Mitochondrial phylogeography of the European wild boar: The effect of climate on genetic diversity and spatial lineage sorting across Europe

Article in Journal of Biogeography · January 2014
Impact Factor: 4.59 · DOI: 10.1111/jbi.12268

19 authors, including:

Laura Iacolina
Aalborg University
58 PUBLICATIONS 265 CITATIONS

Ladislav Paule
University of Technology in Zvolen
113 PUBLICATIONS 1,986 CITATIONS

Christian Gortázar
University of Castilla-La Mancha
371 PUBLICATIONS 8,127 CITATIONS

Massimo Scandura
Università degli Studi di Sassari
87 PUBLICATIONS 621 CITATIONS

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Massimo Scandura
Retrieved on: 21 May 2016
Mitochondrial phylogeography of the European wild boar: the effect of climate on genetic diversity and spatial lineage sorting across Europe

Sibelle T. Vilaça¹, Daniela Biosa², Frank Zachos³,⁴, Laura Iacolina², Julia Kirschning³, Paulo C. Alves⁵,⁶, Ladislav Paule⁷, Christian Gortazar⁸, Zizis Mamuris⁹, Bogumiła Jędrzejewska¹⁰, Tomasz Borowik¹⁰, Vadim E. Sidorovich¹¹, Josip Kusak¹², Stefano Costa¹³, Laurent Schley¹⁴, Günther B. Hartl¹⁴, Marco Apollonio², Giorgio Bertorelle¹⁴ and Massimo Scandura²*

¹Dipartimento di Biologia ed Evoluzione, Università di Ferrara, 44121 Ferrara, Italy, ²Dipartimento di Scienze della Natura e del Territorio, Università di Sassari, 07100 Sassari, Italy, ³Natural History Museum Vienna, A-1010 Vienna, Austria, ⁴Zoological Institute, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany, ⁵CIBIO, Faculdade de Ciências, Universidade do Porto, 4485-661 Vairão, Portugal, ⁶Wildlife Biology Program, University of Montana, Missoula, MT 59812, USA, ⁷Faculty of Forestry, Technical University in Zvolen, 960 53 Zvolen, Slovakia, ⁸SaBio IREC, CSIC-UCLM-JCCM, 13071 Ciudad Real, Spain, ⁹Department of Biochemistry and Biotech, University of Thessaly, 41221 Larissa, Greece, ¹⁰Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland, ¹¹Institute of Zoology, National Academy of Sciences of Belarus, 220072 Minsk, Belarus, ¹²Department of Biology, Faculty of Veterinary Medicine, University of Zagreb, 10000 Zagreb, Croatia, ¹³Laboratorio Chimico della Camera di Commercio di Torino, 10127 Torina, Italy, ¹⁴Administration de la Nature et des Forêts, L-2453 Luxembourg

ABSTRACT

Aim Climate changes in the past had a deep impact on the evolutionary history of many species and left genetic signatures that are often still detectable today. We investigated the geographical pattern of mitochondrial DNA diversity in the European wild boar (Sus scrofa). Our final aims were to clarify the influence of present and past climatic conditions, infer the geographical position of glacial refugia, and suggest post-glacial spatial dynamics.

Location Europe.

Methods D-loop sequences were obtained for 763 individuals from Portugal to western Russia. Phylogenetic, multivariate and interpolation methods were used to describe the genetic and geographical patterns. Climatic suitability during the Last Glacial Maximum (LGM) was predicted using MaxEnt. The effect of present and past suitability on the observed patterns of diversity was evaluated by multiple linear regression.

Results We confirmed the existence of a ubiquitous mitochondrial clade in Europe (E1), an endemic clade in Italy (E2) and a few East Asian haplotypes (A), presumably introgressed from domestic pigs. No Near Eastern haplotypes were detected. Genetic divergence was not simply related to geographical distance. A clear south–north decreasing gradient of diversity was observed, with maximum levels in putative glacial refugia. Latitudinal variation in climatic conditions during the LGM was shown to be a good predictor of current genetic diversity. Moreover, an unexpected similarity between Iberia and eastern Europe was observed, while central European populations showed a higher affinity to the Italian gene pool.

