, 67 Italian domestic pigs from five historical breeds were sampled for comparison: Cinta Senese (coded Pig 1), Sarda (Pig 2), Calabrese (Pig 3), Casertana (Pig 4) and Nera Siciliana (Pig 5). Five commercial pigs (Large White, Pig 6) were also sampled. All the wild boar sampling areas in Italy, possibly with the exclusion of the Castel Porziano Presidential Reserve, and certainly at different degrees, had been subjected to occasional restocking with unknown genetic effects. The samples from four localities (IFlo, IMrp, ICpr and HDif) had already been used in a previous study (Vernesi *et al*. 2003). Total genomic DNA was extracted by using commercial kits (QIAGEN) or the standard phenol–chloroform method (Sambrook *et al*. 1989), followed by concentration in Microcon-30 columns (Amicon), and kept at –20 °C.

Mitochondrial DNA sequencing

Almost the entire control region (CR) was amplified by polymerase chain reaction (PCR) using two primers developed by Alves *et al*. (2003) (Ss.L-Dloop 5′-CGCCATCA-GCACCCAAAGCT3′ and Ss.Hext-Dloop 5′-ATTTTGGGA-GGTTATTGTGTTGTA3′) anchoring at positions 16569 and 1128 of the complete pig mitochondrial DNA (mtDNA) genome (GenBank Accession no. AF034253; Lin *et al*. 1999). Reactions were performed in an Applied Biosystems 2420 thermal cycler, with amplification conditions set at 35 cycles of 92 °C for 1 min, 62 °C for 1 min and 72 °C for 1 min, followed by a final extension step at 72 °C for 10 min. PCR products were purified by Exo/SAP digestion and a 411-bp fragment, including the hypervariable extended termination associated sequences (ETAS) domain (Sbisà *et al*. 1997), was directly sequenced using the forward primer SS.L-Dloop and the BigDye Terminator kit version 3.1 (Applied Biosystems). This region was also selected to maximize the size of possible alignments including already published GenBank sequences. Fragments were finally purified in columns loaded with Sephadex G-50 and run in an ABI PRISM 3100 Avant automatic sequencer (Applied Biosystems). Ambiguous positions were verified by re-sequencing the target region with the internal reverse primer Ss.Hint-Dloop (5′-TGGGCGATTTTAGGTGAGATGGT3′), mapping at position 465 of the pig mtDNA. Sequences were obtained for a subsample of 145 wild boars (between 8 and 12 per sampling location) and 47 domestic pigs from the Italian historical breeds (Pig 1 to Pig 5). Commercial

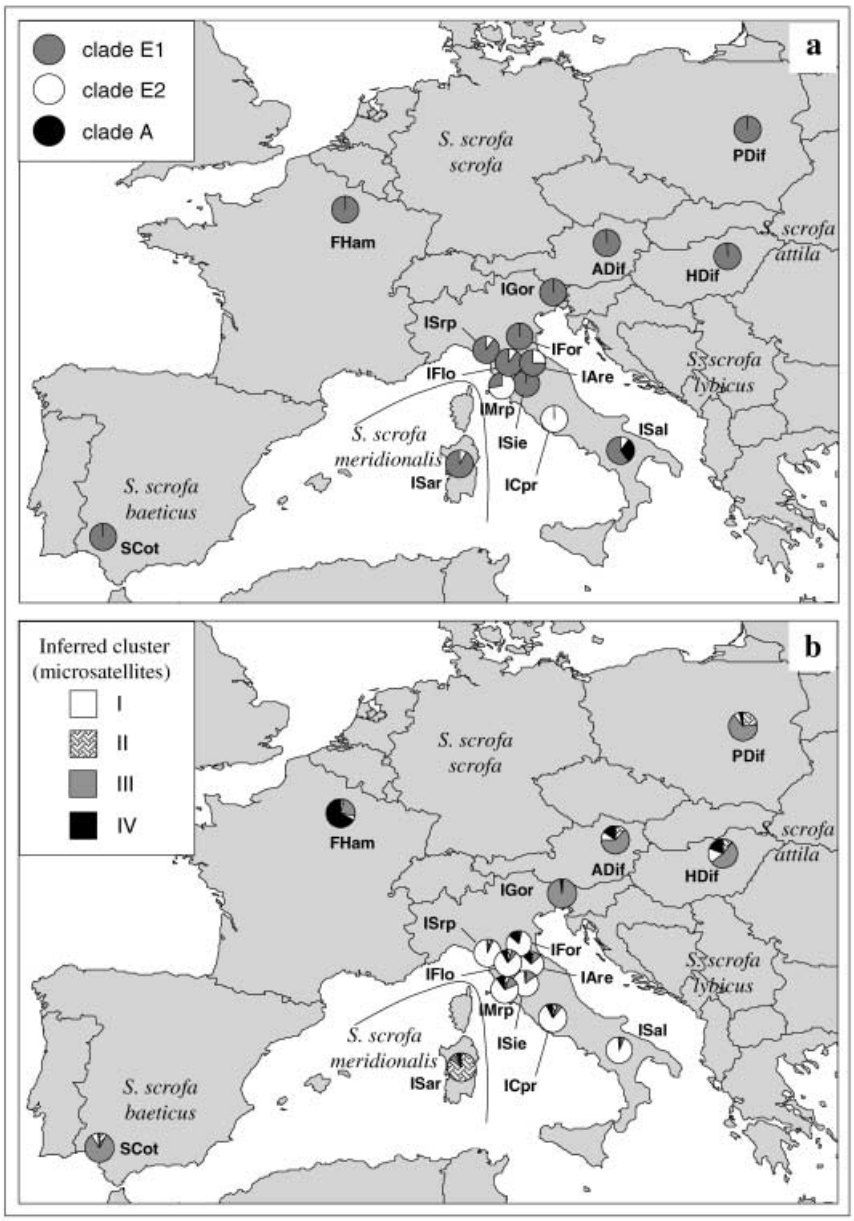

Fig. 1 Geographical locations of wild boar samples. In (a), pie charts represent proportions of each of the three main mtDNA clades (E1, E2 and A) in each sampling area. In (b), pie charts indicate proportions of membership of each sampled population to the four clusters inferred by structure analysis 2 (see text for details). The following abbreviations are used: IAre (Arezzo, Italy), IFor (Forì, Italy), ISie (Siena, Italy), ISal (Salerno, Italy), IFlo (Florence, Italy), IMrp (Maremma Regional Park, Italy), ISrp (San Rossore Regional Park, Italy), ICpr (Castel Porziano Presidential Reserve, Italy), ISar (Sardinia, Italy), IGor (Gorizia, Italy), SCot (El Coto, Spain), FHam (Haute Marne, France), ADif (Austria, different areas), PDif (Poland, different areas), HDif (Hungary, different areas). Subspecies indications on the map are after Groves (1981) and Apollonio *et al*. (1988).

pigs (Pig 6) were not analysed at this marker because of the large amount of data already available in GenBank.

Microsatellite genotyping

A panel of 10 polymorphic microsatellites was selected for the analysis: S026, S215, S355, SW72, SW461, SW857, SW1492, SW2021, SW2496, and SW2532 (details at www.thearkdb.org). Five of them (S026, S215, S355, SW72, and SW857) are in the set recommended by the Food and Agriculture Organization (FAO) to analyse pig diversity (Barker *et al*. 1998), and the rest had been successfully used to study genetic relationships among some European wild boar populations (Vernesi *et al*. 2003). This set was used to genotype all the sampled wild boars and 40 pigs which were raised in wild boar sampling areas (Tuscany and Sardinia). Each PCR was performed in a 10-μL reaction volume, containing 3 μL of DNA solution, 0.5 U of *Taq* DNA polymerase (Euroclone), 1× PCR buffer (Euroclone), 2.5 mM $MgCl_2$, 100 μM of each dNTP and 2 pmol of each primer. The forward primer of each pair was labelled with an ABI fluorescent dye (6-FAM, HEX or TET). The amplification profile was set up with an initial step of denaturation at 95 °C for 3 min, followed by 35 cycles of 92 °C for 45 s, T_a (54–60 °C) for 45 s, and 72 °C for 30 s. A further extension step of 72 °C for 10 min concluded the reaction.

PCR-amplified microsatellite alleles were sized using capillary electrophoresis ABI PRISM automatic sequencers

and internal ROX-500 size standard (Applied Biosystems). The GENEMAPPER software (Applied Biosystems) was used to analyse electrophoretic data.

Genotypes were obtained for all the 252 sampled wild boars, for a subsample of two historical Italian breeds (Pig 1 and Pig 2), and for the Italian commercial pigs (Pig 6).

Mitochondrial DNA data analysis

A total of 192 novel mtDNA CR sequences (411 bp) were obtained (GenBank Accession nos: EU362409–EU362600) and aligned with 612 sequences available in GenBank (Ursing & Arnason 1998; Lin *et al*. 1999; Giuffra *et al*. 2000; Kijas & Andersson 2001; Okumura *et al*. 2001; Kim *et al*. 2002; Randi *et al*. 2002; Alves *et al*. 2003; Yang *et al*. 2003; Gongora *et al*. 2004; Larson *et al*. 2005; Fang & Andersson 2006; Fang *et al*. 2006) using the function CLUSTAL W (Thompson *et al*. 1994) implemented in MEGA 3.0 (Nei & Kumar 2000) and adjusted by eye. Overall, the aligned sequences corresponded to 1065 individual pigs from all over the world (including 231 European wild boars). Haplotypes were collapsed from the entire data set using COLLAPSE version 1.2 (D. Posada, available at http://darwin.uvigo.es), setting deletions as fifth state.

Number of different haplotypes, haplotype (h) and nucleotide (π) diversity, and the mean number of pairwise nucleotide differences between haplotypes (k) were computed using the software ARLEQUIN version 3.01 (Excoffier *et al*. 2005). Allelic richness [$AR_{(r)}$] for each sampled population was calculated from haplotype frequencies using the rarefaction method proposed by El Mousadik & Petit (1996) with the software RAREFAC (R. Petit, www.pierroton.inra.fr/genetics/labo/Software/Rarefac/index.html). The rarefaction size r was set to the smallest sample size among the groups included in the analysis.

A median-joining (MJ) network of haplotypes (Bandelt *et al*. 1999) was created with the software NETWORK 4.1.0.9 (Fluxus Technology), using equal weights for all the mutations and setting the parameter ε to zero, in order to restrict the choice of feasible links in the final network. This approach is especially useful in reconstructing genealogies among closely related taxa, for example, for interpopulation analysis (Bandelt *et al*. 1999). Distributions of pairwise nucleotide differences between haplotypes (mismatch distributions), which are informative on the recent demographic history of a population (Slatkin & Hudson 1991; Rogers & Harpending 1992), were analysed in wild boar and pig populations according to the sudden expansion model as implemented by ARLEQUIN. The age of the expansion was estimated using a generalized nonlinear least-square method, which is based on the minimization of the sum of squared deviations between the observed and the expected mismatch distributions (Schneider & Excoffier 1999). Confidence intervals are obtained using a parametric bootstrap approach based on 1000 simulated samples (Schneider & Excoffier 1999). Finally, Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) were used to test the null hypothesis of demographic stability, under the common assumption (Avise 1995) that the mtDNA control region can be used as a marker of demographic processes, even though natural selection (mainly at linked regions) cannot be excluded (Ballard & Kreitman 1995). The significance of these statistics was evaluated with 1000 simulated samples as implemented in ARLEQUIN.

Microsatellite data analysis

In order to evaluate the levels of genetic variability in the sampled populations, observed and expected (unbiased gene diversity, Nei 1987) heterozygosities were computed with GENETIX version 4.05 (Belkhir *et al*. 2001). In addition, allelic richness and private allelic richness were calculated per population using HP-RARE (Kalinowski 2005), setting rarefaction according to the smallest sample size in each computation. The possible effect of demographic bottlenecks on the microsatellite variation was tested using the method implemented in the software BOTTLENECK version 1.2 (www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html) (Cornuet & Luikart 1996). The test was performed under three alternative models of microsatellite evolution: the infinite allele model (IAM), the stepwise-mutation model (SMM) and the two-phased model (TPM, set with 10% IAM and 90% SMM). Following the authors' suggestions, the Wilcoxon test was used to test the overall differences between the expected heterozygosity and the heterozygosity predicted from the number of alleles.

Deviations from Hardy–Weinberg equilibrium (HWE) were tested for each population and each locus using the Markov chain method proposed by Guo & Thompson (1992), implemented in the software GENEPOP version 3.4 (Raymond & Rousset 1995). Parameters of the Markov chain expressed as dememorizations/batches/iterations were 10 000/100/5000. The significance level was modified for multiple testing across populations and across loci using the sequential Bonferroni correction (Holm 1979). Deviation from linkage equilibrium (LE) was tested for each pair of loci in each population (810 tests in total) using the log-likelihood ratio approach as implemented in the software FSTAT (Goudet 2001) and a sequential Bonferroni correction (Holm 1979). The minimum significance level for the sequential Bonferroni correction should be set to 0.05/810, when the tests are independent, and to higher values when they are not. To avoid excessive rejections of the linkage disequilibrium hypotheses due to nonindependent LE tests, we set the threshold at 0.05/180, that is, we considered the number of loci times the number of populations, as the effective number of independent tests. This choice did not affect our conclusions, since LE was the rule for our microsatellite

markers. We used MICRO-CHECKER version 2.2.3 (Van Oosterhout *et al*. 2004) to detect signs of the possible occurrence of null alleles, that is, homozygote excess evenly distributed among homozygote size classes at one locus.

In order to evaluate levels of genetic heterogeneity among sampling areas, Weir and Cockerham's estimator of F_{ST} (Weir & Cockerham 1984) was computed using the program GENETIX. Significant deviations from zero were tested over 1000 permutations. Molecular distances between alleles and corresponding indices such as R_{ST} (Slatkin 1995) were not used for the analysis of microsatellites to avoid unpredictable results due to probable multiple-step mutations. F_{ST} values were also used to test the relationship between genetic and geographical distances, using the Mantel test (Mantel 1967) as implemented in ARLEQUIN version 3.01 (Excoffier *et al*. 2005).

Pairwise genetic distances between sampling areas were calculated by the program POPULATIONS version 1.2.28 (Langella 2002). A neighbour-joining (NJ) tree of populations, based on the D_A distance (Nei *et al*. 1983), was used to represent the relationships among groups. A consensus tree was obtained by bootstrapping (1000 replicates) distance values over loci.

A Bayesian cluster analysis was carried out using the method implemented in STRUCTURE version 2.1 (Pritchard *et al*. 2000). We first explored which value of K (number of clusters) maximized the likelihood of the data $[P(D \mid K)]$. Simulations were performed by replicating 10 runs for each value of K comprised between 1 and 20, with the following settings: admixture model (initial $\alpha = 1.0$), no population information, correlated allele frequencies, burn-in length: 20 000, Markov chain Monte Carlo (MCMC) length: 1 000 000. Selected burn-in and MCMC lengths allowed the convergence of the chain. All other parameters were set at their default values. The results were then used to evaluate the most likely partition of our data set, adopting the method proposed by Evanno *et al*. (2005), which relies on the second order rate of change of the likelihood function with respect to K. Once defined the most reliable value of K, the genetic contribution of each inferred cluster to the predefined populations as well as to each individual was investigated.

The Bayesian analysis was also used to study the behaviour of each predefined population when the data set was split into a variable number of clusters, starting from $K = 2$ up to the most reliable value of K. For each K, the run providing the highest value of $\ln [P(D)]$ was used. In this descriptive analysis, original groups were assigned to different clades when their composition can be unquestionably assigned to one of the inferred groups. Clearly, this analysis as well as all the tree-based representations implies that early splits (i.e. splits obtained at small K values) can be used to identify the most relevant partitions.

Following the approach introduced by Sacks *et al*. (2004), we also considered the partition which better subdivided the wild boar populations by virtue of their geographical locations. The value of the 'geographical index' (i.e. the average geographical distance between individual locations within clusters divided by the average pairwise distance irrespective of clusters, Sacks *et al*. 2004), which can be computed for each K, is expected to be close to 1 when genetic clusters do not correspond to geographical groups, and lower when each of the K inferred groups includes adjacent populations. When the geographical index stops to decrease for increasing values of K, the geographically meaningful number of groups is reached (Sacks *et al*. 2004).

Results

Mitochondrial DNA variation

A total of 192 mitochondrial CR sequences (411 nucleotides) were analysed: 98 from Italian wild boars, 47 from wild boars sampled in five other European countries, from Spain to Poland, and 47 from five Italian pig breeds. In total, 26 haplotypes (14 of which had never been detected before) and 31 segregating sites (28 substitutions and 3 indels) were identified (see Table S1, Supplementary materials). Different estimates of mitochondrial variability in each population are shown in Table 1. Sample sizes are quite small, and a large variation is expected, and actually observed, across populations at a single locus. However, once we pool the samples into three major groups — wild boars from Italy, wild boars from Europe excluding Italy, and pigs — a clear pattern emerges: genetic diversity in Italy is similar to, or larger than variation observed when several European countries are jointly considered, while pigs show only slightly lower levels of diversity (see Table 1). Sixteen out of 20 haplotypes observed in wild boars are detected in Italy and only seven in the rest of Europe. Of the 28 segregating sites observed in the wild populations, just one is monomorphic in Italy, and 22 in Europe when Italy is not considered. The expectations of these two measures of variation are affected by the sample size, which is clearly larger in the Italian group. But the pattern does not change much when allelic richness, haplotype and nucleotide diversity (i.e. statistics whose expectations are not affected by sample size) are considered (see Table 1).

The high diversity observed in Italian wild boars is due to private mutations and haplotypes, which are mainly related to the exclusive presence in Italy of the three major *Sus scrofa* mtDNA lineages: E1, E2, and A (see Fig. 1a and 2). The codes for these clades were introduced by Giuffra *et al*. (2000) and correspond to clades D1, D4 and D2, respectively, in Larson *et al*. (2005). Clades E1 and A are widely distributed, respectively, in Europe and eastern Asia, while E2 is a second European mtDNA lineage, separated from E1 by five fixed transitions. E2 haplotypes are not observed outside Italy.