Main conclusions The current distribution of mitochondrial genetic diversity was highly influenced by past climatic events, especially those related to the LGM, and is consistent with a major contribution of the Italian peninsula and the Balkans to the post-glacial recolonization of northern areas. More recent processes, such as restocking and extensive hunting, probably acted at rather local scales, without great impact on the global pattern of mitochondrial diversity.

Keywords Climate change, genetic differentiation, glacial refugia, Last Glacial Maximum, mtDNA, phylogeography, Sus scrofa.
INTRODUCTION

The phylogeographical patterns observed in many European mammals are thought to be strongly related to Quaternary climatic fluctuations (Hewitt, 2004). The extensive ice cover of northern regions during ice ages temporarily confined thermophilic species to lower latitudes, whereas northward recolonizations took place during the milder interglacials. In particular, most of the current phylogeographical discontinuities are likely to have arisen during the last glaciation (c. 110,000–12,000 years ago), as suggested also by ancient DNA studies (e.g. Hofreiter et al., 2004). During the Last Glacial Maximum (LGM; 23,000–16,000 years ago), ice caps covered northern regions and the main mountain chains, and most of continental Europe north of 45° N was covered by permafrost (Vandenberghe et al., 2012). Iberia, Italy, the Balkans and the Caspian/Caucasus region acted as refugia for remnant populations and represented the source for the following recolonization of northern Europe for several species (Taberlet et al., 1998; Hewitt, 2004).

In agreement with the traditional 'southerly refugia model', present-day genetic diversity is expected to decrease gradually from refugia to recolonized areas. A progressive loss of genetic diversity at increasing distance from refugia is expected if the recolonization was fast and if it followed a 'leading-edge expansion model', where only populations at the periphery of refugial regions contributed to repopulating the rest of the continent (Hewitt, 2004).

The Eurasian wild boar (Sus scrofa Linnaeus, 1758) is a temperate species currently widespread in Europe, where it is one of the most important game species (Apollonio et al., 2010). Apart from human exploitation, climate is the main limiting factor for it, either through its effect on physiology and metabolism or through its indirect effect on food availability and accessibility (Geisser & Reyer, 2005; Melis et al., 2006). First, harsh winters and hot summers represent a problem for the thermoregulation of piglets (Berg et al., 2006). Second, highly energetic food, like acorns or agricultural crops, has to be consumed throughout the year in order to survive the limiting season (Schley & Roper, 2003), when frozen or very dry soils can limit access to underground food (Herrero et al., 2004). As a consequence, under extreme climatic conditions, the species may have a high mortality rate.

Such a limited adaptation to severely cold conditions suggests that the geographical distribution of the wild boar in Europe was largely influenced by Pleistocene and Holocene climatic oscillations. Fossil remains indicate that the species survived during the LGM in Iberia, southern France, the Italian peninsula and the Balkans (Sommer & Nadachowski, 2006). Genetic data, based on a numerically and geographically restricted sampling, support the view that, and possibly the Balkans, acted as major genetic reservoirs (Larson et al., 2005; Scandura et al., 2008; Alves et al., 2010; Alexandri et al., 2012). When the climate became milder, the recolonization of the continent would have started from these southern refugia, with their contributions probably depending on the effective size of relict populations and on local environmental conditions.

Here, we investigate the distribution of wild boar mitochondrial DNA (mtDNA) lineages in Europe, considering as explanatory variables the geographical location of the samples and the present and past climatic conditions. We assembled a large dataset by adding 467 new sequences of European wild boar to previously published ones, thus filling important gaps in Central and Eastern Europe and covering the area from the Atlantic coast of Portugal to western Russia. This large dataset allowed us to address five specific questions. (1) How is genetic diversity geographically partitioned in Europe? (2) Does genetic diversity show a latitudinal cline, being higher in southern glacial refugia? (3) Does genetic diversity correlate with present and past climatic suitability? (4) Does genetic divergence increase with spatial distance? (5) Can past climatic changes and subsequent demographic and range fluctuations explain the observed geographical partitioning of mitochondrial lineages?