Table 1 Genetic variability observed in mtDNA control region sequences (411 bp) in European wild boar populations and in five Italian pig breeds. Values in round brackets correspond to analyses performed excluding the Asian H15 haplotypes (see text). Rarefaction sizes used to compute the allelic richness for each wild boar population and for major groupings are reported in square brackets next to the estimated value

Population	No. of sequences	No. of different haplotypes	Allelic richness	No. of polymorphic sites	Haplotype diversity (h)	Nucleotide diversity (π)
IAre	8	4	3.00 [8]	12	0.821	0.013
IFor	10	3	1.98 [8]	5	0.711	0.007
ISie	8	1	0.00 [8]	0	0.000	0.000
ISal	10	6 (5)	4.36 [8]	20 (14)	0.889	0.021
			(4.00[7])		(0.905)	(0.013)
IFlo	10	4	2.78 [8]	11	0.778	0.008
IMrp	11	2	0.99 [8]	9	0.436	0.010
ISrp	10	2	0.80 [8]	9	0.200	0.004
ICpr	10	3	1.80 [8]	2	0.600	0.002
ISar	12	7	4.31 [8]	15	0.864	0.007
IGor	9	1	0.00 [8]	0	0.000	0.000
SCot	9	1	0.00 [8]	0	0.000	0.000
FHam	10	2	0.80 [8]	1	0.200	0.000
ADif	10	4	2.60 [8]	5	0.711	0.006
PDif	8	2	1.00 [8]	2	0.571	0.003
HDif	10	2	0.98 [8]	2	0.356	0.002
WB Italy	98	16 (15)	10.94 [47]	27 (21)	0.874	0.013
			(10.21 [7])		(0.866)	(0.012)
WB Europe (non-Italy)	47	7	6.00 [47]	6	0.829	0.005
WB overall	145	20 (19)	13.08 [47]	28 (22)	0.902	0.011
			(12.47 [47])		(0.898)	(0.010)
Pig 1	10	2	0.78 [5]	2	0.356	0.002
Pig 2	5	3	2.00 [5]	4	0.800	0.005
Pig 3	6	2	0.83 [5]	2	0.333	0.002
Pig 4	11	2	0.88 [5]	1	0.436	0.001
Pig 5	15	5	1.90 [5]	5	0.705	0.005
DP overall	47	10	9.00 [47]	8	0.827	0.005

WB, wild boar; DP, domestic pig.

The sequences in our data set were then aligned with 612 *S. scrofa* sequences available in GenBank and corresponding to European wild boars and pig breeds. In the overall data set, composed of 804 sequences, a total of 114 different haplotypes (coded as H1–H114) were identified. A complete list of sequences and corresponding haplotypes used in the present study is reported in Table S2, Supplementary material. Of the 70 haplotypes observed in European *S. scrofa* (n = 646 individuals), those with the highest frequency and the widest geographical distribution are H22 (137 individuals), H23 (64) and H29 (42), all of which belong to clade E1 and are shared by wild and domestic individuals (see the network in Fig. 2, singletons are not represented). Of the haplotypes classified in the clade E2, H11 (18) and H12 (13) are the most frequent and are both found in four sampling localities in continental Italy. The two most frequent haplotypes of clades E1 and E2 had been already found in a few Italian museum specimens (dating late 18th–early 19th century) analysed by Larson *et al*. (2005). The Asian clade (clade A) is present in European pig breeds and also in a few wild individuals. In Italy, this clade is observed only in three wild boars from Southern Italy (ISal), all sharing the same haplotype H15. This sequence is not uncommon in European pig breeds (see Fig. 2), and was probably transferred to the wild through hybridization events deliberately induced between domestic breeds and Italian wild boar stocks reared in captivity. The haplotype H15 was excluded from further analyses on account of its belonging to a divergent and exotic clade. The genetic variation observed in ISal and overall in Italy when H15 is excluded is also reported in Table 1: the previously described general pattern does not change. None of the pig samples, including those from the five Italian breeds analysed in this study, shows E2 haplotypes (see Table S1).

The mismatch analysis was initially performed on the two groups of samples which were most clearly differentiated at mtDNA sequences according to the presence/absence of clade E2: Italy and Europe excluding Italy. The distributions

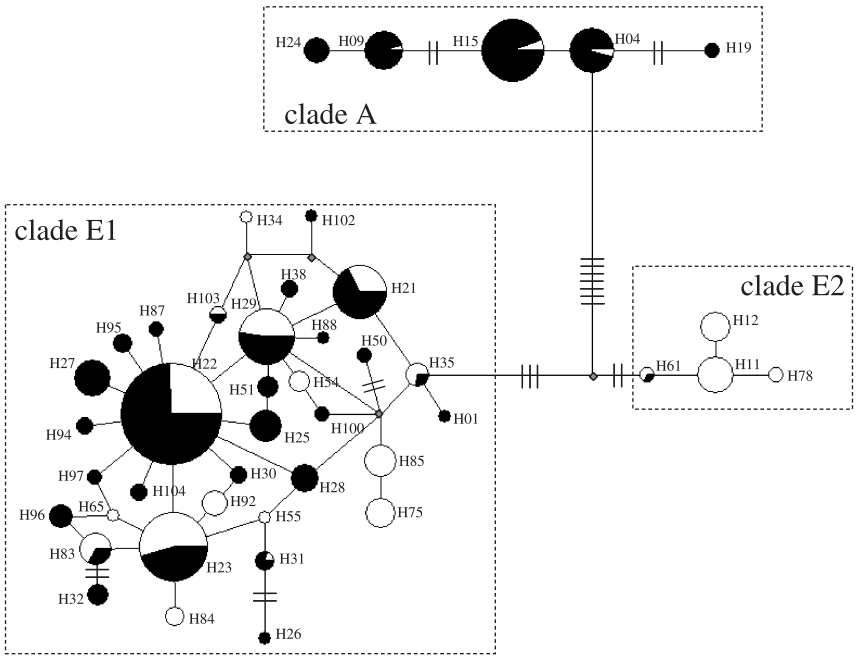

Fig. 2 Median-joining network (Bandelt *et al*. 1999) based on the joint mtDNA data set (original and published sequences). Only haplotypes with a frequency ≥ 2 in the data set were included. Circles indicate sequences observed in wild boar (white) or domestic pig (black) individuals. Size of circles is proportional to the haplotype frequency. Gray diamonds are median vectors, that is haplotypes not observed in the data. Solid branches connecting circles represent single nucleotide changes; lines fragmented by *i* by dashes indicate *i* inferred changes.

obtained either including or leaving out the GenBank sequences (which were possibly affected by non-random sampling) are almost identical. Therefore, only the results obtained with the larger data set are presented. The mismatch distribution in European wild boars ($n = 109$) is smooth and unimodal (Fig. 3a), as expected in the case of a past demographic expansion. The Fu's neutrality test supports the expansion hypothesis ($F_S = –8.25$, $P < 0.01$), whereas the Tajima's D is negative but not significant ($D = –0.84$, $P = 0.22$). The estimated expansion age is 1.36 mutational units (95% CI = 0.58–1.91). On the contrary, the shape of the distribution is clearly ragged and multimodal when the Italian samples are analysed, and the neutrality tests are not significant (Fig. 3b; $n = 116$; $D = 0.95$, $P = 0.86$; $F_S = –0.004$, $P = 0.56$).

The possibility that the clades E1 and E2 observed in Italy simply correspond to two different populations which recently mixed was considered by performing separate mismatch analyses on these clades. According to the results (Fig. 3c, d), the pattern of E1 variation in Italy is not the same as in Europe, since the mismatch distributions have different shapes and the neutrality tests provide different results ($n = 78$; $D = –0.73$, $P = 0.23$; $F_S = –2.30$, $P = 0.17$). The mismatch distribution of the Italian clade E2 is unimodal (Fig. 3d), but Tajima's D and Fu's F_S are far from significance ($n = 38$; $D = –0.11$, $P = 0.48$; $F_S = –0.88$, $P = 0.26$).

Finally, the mismatch distribution was computed for two groups of pig samples, the first including all the European breeds (still excluding the Asian clade) and the second including the Italian breeds only. As regards the former group (European breeds, Fig. 3e), the shape is very similar, although centred at slightly different pairwise difference values, to the distribution observed in the European wild boar (Fig. 3a). Again, neutrality tests point at a demographic expansion ($n = 305$; $D = –1.54$, $P < 0.05$; $F_S = –26.2$, $P < 0.001$), and the estimated expansion age is 0.85 mutational units (95% CI: 0.38–1.12). As regards the latter group (Italian breeds, Fig. 3f), the shape of the distribution is very different from that observed in the Italian wild boar. It partially resembles the European pig breeds distribution (although unusually flat in the central classes), whereas neutrality tests point at a different pattern of genetic variation and do not support the expansion hypothesis ($n = 47$; $D = 0.06$, $P = 0.57$; $F_S = –2.40$, $P = 0.13$).

Genetic variation at microsatellite loci

Between 7 (S026 and S215) and 22 (SW2496) alleles per locus are found across the 292 genotyped individuals. Mean levels of heterozygosity (Table 2) are moderate, averaging 0.57 (SD = 0.05) in the 15 wild boar populations and 0.62 (SD = 0.11) in three pig breeds. Average levels of both heterozygosity and allelic richness are relatively homogeneous across wild boar populations, ranging between 0.47 and 0.62 and between 2.6 and 3.5, respectively. Genetic variation is similar or slightly higher in the joined Italian samples than in the pooled European group (Table 2). An influence on this pattern of individuals from ISal introgressed with Asian pig genomes can be probably excluded. In fact, most populations show higher levels of genetic variation than ISal (Table 2). As regards single populations, Sardinian

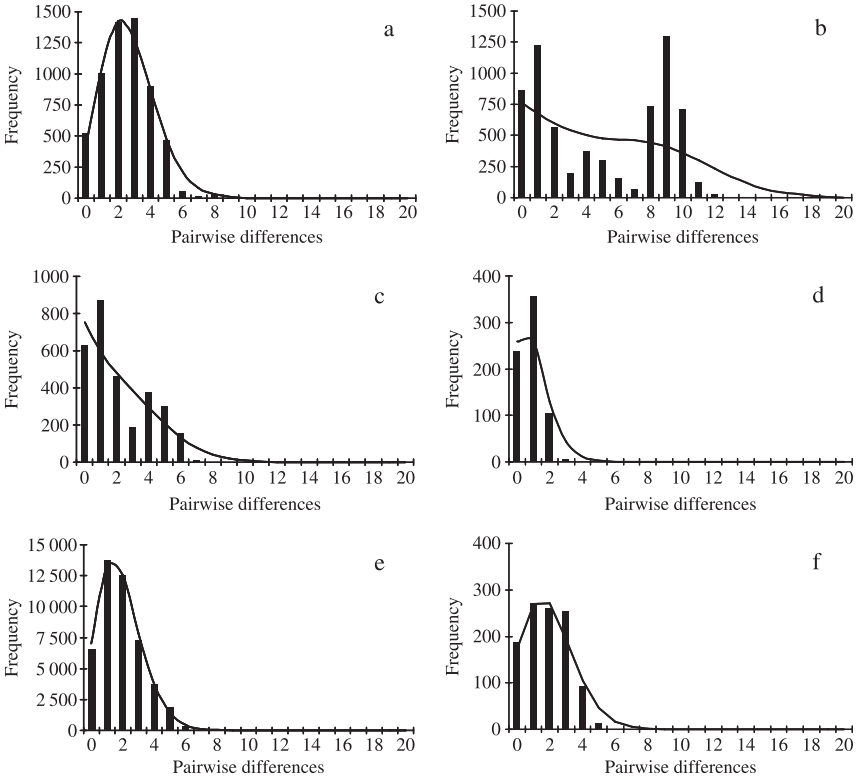

Fig. 3 Mismatch distributions based on pairwise site differences between sequences. (a) European wild boars (excluding Italy); (b) Italian wild boars, all haplotypes; (c) Italian wild boars, clade E1 only; (d) Italian wild boars, clade E2 only; (e) European domestic pigs, clade E1 only; (f) Italian domestic pigs. The expected curve (solid line) was obtained from simulated values computed from the data under the model of demographic expansion. Haplotypes of the Asian clade (A) were excluded from the analysis.

Table 2 Genetic variability detected at microsatellite loci in European wild boar and pig populations. Expected heterozygosity was calculated as gene diversity (Nei 1987). Allelic richness and private allelic richness were calculated with rarefaction set at 14 genes for each population and at 76 genes for groups analyses. Pig 6 was excluded from this analysis for its low sample size

Population	*N*	Average expected heterozygosity	Average observed heterozygosity	Average allelic richness	Average private allelic richness
IAre	10	0.53	0.51	3.84	0.01
IFor	10	0.61	0.59	4.18	0.09
ISie	10	0.62	0.61	3.95	0.10
ISal	10	0.55	0.47	3.56	0.02
IFlo	20	0.62	0.57	4.19	0.13
IMrp	11	0.49	0.45	3.63	0.04
ISrp	15	0.51	0.47	3.04	0.00
ICpr	19	0.54	0.52	3.67	0.01
ISar	41	0.61	0.50	4.27	0.54
IGor	19	0.47	0.47	3.10	0.08
SCot	15	0.60	0.59	3.82	0.26
FHam	20	0.58	0.57	3.88	0.17
ADif	13	0.62	0.55	4.27	0.05
PDif	19	0.61	0.59	3.84	0.07
HDif	20	0.61	0.59	4.45	0.14
WB Italy	165	0.65	0.51	7.96	1.43
WB Europe (non-Italy)	87	0.66	0.58	7.54	1.02
WB overall	252	0.66	0.53	8.44	3.18
Pig 1	22	0.51	0.47	3.31	0.20
Pig 2	13	0.61	0.57	4.26	0.44
Pig 6	5	0.73	0.61	—	—
DP overall	40	0.62	0.52	6.85	1.21

WB, wild boar; DP, domestic pig.

wild boars (ISar) and Sardinian pigs (Pig2) show higher levels of allelic diversity and the highest proportions of private alleles. The lowest values of diversity are found in Northeastern Italy (IGor) and in the enclosed population in San Rossore Regional Park (ISrp).

Averaging across loci, observed heterozygosity is slightly smaller than expected under HWE in all populations. In the locus-by-locus analysis, the number of significant tests after Bonferroni correction is six (out of 180): one in a pig sample (where substructuring, inbreeding, or selection are likely), one in Florence (where translocations are documented, historically and genetically, Vernesi *et al*. 2003), and two each in the probably heterogeneous samples from Poland and Sardinia. Deviations from linkage equilibrium are significant only in 26 pairwise tests (out of 810), concentrated in the pig and the Polish samples. The possible presence of null alleles, tested separately for each locus and each population with MICRO-CHECKER, is limited (13 significant results), randomly distributed across loci, and mainly concentrated again in the Sardinian and the pig samples.

The bottleneck test fails to identify the genetic signature of a demographic decline, no matter the model of microsatellite evolution selected for this analysis. The relative excess of heterozygosity expected in bottlenecked populations (Cornuet & Luikart 1996) is not observed either in the single-locus analysis or in the Wilcoxon test combining the results for the different loci. In contrast, a general deficiency of heterozygosity is found in some populations, but this result is not always consistent across the different mutation models.

Genetic differentiation among populations

The genetic divergence between populations was analysed using the multilocus microsatellite data set. The overall F_{ST} values are relatively high and significant (0.14 and 0.15, excluding and including domestic pigs, respectively; $P < 0.001$ in both cases). Genetic differentiation due to differences between wild and domestic groups is significant but very limited ($F_{CT} = 0.030$, $P = 0.01$). In the wild boar, pairwise F_{ST} values range between 0.00 and 0.31 across Europe (including Italy) and are comparable to those observed between wild boar and domestic pig samples (range 0.10–0.25). Most pairwise values are similar to each other and similar to the global F_{ST}, which is in agreement with a moderately structured tree of populations with short (and moderately supported) internal branches (Fig. 4). No significant correlation was detected between genetic and geographical distances (Mantel test: $r = 0.153$; $P = 0.19$).

When the population structure is analysed with the Bayesian method implemented in the program STRUCTURE, some levels of partitioning in geographically meaningful groups clearly emerge. In the following, we will consider two different partitions which are identified by means of

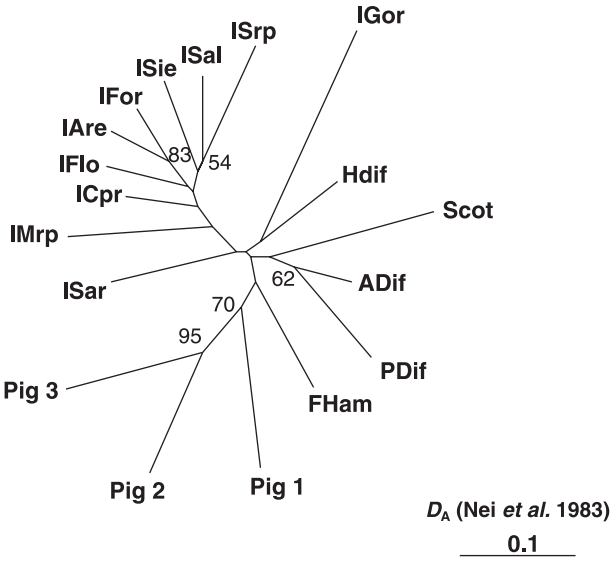

Fig. 4 Neighbour-joining tree based on the Nei's *et al*. (1983) distance (D_A) between populations computed on microsatellite data. Bootstrap support at internodes is shown if > 50%.