MATERIALS AND METHODS

Sampling

Tissue samples provided by local hunters were collected in 16 countries (Germany, Luxembourg, France, Portugal, Italy, Greece, Croatia, Bosnia–Herzegovina, Serbia, Slovakia, Romania, Bulgaria, Poland, Belarus, Ukraine and Russia). Areas that were under-represented in previous studies (Larson et al., 2005; Scandura et al., 2008; Alves et al., 2010; Alexandri et al., 2012) were specifically targeted. We also collected information on the historical distribution of the wild boar in most of Europe, and on how translocations, habitat fragmentation and overhunting might have influenced its natural distribution and demography (see Appendix S1 in Supporting Information).

Sequencing

We sequenced a total of 467 wild boar specimens from 36 locations. Sequences from Tunisia (n = 77), previously published by Hajji & Zachos (2011), only partly overlapped with the D-loop alignment and were therefore extended by re-sequencing. Total genomic DNA was isolated using a commercial DNA isolation kit (Sigma-Aldrich, St Louis, MO, USA; Qiagen, Hilden, Germany). Laboratory analyses consisted of amplifying the mitochondrial D-loop and sequencing a fragment of 411 base pairs (bp) with the BigDye Terminator 3.1 methodology (Applied Biosystems, Foster City, CA, USA) after purification with ExoSAP (ThermoFisher Scientific, Waltham, MA, USA). Detailed methodological procedures are described in Scandura et al. (2008).

Electropherograms were visually inspected, and base calls edited in FinchTV 1.2 (Geospiza, Seattle, WA, USA). By virtue of the quality of the electropherograms and the shortness
of the region, most sequences were obtained with a single (forward) primer. Nonetheless, to ensure accuracy of nucleotide identification, a subset of samples was sequenced in the reverse direction, as were all individuals assigned to singleton haplotypes and all samples showing double peaks at any nucleotide position.

**Phylogenetic analysis of D-loop sequences**

The 467 new sequences were aligned using the ClustalX algorithm implemented in MEGA 4.0 (Tamura et al., 2007) together with 632 sequences retrieved from GenBank. The downloaded sequences represented animals classified as wild *S. scrofa* from Europe (*n = 361*), North Africa (*n = 77*) and Asia (*n = 194*) (sequences are listed in Appendix S2).

The final dataset combined the published and newly produced sequences of the mitochondrial control region (411-bp region) and comprised sequences from 1099 wild boars from three continents. Haplotypes were collapsed in Collaps e 1.2 (Posada, 2011), and new haplotypes were deposited in GenBank (accession numbers K C608827–K C608847). A median-joining (MJ) network of haplotypes (Bandelt et al., 1999) was then built in Network 4.6 (Fluxus Technologies, Clare, UK).

The most appropriate model of nucleotide change was selected using jModelTest 0.1.1 (Posada, 2008), including a sequence of Sus barbatus as outgroup (GenBank accession number A J314540). The best model, according to both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), was the HKY model (Hasegawa et al., 1985) with gamma-distributed (G) rate variation across sites.

Bayesian phylogenetic analyses were carried out in MrBayes 3.2 (Ronquist & Huelsenbeck, 2003) using the HKY+G model of sequence evolution and two independent runs of four Markov chains (one cold and three heated) over 1,000,000 generations and sampling every 100 generations. The first 25% of the sampled trees and estimated parameters were discarded as burn-in. Plots of log-likelihood scores against generation times were used to identify the point at which log-likelihood values reached stationarity. The final consensus tree was drawn in MEGA.