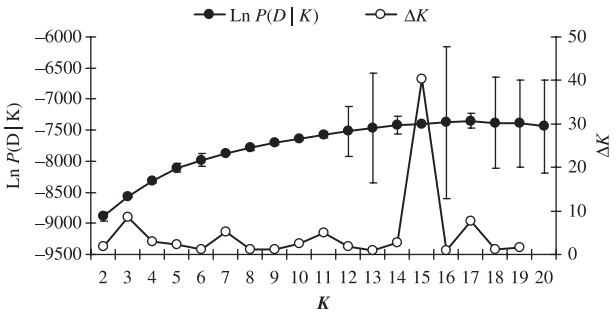

Fig. 5 Posterior probability of the data {ln [$P(D \mid K)$]} and values of ΔK (Evanno *et al*. 2005) as a function of K (number of clusters), as resulting from the simulations in STRUCTURE (structure analysis 1).

the likelihood (structure analysis 1) and the 'geographical index' (structure analysis 2) approaches. The more rigorous likelihood approach identifies 15 clusters (Fig. 5). Accordingly, a detailed analysis of the relationship between inferred clusters and original populations is provided. The 'geographical index' approach identifies four clusters and could thus be regarded as a synthetic representation of the major geographical groups.

Structure analysis 1: likelihood-based partition (Table 3 and Fig. 6)

The large number of clusters and the contribution of each population to each inferred cluster confirm the relatively large genetic divergence between most of the samples. All the populations from Central-Southern Italy, with the exception of IMrp, have major components in the same

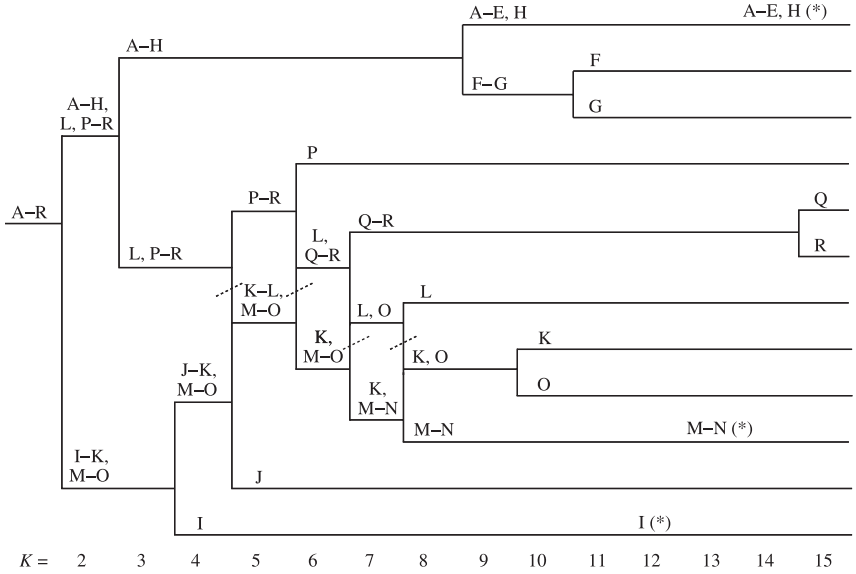

Fig. 6 Diagram showing how populations split into the clusters inferred by the program STRUCTURE as the value of *K* increases. Populations are referred to as: A, IAre; B, IFor; C, ISie; D, ISal; E, IFlo; F, IMrp; G, ISrp; H, ICpr; I, ISar; J, IGor; K, SCot; L, FHam; M, ADif; N, PDif; O, HDif; P, Pig 1; Q, Pig 2; R, Pig 3. Splitting of one population or a group of populations into multiple exclusive clusters is indicated by an asterisk. Reticulations indicate that the attribution of some original groups is not consistent across all *K* values; dashed lines indicate likely solutions, based on clustering at lower *K* values.

Table 3 Partition of the 18 sampled populations into the 15 clusters inferred by the program STRUCTURE (structure analysis 1). This was the most supported value of *K*, obtained by simulated data assuming the admixture model and ignoring population information (10 replicated runs, each with 1 000 000 iterations of data collection after a burn-in of 20 000 iterations). Proportions higher than 0.1 are in bold. Letter codes used in Fig. 6 are reported

			Inferred clusters														
Code	Pop	*N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	IAre	10	0.057	0.020	0.016	0.018	0.058	0.015	0.050	0.019	0.006	**0.370**	0.019	0.020	0.016	0.029	**0.286**
B	IFor	10	0.040	0.009	0.039	0.019	0.028	0.011	0.015	0.075	0.005	**0.328**	0.008	0.019	0.096	0.044	**0.265**
C	ISie	10	0.020	0.007	0.010	0.036	0.050	0.025	0.010	0.082	0.013	**0.331**	0.013	0.007	0.030	0.014	**0.353**
D	ISal	10	0.024	0.013	0.029	0.009	0.012	0.017	0.010	**0.257**	0.019	**0.245**	0.011	0.010	0.060	0.007	**0.276**
E	IFlo	20	0.036	0.012	0.017	0.051	0.010	0.016	0.014	0.048	0.013	**0.333**	0.038	0.056	0.041	0.021	**0.295**
F	IMrp	11	0.016	0.009	0.015	0.007	0.014	0.036	0.014	0.064	0.038	0.057	0.007	0.011	**0.652**	0.009	0.052
G	ISrp	15	0.009	0.006	0.009	0.004	0.007	0.009	0.008	**0.862**	0.006	0.025	0.005	0.007	0.010	0.007	0.027
H	ICpr	19	0.048	0.015	0.012	0.013	0.015	0.080	0.033	0.083	0.030	**0.251**	0.009	0.020	0.043	0.041	**0.308**
I	ISar	41	0.020	**0.529**	0.013	0.010	0.015	**0.232**	0.016	0.021	0.031	0.011	0.034	0.027	0.013	0.018	0.011
J	IGor	19	0.011	0.007	0.005	0.006	0.019	0.009	0.010	0.005	**0.893**	0.005	0.006	0.005	0.006	0.006	0.006
K	SCot	15	0.015	0.014	0.013	0.015	**0.802**	0.008	0.007	0.054	0.008	0.012	0.012	0.010	0.012	0.005	0.013
L	FHam	20	0.020	0.010	**0.702**	0.030	0.029	0.015	0.016	0.010	0.009	0.009	0.011	0.055	0.036	0.037	0.010
M	ADif	13	**0.181**	0.020	0.055	**0.178**	0.040	0.012	0.032	0.016	0.034	0.019	**0.196**	0.035	**0.133**	0.029	0.022
N	PDif	19	**0.115**	0.012	0.020	**0.363**	0.061	0.012	0.008	0.022	0.031	0.010	**0.294**	0.007	0.015	0.019	0.011
O	HDif	20	**0.549**	0.016	0.044	0.010	0.026	0.017	0.030	0.037	0.093	0.031	0.033	0.019	0.029	0.034	0.031
P	Pig 1	22	0.013	0.011	0.011	0.011	0.007	0.006	0.015	0.011	0.012	0.014	0.014	0.063	0.009	**0.790**	0.014
Q	Pig 2	13	0.013	0.009	0.007	0.006	0.005	0.009	**0.597**	0.007	0.009	0.009	0.006	**0.289**	0.007	0.017	0.009
R	Pig 3	5	0.013	0.009	0.021	0.036	0.038	0.012	**0.129**	0.010	0.007	0.019	0.023	**0.643**	0.009	0.013	0.018

three inferred clusters (8, 10 and 15, see Table 3). IMrp has one major component in cluster 13, poorly shared by all the other samples. Sardinian and northeastern Italian samples have almost 'private' components, namely clusters 2 and 6, and cluster 9, respectively. Of the European populations, the Spanish and the French samples correspond to private clusters 5 and 3, respectively, while Hungarian, Polish and Austrian wild boars share three cluster components (1, 4, and 11). No one of the wild boar populations shows a relevant proportion of any of the three clusters associated to domestic pigs, although the analysis conducted on an individual basis reveals some levels of admixture. Provided that the three pig samples have a cumulative proportion of clusters 7, 12, and 14, comprised between 78% and 91%, and that the homologous proportion in wild boar populations averages 6%, we arbitrarily classified single wild boars having a cumulative proportion of these clusters > 18% (three times the 'background' proportion observed in

wild boars across Europe) as hybrids. Indeed, the higher similarity between these individuals and the pigs could be possibly explained by introgression and not by common ancestry. Using this threshold, 7% and 9% of the wild boars in Italy and in Europe excluding Italy, respectively, seem to have genomes affected by pig introgression. In Sardinia, only two specimens (5%) fell into this category, but their pig contribution is very high (> 80%). According to the same approach, about 7% of wild boar individuals collected in Italy seems to be affected by introgression of Eastern European wild boar genomes.

The results of the Bayesian analysis obtained with different numbers of inferred groups are summarized hierarchically in Fig. 5. By progressively increasing the number K of components, from $K = 1$ up to $K = 15$ (the X-axis in the diagram), we analysed the partition of each original group into the K-inferred clusters, thus identifying decreasingly important subgroups of samples. Each branch in the figure represents a group inferred at the corresponding K value, while nodes represent the progressive splits of inferred groups when K increases. Original groups were assigned to the branches of the diagram (i.e. to the inferred groups) when their composition was mostly associated to a single inferred group. The most basal split, for $K = 2$, does not separate the domestic and the wild forms as expected: one group actually includes wild boars from Central-Southern Italy and France as well as domestic pigs, whereas the wild boar populations from Sardinia and Northeastern Italy group with the rest of the European populations. For $K = 3$, Central-Southern Italian wild boar populations are separated from pigs, unlike the French one which still groups with the domestic breeds. The Sardinian population is the first original population to emerge separately with a nonshared inferred group, at $K = 4$. The following four steps are somewhat confused because of the French and the Hungarian populations' ups and downs in the diagram branches. This erratic pattern produces the reticulations between $K = 5$ and $K = 8$, which can be, however, easily resolved by considering the earlier assignments of these two populations (see dashed lines in the diagram). Like in previous analyses, the next splits support the divergence of the Northeastern Italian sample (IGor = 5), the clustering of all Central-Southern Italian populations (subdivided into three components only for $K > 10$), and the grouping of most of the other European populations (which starts to split for $K > 7$). Interestingly, the French population is the last group which separates from the domestic pig (only for $K > 7$).

Structure analysis 2: geographically based partitions (Fig. 1b)

This analysis allows for a map-based representation of a reduced number of genetically homogeneous, geographically related groups (see Fig. 1b). It can also be regarded as a zoom of the diagram in Fig. 6 for $K = 4$. All populations from Central-Southern Italy have major components in cluster I, with a cumulative contribution by other clusters in all cases < 30%. Similarly, the Sardinian samples are assigned to the almost 'private' cluster II. Cluster III corresponds to the major genetic component in the rest of the European populations, with the exception of the French sample where the cluster IV prevails. Cluster IV represents the domestic pigs (90% of individual pigs have an assignment probability to this cluster of ≥ 80%). This result is not to be taken as an evidence of pig introgression in France, since in the more accurate structure analysis 1 (see above), none of the French individuals was identified as hybrid. Instead, it supports the hypothesis of a higher-than-average genetic relationship between domestic pigs and French wild boars (see also Figs 4 and 6). When compared with the distribution of the mtDNA clades (see Fig. 1a), the results of this analysis confirm the remarkable divergence between European and Central-Southern Italian wild boars. They additionally suggest that, notwithstanding the frequency of the typical Italian mtDNA clade (E2), large contributions of exotic gene pools can be excluded in all Italian populations.

Discussion

By the analysis of one mitochondrial and 10 independent nuclear markers in wild boars collected across Europe, we evaluated the possible contribution of different natural and human-related processes in shaping the present genetic diversity of the species in the old continent. The role of past wide-scale events, like range and size fluctuations occurring during the last glacial and postglacial periods, are pointed to as the main force leading to the observed levels of differentiation in Europe. In addition, the comparison of the genetic diversity of wild and domestic pigs is compatible with the idea of a domestication centre in Central Europe, but it does not provide support to the hypothesis of an independent domestication in the Italian peninsula.

Postglacial dispersal

Our comparative analysis of wild boar control region sequences shows the signature of a past demographic expansion, which could well have followed the range contraction occurring in Europe during the last glaciation. Assuming a sudden expansion model, this event can be dated back to approximately 1.36 mutational units in the past. This figure is obviously imprecise, because of stochastic variations in the coalescent and the sampling processes, and its conversion into years appears even more difficult. At least a fivefold range of variation can be found in the per-site/per-year mutation rates which have been used in several studies on the hypervariable mtDNA control region

in mammals. For the analysis of ungulates in particular, Randi *et al*. (1998) suggested a range between 4% and 8% sequence divergence per site per million years, while Birungi & Arctander (2000) suggested an interval between 10% and 20%. Accordingly, if we consider only the errors associated to the calibration, the expansion age of the wild boar should be dated back to between 16 000 BP and 80 000 BP. Several authors argued for a mutation rate closer to the upper limit of these intervals (e.g. Bradley *et al*. 1996 and Kim *et al*. 2002 opted for a 15% rate), and we therefore believe that the molecular data we analysed support the hypothesis of a demographic expansion of the wild boar following the last glaciation. Interestingly, the age of the expansion estimated from the similarly unimodal mismatch distribution in the European domestic pig data set (Fig. 3e) is about 40% younger, that is, 10 000 years ago using a 20% rate. Considering the errors associated to these estimates, also for the reason that the method assumes a single demographic event, these data appear consistent with a pig Neolithic expansion associated to the domestication process, or they might actually correspond to a sort of average that keeps also the signs of the postglaciation expansion of their wild ancestors.

Two additional results support the hypothesis that the pattern of genetic diversity in the European wild boar was shaped by the last postglacial colonization event from one or more southern refugia. First, the divergence between European samples outside Italy is similar in different comparisons, and no correlation is observed between genetic and geographical distances. This finding is not consistent with a stable equilibrium between drift and migration (i.e. under isolation by distance), while it is expected in case of a rapid colonization and a simultaneous demographic expansion from a common source population. Second, the mitochondrial variation in Italy is not compatible with a demographic expansion, and both mtDNA and nuclear markers show a similar or even higher level of variation in Italy than in the rest of the European regions jointly analysed. This result is remarkable at the mtDNA: two major clades, E1 and E2, separated by at least 50 000 years, are observed in Italy, whereas only E1 is found elsewhere. The simplest explanation for this pattern seems to be the process of contraction into southern refugia and the following re-expansion towards northern areas. The recent finding (Larson *et al*. 2007) that E2 haplotypes were present in the present-day Croatia about 11 000 years ago is not unexpected under this view, considering that Northern Adriatic was not submerged during the last glaciation (Van Andel 1989). According to this scenario, Italian diversity represents a large fraction of the wild boar preglaciation diversity, which was then preserved without any major impact by subsequent demographic processes. Different hypotheses about the location of source refugia appear equally likely: it could be Italy itself, with the loss of rare E2 haplotypes during the colonization, or some other southern areas in Europe, or a combination of these two. Actually, the Italian peninsula seems to have played for several species a minor role in the recolonization processes than the Iberian and Balkan peninsulas, probably on account of the Alps being a greater physical barrier to the dispersal of individuals (Hewitt 2000). A wider sampling in Iberia and in the Balkans will be necessary to identify which refugium area contributed most to the present gene pool of the European wild boar. Similarly, more locations throughout Europe should be screened for the presence of E2 haplotypes, as to exclude that they occur outside Italy at low frequency.

Domestication and hybridization with domestic pigs

Wild boars and domestic pigs share the most common mtDNA haplotypes and microsatellite alleles, and the population tree (Fig. 4) only weakly resolves the two groups. Also the Bayesian clustering analysis confirms that wild and domestic forms are not more divergent than other pairs of wild boar populations. It seems therefore that the differentiation from the wild boar during the domestication, remarkable at morphological traits and thus probably also at their genetic determinants, was not accompanied by a strong founder effect. As already suggested in relation to other domestication events which implied a long co-existence between the domestic and the wild forms (e.g. dog and cattle), occasional and/or deliberate hybridizations could have played a role in reducing their genetic divergence (Vilà *et al*. 2005; Beja-Pereira *et al*. 2006). A fraction between 5% and 10% of the wild boar individuals we analysed shows the effects of pig introgression, but the global contribution of pig genomes in the wild populations is clearly lower (as supported by the population analysis). Remarkably, we even found wild boar individuals in Southern Italy with Asian pig mtDNA, usually observed in some ameliorated European breeds crossbred with Asian pigs (Giuffra *et al*. 2000). This evidence is consistent with the observation by Fang *et al*. (2006), who assayed an approximately 10-fold lower frequency of Asian haplotypes in wild boar populations than in domestic breeds in Europe (3% vs. 30%). This result supports the view that some levels of hybridization between wild boars and domestic pigs occurred in the past and possibly still occur today.

Domestication: the origin of European breeds

The similarity between European pig breeds and wild boars, when considered in connection with the finding that Middle Eastern mtDNA lineages are not observed in European pig breeds (Larson *et al*. 2005), unequivocally suggests that modern breeds in Europe descend from local wild populations. Additionally, the population tree and the Bayesian inference of population clustering based on

nuclear markers indicate that the wild boar population to be most closely related to the pig is located in northern France. This result is compatible with the hypothesis of Larson *et al*. (2005, 2007) based on modern and ancient mtDNA sequences, whereby Central Europe was an important domestication centre. A second domestication centre for the European breeds was actually proposed by Larson *et al*. (2005) to be located in Southern Europe, and more specifically in mainland Italy. This hypothesis, although intriguing, was supported only by two museum Sardinian specimens with possible feral origin showing E2 mtDNA sequences. Similarly, Larson *et al*. (2007) found two Bronze Age Sardinian individuals morphologically classified as 'domestic/feral' showing E2 sequences, although all ancient Italian samples firmly attributed to domestic pigs had European E1 haplotypes. In our analysis at both mtDNA and nuclear markers, we do not find any specific relationship between Italian pigs and Italian wild boars, and no E2 haplotype is found in the 47 pigs belonging to the five local breeds we analysed. Although the pattern of genetic variation in the Italian breeds suggests a different demographic history from that of other European breeds (no signs of demographic expansion are detected), our data seem to exclude an independent domestication event in Italy, or, at least, that pigs possibly domesticated in Italy have left descendants in modern breeds.

Overhunting and demographic decline

The genetic impact of the demographic decline that affected several wild boar populations in Europe during the last two centuries seems to be extremely limited, if any.