**Spatial differentiation and isolation by distance in Europe**

To assess the spatial differentiation among European wild boar, we excluded all non-European sequences and those for which no detailed geographical information was available. Asian haplotypes found in wild boars sampled in Europe were excluded, because they were attributed to local introgression with domestic pigs (Scandura et al., 2011). In total, 763 sequences were considered from 77 sites across 19 countries (Appendices S2 & S3a). Some sites with sample size below 10 were pooled. The choice of pooling was based on the geographical location and the genetic composition: only adjacent sites not separated by physical barriers and showing similar haplotypes were grouped. One population in southern Italy (ISal), having *n = 7* (after removal of three Asian haplotypes), was kept separate, as its allelic composition differed markedly from the nearest populations. After pooling, the final number of populations used in the statistical analyses was 39 (Appendix S3b).

The genetic structure in Europe was analysed with spatial factor analysis (spFA; Frichot et al., 2012) in R (R Development Core Team, 2012). This analysis uses the geographical information (coordinates) to correct for the effects of spatial autocorrelation in the exploratory analysis of genetic data (allele frequencies). Compared to similar methods, spFA appears to better remove the distortion introduced by the decay of genetic similarity with geographical distance when the genetic structure is inferred from a principal components analysis (Frichot et al., 2012).

The software Arlequin 3.5 (Excoffier & Lischer, 2010) was used to run a Mantel test (Mantel, 1967) to investigate whether an isolation-by-distance (IBD) model could explain the geographical pattern observed in Europe. The occurrence of IBD was tested for 38 European populations (Sardinia was excluded) by looking at the correlation between spatial distances and linearized genetic distances (*φST*/(1 − *φST*)). Geographical distances were computed using either the linear Euclidean distance or the minimum land distance between two sampling areas, considering the sea and high elevations (> 2000 m a.s.l.) to represent barriers for wild boar.

**Genetic diversity**

Haplotype diversity (*Hk*; Nei, 1987) and nucleotide diversity (*π*) were computed in Arlequin for the 39 populations. As sample sizes differed considerably among areas, an unbiased estimate of allelic richness (AR) was also calculated with Contrib 1.0 (Petit et al., 1998). To summarize the spatial distribution of genetic diversity, *Hk* values for 38 of the 39 populations were interpolated using the ordinary kriging (OK) method and the Spatial Analyst extension in ArcGis 10 (ESRI, Redlands, CA, USA). Ordinary kriging is a geostatistical interpolator method that creates a smooth surface even when sampling is spatially uneven. The sample from Russia – the northernmost sampling region in our study – was excluded from this analysis. When included, it generated an isolated point of very high diversity, deforming the global pattern in the north-eastern area. This population resulted from pooling different sampling sites within a radius of 540 km, all of which had small sample sizes. Such forced grouping, coupled with the fact that this area was possibly affected by the post-glacial expansion of more eastern (unsampled) populations, may have led to biased estimates of genetic diversity and we therefore excluded Russia from this analysis as a precautionary measure.

Fu’s *F*0 (Fu, 1997) and Tajima’s *D* (Tajima, 1989) statistics were calculated in Arlequin. Significance was assessed by randomly generating samples under the hypothesis of selective neutrality and demographic stability. Excluding
strong selective effects on the mtDNA region analysed, P-values smaller than the significance threshold of 0.00128 (following the Bonferroni approach for multiple testing) can be considered evidence of deviation from demographic stability.

Modelling the species’ present and past (LGM) range

To evaluate the influence of past range variation on the present wild boar genetic diversity in Europe, we assessed the ecological suitability for the species at the time of the LGM, identifying putative refugia. We used the machine learning method based on maximum entropy implemented in the program MaxEnt 3.3.3 (Phillips et al., 2006) to predict the present wild boar distribution and that during the LGM. We used as presence data the geographical coordinates of the 77 sampling sites of this study, the sites used in Melis et al. (2006), and those available from the Global Biodiversity Information Facility (GBIF) database. As the latter was largely biased due to over-representation of points in some areas (e.g. France), we sub-sampled the GBIF locations to obtain an even density of points across countries and to produce a good representation of the different environmental contexts. A total of 215 locations were used, covering all the parts of Europe relevant to our study.