In general, the level of genetic variations at the mtDNA sequences in the whole data set is within the range observed in other ungulate populations (Vernesi *et al*. 2002; Feulner *et al*. 2004; Randi *et al*. 2004). The average microsatellite heterozygosity (0.66) is only slightly lower than the average value observed in 14 non-endangered species (0.70) and much higher than the average value of 0.41 found across 14 threatened taxa (Frankham *et al*. 2004). The genetic variation, as expected in nonpanmictic species, is lower within single localities, with large differences between samples especially at the mtDNA region. This is probably a consequence of the increase of drift effects due to the smaller effective population sizes at this marker, and possibly this is also related to our small sample sizes. We note however that the levels of variation are relatively high almost everywhere, also within samples, with only two restricted groups in as many regional parks (ISrp and IMrp) and one sample with an acknowledged history of recent recolonization (IGor), having very low variation both at mitochondrial and microsatellite loci.

More specifically, the mismatch analysis and two neutrality tests applied to mtDNA sequences are all indicative of either stability or expansion, and a specific method developed to identify a bottleneck using microsatellite markers does not reveal the deviation expected in case of demographic decline. In other words, both the levels and the patterns of diversity observed at mtDNA sequences and microsatellites are consistent with the idea that the population size and distribution range contractions did not affect the overall genetic variability.

Recent expansions and translocations

The possibility to detect the genetic effects of the rapid growth occurred in the last 50 years in several regions is questionable. Provided that a few generations of expansion are not sufficient to accumulate enough mutations in the gene genealogy, to capture statistically the reduced drift effects which are typical of this process is very difficult. However, it is interesting to note that a first attempt to identify these effects using an approximate Bayesian computation (ABC) approach (Beaumont *et al*. 2002), which can be used to model complex demographic processes, suggests that a large growth rate occurred in the last 10–20 generations and was larger in Central Italy than anywhere else in Europe (results not shown). Therefore a tentative conclusion would be that Italy still preserves the preglaciation diversity, which was not severely affected by the decline occurred from the Middle Ages until the end of World War II, and was recently frozen by the modern expansion.

As for the translocation events, it was suggested that poorly documented and usually uncontrolled restocking plans drastically affected genetic variation in Central and Southern Italy, and also speeded up the recent expansion process (Randi 2005; Apollonio *et al*. in press). Since translocations were carried out using also animals from Central Europe, this might have artificially contributed to the higher genetic variation observed in Italy. However, in agreement with Vernesi *et al*. (2003), we can exclude a major impact of such human-mediated migrations. If this phenomenon were extensive, we would have observed (i) the clustering of some Central and Southern Italian samples with some Central European groups; (ii) the presence of inferred clusters (in the Bayesian analysis) shared by some Central and Southern Italian and Central European groups; (iii) a large proportion of individuals from Central and Southern Italy assigned to other European populations. None of these predictions is met by our genetic analyses, and only a limited fraction (7%) of individuals sampled in Italy have significant proportions of their genomes that can be related to wild boar populations abroad. In addition, the plausible hypothesis that the native boars in Italy had E2 haplotypes, while E1 haplotypes were introduced by recent translocations of wild boars from Central Europe is contradicted by the fact that the variation patterns of E1 sequences in Italy and in the rest of Europe are different. We therefore conclude

that restocking from Central Europe had a limited genetic impact and an accordingly a marginal role in the recovery of the Italian population.

Conclusions and management implications

In conclusion, the most important event in shaping the observed pattern of diversity seems to have been the last glaciation, which was followed by a sudden demographic and spatial expansion from one or more southern refugia. The genetic signature of more recent processes, which were mostly related to human activities, can be detected but it appears marginal. Clearly, a wider sampling of European populations, including several locations in the Iberian Peninsula and in the Balkans, is necessary to better clarify postglacial dynamics. However, our data point to a single area of discontinuity which corresponds to the Alps. Wild boars sampled south of this chain show, indeed, higher levels of overall genetic variation, a private mtDNA haplogroup and endemic diversity at microsatellite loci.

The wild boar is a rather invasive species, with a relevant impact on biodiversity, agriculture and livestock, and it is also of large interest for hunters. Therefore, it is important to consider the implications of our results for the management of this species. We believe that (i) Italian populations represent a reservoir of genetic diversity in Europe which should be preserved; for example, following the arguments in Petit *et al*. (1998), in the eventual case of extinction of the wild boar in Italy, the reduction of nuclear allelic richness would be about twice as much as in the case of the simultaneous extinction in five other European regions (11% compared to 6%), and a high divergent mtDNA clade would be lost; however, since almost all the Italian groups are genetically very similar, and even documented demographic reductions did not affect significantly the genetic variation, hunting is still recommended to reduce the population size in some areas; (ii) animals should not be translocated from one European region to another, especially across the Alps; even though the level of differentiation is probably not enough to maintain the current subdivision into subspecies, some level of local adaptation is expected and should not be compromised by hybridization; accidental escapes from wild boar farms should also be prevented for the same reason; (iii) artificial crossbreeding with domestic pigs should be avoided and genetic controls in wild boar farms should be enforced. A different matter are those situations where the genetic exchange between wild and domestic pigs has a historical background, due to the long-lasting practice to rear pigs in a natural state (e.g. in Sardinia). In these cases, the present genetic identity of the wild population is influenced by the prolonged gene flow between the two forms, and thus, in absence of significant introgressions from other areas, they can be referred to as a joint evolutionary unit.

Acknowledgements

We are deeply grateful to all those who collaborated with us in collecting samples: Henrik Okarma (Poland), Laszlo Sugar (Hungary), Juan Carranza (Spain), Franz Suchentrunk (Austria), Eric Baubet (France), Roberto Mazzoni Della Stella (Siena), Stefano Antonacci (Gorizia), Paolo Varuzza (Salerno), Luca Tonini (Maremma Regional Park), Riccardo Bozzi and several local hunters and pig breeders. The Istituto Nazionale per la Fauna Selvatica (INFS) kindly allowed us to collect samples from the Castel Porziano Presidential Reserve. We also thank Lidia Migliori, Giovanni Manca and David Caramelli for technical assistance, and Umberto Albarella for useful suggestions to an early draft of the manuscript. Three anonymous referees made constructive and valuable comments on this paper. The study was possible thanks to the financial contribution of the Sardinian Regional Government, the University of Ferrara, The Italian Ministry for University Research (MIUR PRIN 2005) the Centre for Alpine Ecology, and the Autonomous Province of Trento. L. Iacolina benefited by a fellowship sponsored by the Sardegna Resorts srl.

References

Alves A, Ovilo C, Rodriguez MC, Siliò L (2003) Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and others domestic and wild pig populations. *Animal Genetics*, **34**, 319–324.

Apollonio M, Randi E, Toso S (1988) The systematics of the wild boar (*Sus scrofa* L.) in Italy. *Bollettino Di Zoologia*, **3**, 213–221.

Apollonio M, Ciuti S, Pedrotti L, Banti P (in press) Ungulate management in Italy in the XXI Century. In: *Ungulate Management in Europe in the XXI Century* (eds Apollonio M, Andersen R, Putman R). Cambridge University Press, Cambridge, UK.

Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, **9**, 686–690.

Avise JC (2004) *Molecular Markers Natural History, and Evolution*, 2nd edn. Sinauer & Associates, Sunderland, Massachusetts.

Ballard JW, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution*, **10**, 485–488.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barker JSF, Hill WG, Bradley D, Nei M, Fries R, Wayne RK (1998) *Measurement of Domestic Animal Diversity (MoDAD): Original Working Group Report*. FAO, Rome.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in population genetics. *Genetics*, **162**, 2025–2035.

Beja-Pereira A, Caramelli D, Lalueza-Fox C *et al*. (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences, USA*, **103**, 8113–8118.

Belkhir K, Borsa P, Chikhi L, Raufauste N, Bonhomme F (2001) *GENETIX 4.02, logiciel sous WindowsTM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions. CNRS UMR 5000, Université de Montpellier II, Montpellier (France). Available from URL: http://www.genetix.univ-montpz.fr/genetix/genetix.htm

Bieber C, Ruf T (2005) Population dynamics in wild boar *Sus scrofa*: ecology, elasticity of growth rate and implications for the management of pulsed resource consumers. *Journal of Applied Ecology*, **42**, 1203–1213.

Birungi J, Arctander P (2000) Large sequence divergence of mitochondrial DNA genotypes of the control region within populations of the African antelope, kob (*Kobus kob*). *Molecular Ecology*, **9**, 1997–2008.

Boitani L, Trapanese P, Mattei L, Nonis D (1995) Demography of a wild boar (*Sus scrofa*, L.) population, Tuscany, Italy. *Gibier Faune Sauvage*, **12**, 109–132.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences, USA*, **93**, 5131–5135.

Briedermann L (1990) *Schwarzwild*. VEB Deutscher Landwirtschaftsverlag, Berlin, Germany.

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

Danilkin AA (2001) The wild boar: an unprecedented spread or restoration of the species range? *Doklady Biological Sciences*, **380**, 457–460.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.

Epstein J, Bichard M (1984) Pig. In: *Evolution of Domesticated Animals* (ed. Mason IL), pp. 145–162. Longman, New York.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1803–1810.