As climate represents the driving factor influencing other environmental variables that affect wild boar presence (such as habitat type, water and food availability), climatic variables were used to construct the climate prediction models. Specifically, we used annual mean temperature (BIO1), temperature seasonality (BIO4), annual temperature range (BIO7), mean temperature of the warmest quarter (BIO10), mean temperature of the coldest quarter (BIO11), annual precipitation (BIO12), precipitation of the wettest quarter (BIO16), and precipitation of the driest quarter (BIO17). Current and LGM data were downloaded from WorldClim 1.4 (Hijmans et al., 2005). Two different general circulation models were adopted for the LGM estimations – the Community Climate System Model (CCSM) and the Model for Interdisciplinary Research on Climate (MIROC). As snow cover is a crucial limiting factor for wild boar (Melis et al., 2006), we also included two variables from the Stage Three Project (van Andel, 2002): snow depth (in centimetres); and the number of days per year with snow cover. Layers were cropped to span from latitude 68° N to 33.8° N and from longitude 12° W to 51.7° E. All layers were used in their original spatial resolution (2.5 arc-minutes) and projection (WGS84 datum). Models were run in MaxEnt using default settings. To establish a threshold and transform the continuous maps into binary presence/absence maps, we used the 10% of training presences as a threshold value. We ran 10 replicates, and the average across all runs was calculated. To evaluate model performance, we used the area under the curve of the receiver operating characteristic (AUC), which measures the ability of a prediction to discriminate presence from absence (Elith et al., 2010) and ranges from 0.5 to 1. An AUC value of 0.5 indicates that the model has no predic-tive ability, whereas a perfect discrimination between suitable and unsuitable cells will achieve the maximum AUC, i.e. 1.0 (Morueta-Holme et al., 2010). We also performed a multivariate similarity surface (MESS) analysis when projecting in MaxEnt, following Elith et al. (2010). The MESS analysis allows novel climatic regions to be distinguished, by predicting values that fall outside the training range of variables.

To validate the palaeoclimatic map, we used fossil records compiled by William Davies (available at the Stage Three Project website: http://www.esc.cam.ac.uk/research/research-groups/oistage3) and also by Sommer & Nadachowski (2006). We only considered archaeological sites where the presence of S. scrofa had been reported. Two time intervals were considered: LGM (23,000–16,000 years ago) and older than 23,000 years ago. The occurrence of wild boar remains in a site dating back to the glacial period was considered as indicative of its presence in a glacial refugium.

Correlation between genetic diversity and environmental variables

To understand how present and past habitat suitability in Europe can explain the detected pattern of genetic diversity across the continent, we compared the effects of five predictive variables (latitude; longitude; present suitability; LGM suitability according to the MIROC model; LGM suitability according to the CCSM model) on Hk and AR by multiple linear regression in R (R Development Core Team, 2012). Data for Russia were not considered in the models, because genetic diversity can be biased in this population (see above). Both single effects and joint effects of variables were tested and the most parsimonious model was selected on the basis of the Akaike information criterion (AIC, ; Symonds & Moussalli, 2011).

RESULTS

Mitochondrial DNA diversity and phylogeography

The MJ network based on the large alignment including the Asian and North African haplotypes (in total 1099 individuals and 87 different haplotypes; Appendix S2) confirmed three major groups (Fig. 1): an Asian clade (corresponding to clade A from Giuffra et al., 2000); a pan-European (clade E1) clade; and a clade found only in Italy (clade E2). Only seven Asian haplotypes were observed among the 828 European sequences: four in Italy and one each in Germany, Luxembourg and Belgium. The European and Asian clades were separated by six fixed mutations (at positions 15545, 15568, 15573, 15583 and 15732 in the reference mtDNA sequence; Ursing & Arnason, 1998).