Fang M, Berg F, Ducos A, Andersson L (2006) Mitochondrial haplotypes of European wild boars with $2n = 36$ are closely related to those of European domestic pigs with $2n = 38$. *Animal Genetics*, **37**, 459–464.

Feichtner B (1998) Ursachen der Streckenschwankungen beim Schwarzwild im Saarland. *Zeitschrift für Jagdwissenschaft*, **44**, 140–150.

Fernandez-Llario P, Mateos-Quesada P (1998) Body size and reproductive parameters in the wild boar *Sus scrofa*. *Acta Theriologica*, **43**, 439–444.

Feulner PGC, Bielfeldt W, Zachos FE *et al*. (2004) Mitochondrial DNA and microsatellite analyses of the genetic status of the presumed subspecies *Cervus elaphus montanus* (Carpathian red deer). *Heredity*, **93**, 299–306.

Frankham R, Ballou JD, Briscoe DA (2004) *A Primer of Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Geisser H, Reyer H-U (2004) Efficacy of hunting, feeding, and fencing to reduce crop damage by wild boars. *Journal of Wildlife Management*, **68**, 939–946.

Geisser H, Reyer H-U (2005) The influence of food and temperature on population density of wild boar *Sus scrofa* in the Thurgau (Switzerland). *Journal of Zoology London*, **267**, 89–96.

Genov P, Nikolov H, Massei G, Gerasimov S (1991) Craniometrical analysis of Bulgarian wild boar (*Sus scrofa*) populations. *Journal of Zoology, London*, **225**, 309–325.

Giuffra E, Kijas JMH, Amarger V *et al*. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**, 1785–1791.

Gongora J, Fleming P, Spencer PBS *et al*. (2004) Phylogenetic relationships of Australian and New Zealand feral pigs assessed by mitochondrial control region sequence and nuclear GPIP genotype. *Molecular Phylogenetics and Evolution*, **33**, 339–348.

Goudet J (2001) FSTAT, *a Program to Estimate and Test Gene Diversities and Fixation Indices Version 2.9.3*. Available from URL: http://www.unil.ch/izea/softwares/fstat.html.

Groves C (1981) *Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus Sus*. Research School of Pacific Studies, Australian National University, Canberra, Australia.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics*, **48**, 361–372.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hofreiter M, Serre D, Rohland N *et al*. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences, USA*, **101**, 12963–12968.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 54–70.

Jedrzejewska B, Jedrzejewski W, Bunevich AN, Milkowski L, Krasinski A (1997) Factors shaping population densities and increased rates of ungulates in Bialowieza Primeval Forest (Poland and Belarus) in the 19th and 20th centuries. *Acta Theriologica*, **42**, 399–451.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kijas JMH, Andersson L (2001) A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution*, **52**, 302–308.

Kim K-I, Lee J-H, Li K *et al*. (2002) Phylogenetic relationship of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. *Animal Genetics*, **33**, 19–25.

Langella O (2002) POPULATIONS 1.2.28. *Population Genetic Software (Individuals or Populations Distances, Phylogenetic Trees)*. Available from URL: http://www.bioinformatics.org/project/?group_id=84.

Larson G, Dobney K, Albarella U *et al*. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Larson G, Albarella U, Dobney K *et al*. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences, USA*, **104**, 15276–15281.

Lin C-S, Sun Y-L, Liu C-Y *et al*. (1999) Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene*, **236**, 107–114.

Lucchini V, Meijaard E, Diong CH, Groves CP, Randi E (2005) New phylogenetic perspectives among species of South-east Asian wild pig (*Sus* sp.) based on mtDNA sequences and morphometric data. *Journal of Zoology*, **266**, 25–35.

Magri D, Vendramin GG, Comps B *et al*. (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199–221.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, UK.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Okumura N, Kurosawa Y, Kobayashi E *et al*. (2001) Genetic relationship amongst the major non-coding regions of mitochondrial DNAs in wild boars and several breeds of domesticated pigs. *Animal Genetics*, **32**, 139–147.

Oliver WLR, Brisbin IL, Takahashi S (1993) The Eurasian wild pig, *Sus scrofa*. In *Pigs, Peccaries and Hippos – Status Survey and Action Plan*. (ed. Oliver WLR), pp. 112–121. IUCN/SSC, Gland, Switzerland.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit RJ, Aguinagalde I, de Beaulieu J-L *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

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

Randi E (2005) Management of wild ungulate populations in Italy: captive-breeding, hybridisation and genetic consequences of translocations. *Veterinary Research Communications*, **29**, 71–75.

Randi E, Apollonio M, Toso S (1989) The systematics of some Italian populations of wild boar (*Sus scrofa* L.): a craniometric and electrophoretic analysis. *Zeitschrift für Säugetierkunde*, **54**, 40–56.

Randi E, Mucci N, Pierpaoli M, Douzery E (1998) New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome *b* gene. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 793–801.

Randi E, D'Huart JP, Lucchini V, Aman R (2002) Evidence of two genetically deeply divergent species of waterhog, *Phacochoerus africanus* and *P. aethiopicus* (Artiodactyla: Suiformes) in East Africa. *Mammalian Biology*, **67**, 91–96.

Randi E, Alves PC, Carranza J, Milošević-Zlantanović S, Sfougaris A (2004) Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Molecular Ecology*, **13**, 3071–3083.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rogers A, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Sacks BN, Brown SK, Ernest HB (2004) Population structure of California coyotes corresponds to habitat-specific breaks and illuminates species history. *Molecular Ecology*, **13**, 1265–1275.

Sáez-Royuela C, Tellería JL (1986) The increased population of the wild boar (*Sus scrofa* L.) in Europe. *Mammal Review*, **16**, 97–101.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, **205**, 125–140.

Schley L, Roper TJ (2003) Diet of wild boar *Sus scrofa*. Western Europe, with particular reference to consumption of agricultural crops. *Mammal Review*, **33**, 43–56.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Singer FJ, Swank WT, Clebsch EEC (1984) Effects of wild pig rooting in a deciduous forest. *Journal of Wildlife Management*, **48**, 464–473.

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

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Ursing BM, Arnason U (1998) The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Journal of Molecular Evolution*, **47**, 302–306.

Van Andel TH (1989) Late Quaternary sea-level changes and archaeology. *Antiquity*, **63**, 733–745.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: a software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vernesi C, Pecchioli E, Caramelli D *et al*. (2002) The genetic structure of natural and reintroduced roe deer (*Capreolus capreolus*) populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. *Molecular Ecology*, **11**, 1285–1297.

Vernesi C, Crestanello B, Pecchioli E *et al*. (2003) The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. *Molecular Ecology*, **12**, 585–595.

Vilà C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends in Genetics*, **21**, 214–218.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Welander J (2000) Spatial and temporal dynamics of wild boar (*Sus scrofa*) rooting in a mosaic landscape. *Journal of Zoology*, **252**, 263–271.

Yang J, Wang J, Kijas J, Liu B, Han HYuM, Yang H, Zhao S, Li K (2003) Genetic diversity present within the near-complete mtDNA genome of 17 breeds of indigenous Chinese pigs. *Journal of Heredity*, **94**, 381–385.

This research is a result of the collaboration among the Department of Zoology and Evolutionary Genetics of the University of Sassari, the Department of Biology and Evolution of the University of Ferrara, the Centre of Alpine Ecology in Trento, and the Biosfera Association in Florence. Massimo Scandura is a researcher and Laura Iacolina and M. Francesca Di Benedetto are PhD students, members of the team led by Marco Apollonio, interested in the behavioural ecology and population genetics of wild mammals. Elena Pecchioli and Barbara Crestanello are researchers interested in conservation genetics of wild vertebrates in mountain ecosystems. Vincenzo Russo and Roberta Davoli work on gene expression in cattle and pig breeds at the Zootechnical Center of the University of Bologna. Giorgio Bertorelle is interested in the reconstruction of demographic and selective processes using population genetics data in different species.

Supplementary materials

The following supplementary material is available for this article:

Table S1 Segregating sites and different haplotypes observed at the mtDNA control region (411 bp) of 145 European wild boar and 47 pig samples sequenced in this study. Haplotype frequencies observed in wild boar sampling areas and pig breeds are shown. The following abbreviations are used for wild boar sampling areas: IAre (Arezzo, Italy), IFor (Forlì, Italy), ISie (Siena, Italy), ISal (Salerno, Italy), IFlo (Florence, Italy), IMrp (Maremma Regional Park, Italy), ISrp (San Rossore Regional Park, Italy), ICpr (Castel Porziano Presidential Reserve, Italy), ISar (Sardinia, Italy), IGor (Gorizia, Italy), SCot (El Coto, Spain), FHam (Haute Marne, France), ADif (Austria, different areas), PDif (Poland, different areas), HDif (Hungary, different areas). The following abbreviations are used for domestic pig breeds: Pig 1 (Cinta Senese), Pig 2 (Sarda), Pig 3 (Calabrese), Pig 4 (Casertana) and Pig 5 (Nera Siciliana).

Table S2 List of the 804 *Sus scrofa* mitochondrial control region sequences downloaded from GenBank or obtained by the authors which were used in the study.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/
10.1111/j.1365-294X.2008.03703.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.