The analysis of 763 European sequences resulted in 50 haplotypes, corresponding to two A haplotypes, 41 E1 haplotypes and seven E2 haplotypes (Appendix S2). The Bayesian tree restricted to these haplotypes (Fig. 2a) gave high (≥ 95%)
posterior probabilities to the three clades, and also suggested the existence of an additional subgroup, E1a, within E1, matching the A-side group reported in other studies (Larson et al., 2007; Scandura et al., 2011; Alexandri et al., 2012). This subclade showed high frequencies in Italy, France, Germany, Austria and the north Adriatic (Fig. 2b); H022 and H023 were its most frequent haplotypes (matching A and BK, respectively, in Larson et al., 2005). Several private haplotypes occurred in Iberia, but the most common (H021, matching haplotype E in Larson et al., 2005) was shared with eastern populations. The derived E1a clade is very rare in both eastern and western areas, where the most basal haplotypes prevail (roughly matching the previously reported C-side group; Larson et al., 2007; Scandura et al., 2011) (Fig. 2b).

Besides the private haplogroup E2, Italian wild boar showed an exclusive E1 haplotype that was spread across the peninsula (H075). North Adriatic populations (from Croatia and north-eastern Italy) were dominated by a private haplotype (H083, subclade E1a), which differed by a single mutation from H023 (Fig. 1). Greece showed a large proportion of a private sequence (H128, 45%), matching haplotype G in Alexandri et al. (2012), which was reported to be common in this region. France and Germany had low numbers of haplotypes when compared to other European populations, whereas Sardinia had the largest number of haplotypes (14), most of which were private (64%).

In the spFA of wild boar populations (Fig. 3), factor 1 separated three geographical regions: Italy/Central Europe, Iberia, and the Balkans/Eastern Europe. A few exceptions to this pattern were observed: southern Portugal (population 21) appeared closer to eastern populations, while Bulgaria and Greece (populations 3 and 15, respectively) were intermediate between Iberian and Italy/Central Europe populations. The highest genetic distances were observed between Italy/Central Europe and eastern populations and not, as predicted by a simple geographical pattern, between the two disjointed groups separated by the largest geographical distances (Iberia and the Balkans/Eastern Europe). Sardinian wild boar (population 27) appeared intermediate between populations from mainland Italy and Austria.
Figure 2 Mitochondrial DNA haplogroups observed in 763 European wild boar (Sus scrofa). (a) Bayesian tree of the 50 mitochondrial D-loop haplotypes. The tree is rooted using a homologous sequence of Sus barbatus. Branch colours represent different haplogroups. Posterior probabilities and major clades are indicated on internodes. The E1 haplotypes not belonging to the well supported E1a clade were grouped into one haplogroup (E1c) for convenience. E1a and E1c roughly correspond to the haplogroups appearing in previous papers and referred to as A-side and C-side, respectively (Larson et al., 2007; Scandura et al., 2011; Alexandri et al., 2012). (b) Map showing the frequency of the three haplogroups (E1a, E1c and E2) in each of the 39 European populations (numbers indicate population codes).
No significant correlation was detected between linearized genetic distances and geographical distances (Euclidean distance, $r = 0.0003$, $P = 0.74$; least-cost distance, $r = -0.033$, $P = 0.78$).

The highest levels of genetic diversity were observed in Italy and Sardinia, mainly due to the presence of the divergent E2 haplotypes (Appendix S3b). The map of interpolated haplotype diversities throughout Europe based on 38 populations highlighted the pattern of higher diversity in southern regions, with a maximum in Greece, Italy and eastern Spain (Fig. 4). A similar pattern was observed for the distribution of allelic richness (results not shown). Fu’s $F$s and Tajima’s $D$ neutrality tests were not significant for the majority of the populations (Appendix S3b).

**Present and past (LGM) range**

Under the current distribution, the AUC values for the training and test data showed satisfactory values (0.886 and 0.817, respectively). AUC values above 0.8 are considered an excellent model prediction (Hosmer & Lemeshow, 2000). The MaxEnt estimation for the present was consistent with the current distribution of the wild boar and was able to predict even newly colonized areas, such as Finland and Sweden. Regarding the LGM distribution, both models predicted the occurrence of the wild boar in Iberia, southern France, Italy and the Balkans. The CCSM model showed smaller areas of
climatic suitability, especially in Italy and France, while the MIROC showed wider refugia for the wild boar (Fig. 5).

In fitting the model to the current wild boar distribution, it was possible to evaluate the relative contribution of variables to the presence/absence of the species and take their weight into account in constructing the predictive model. The most important variable was snow depth, followed by mean annual temperature. The standard deviation for the estimations was low, and most MESS values were positive (i.e. present in the training range). Negative MESS values were observed in northern areas, mainly because snow depth during the LGM could reach values currently not observed in Europe, and thus out of the training range of the model.

Predictions for the LGM were consistent with fossil records (Appendix S3c). The presence predicted by the MIROC model best reflected the fossil distribution, especially in southern France, where the model estimated a larger suitable area than the CCSM model.

Comparing the past suitability obtained with MaxEnt with the current distribution of haplotype diversity across the sampled populations (Fig. 4), it emerges that areas with higher diversity correspond to areas of predicted presence during the LGM, which were located at lower latitudes. Among the considered models (including geographical coordinates and present and past climatic suitability for wild boar as factors), the best-fitting model to explain haplotype diversity was that using latitude alone (Table 1), accounting for 17% of the overall variance ($B = -0.02$, $t = -2.99$, $P = 0.006$, adjusted $R^2 = 0.172$). It is, however, noteworthy that the best five models in Table 1 include geographical location and MIROC, the latter alone representing a good predictor of haplotype diversity ($B = 0.38$, $t = 2.84$, $P = 0.007$, adjusted $R^2 = 0.160$), as expected by virtue of its high correlation with latitude ($R = -0.62$, $t = -4.78$, $P < 0.01$). Conversely, current assessed suitability alone had no effect on haplotype diversity ($B = -0.17$, $t = -0.62$, $P = 0.542$, adjusted $R^2 = -0.017$).

DISCUSSION

Mitochondrial DNA sequences from throughout Europe allowed us to reconstruct the wild boar’s phylogeographical pattern and to identify the relevant factors that are likely to have generated it. Our main findings can be summarized as follows: (1) as a general pattern, mtDNA genetic diversity...
decreases northwards; (2) genetic diversity is better explained by climatic suitability during the Last Glacial Maximum than at the present day; (3) genetic and geographical distances are not correlated; (4) Italy and Central Europe show clear genetic similarities; (5) populations at the longitudinal extremes (Iberia and Eastern Europe) are genetically more similar than expected considering their geographical position; and (6) genetic introgression from domestic breeds into the wild boar lineages appears to be limited, due to low frequency of Asian haplotypes, restricted to a few localities.

South–north gradient of genetic diversity

Wild boar populations from northern areas, especially in central Europe, show very low genetic diversity. Southern areas, on the contrary, are more variable, and the Italian peninsula has the highest values of haplotype and nucleotide diversity. High levels of variation are also observed in Iberia (increasing from Portugal to eastern Spain) and in the Balkans, where Greece shows the highest values, in agreement with Alexandri et al. (2012). The Sardinian population shares the major mtDNA groups with the Italian mainland, but many haplotypes are exclusive to the island (see also Scandura et al., 2008).

The existence of higher genetic diversity at lower latitudes suggests that southern areas had an important role as genetic reservoirs during the last glaciation. This effect is confirmed by the correlation between haplotype diversity and the predicted climatic suitability for the species during the LGM. The current suitability, on the other hand, turned out to be a poor predictor of genetic variation, suggesting that an equilibrium situation (where high suitability corresponds to large populations and high genetic diversity) has not been reached after at least 15,000 years of warmer conditions.

An unusual phylogeographical pattern

The geographical distribution of clades and individual haplotypes from Iberia to the Balkans and western Russia shows some specific features not observed in other species. Excluding a few Asian sequences that are likely to be related to introgression events from domestic animals, only two major clades are observed, E1 and E2. E1 is widespread, whereas E2 is found only in the Italian peninsula and Sardinia. When the frequencies of single haplotypes and phylogenetic clades are analysed, a clear genetic affinity emerges between Italian and central European areas, and the eastern and western regions appear distinct but still showing some degree of genetic affinity.

In fact, the two longitudinal extremes in our sample (Iberia and Eastern Europe) share their most frequent haplotypes, a pattern already appreciated in other studies (Scandura et al., 2008; Alexandri et al., 2012) but with a far smaller sample of sequences. The spFAD also highlighted this similarity, showing few reciprocal mismatches between Iberian and East European populations and, more generally, supporting the genetic proximity of eastern and western areas, which is not compatible with a simple pattern of isolation-by-distance.

This global pattern also seems incompatible with what we know about recent translocations, which were common but probably only left a minor genetic signature at a local scale (e.g. Vernesi et al., 2003). Instead, we believe that the last glaciation and the subsequent recolonization processes from southern refugia were important factors in generating this geographical pattern.

Location and role of LGM refugia and post-glacial recolonization routes

The results obtained with MaxEnt, integrated with fossil data (Sommer & Nadachowski, 2006; Appendix S3c), point to the following refuge areas: Balkans, Italy, Southern France and Iberia. These refugia are shared by many other species (Hewitt, 2004; Schmitt, 2007; Sommer & Zachos, 2009) including ungulates and, given their current genetic diversity, we interpret them as LGM genetic reservoirs. For example, both spatial predictive models (Stockwell & Peters, 1999) and fossil records (Banks et al., 2008; Sommer et al., 2008) suggest that, during the LGM, the red deer, Cervus elaphus, took refuge in the same areas identified for the wild boar, and the same is true for the second most widespread European ungulate, the roe deer, Capreolus capreolus (Sommer et al., 2009).

But what happened in terms of connectivity and migration before, during, and after the last glaciation? The simplest hypothesis to justify the current geographical distribution of wild boar mtDNA lineages requires a stronger isolation of the Italian populations than the other European groups, probably occurring both before and during the last glaciation, and two major post-glacial colonization routes: one starting from the Italian and/or the south-western France refugium, and leading wild boars into central Europe, and another starting from the Balkans with wild boars colonizing the north-eastern regions. Conversely, the contribution of Iberian populations seems negligible.

The similarity between Eastern Europe and Iberia may therefore reflect the pre-LGM distribution, when Iberia, Central Europe and Eastern Europe might have formed a single, possibly panmictic group. This scenario is consistent with previous findings revealing weak phylogeographical structuring in pre-LGM populations of European mammals (Hofreiter et al., 2004). On the other hand, the suggested lack of differentiation across Europe in the interglacial implies that the sharing of E1a sequences between Central Europe and Italy should be mostly due to gene flow from the peninsula, while more recent secondary contacts (natural recolonization or translocations from France) could have played a minor role.

We hypothesize that when the ice cap retreated, the recolonization of suitable regions by the southern remnant populations was driven by a density-dependent leading-edge pulse. Rapid expansions were sustained by dense populations, where
the major source of migrants was represented by their northern portions (i.e. the edge). Such a mechanism could explain the genetic longitudinal discontinuity we observed in Europe. In particular, it can explain the glaring dissimilarity between populations across the Pyrenees, where none of the seven haplotypes observed in around 80% of the Iberian individuals was found in France, Luxembourg or Germany. Accordingly, the recolonization of central Europe could have started from high-density populations inhabiting the Italian peninsula and southern France. This expansion presumably prevented further expansion from the Iberian refugium, possibly hosting less abundant populations. In this scenario, dispersal across the Alps would have been assured by the existence of several passes lying at medium elevations (< 2000 m a.s.l.), well accessible for this species. Simultaneously, wild boars from the Balkans would have recolonized north-eastern regions, with a minor contribution to Central Europe. As a result, the two extremes of Europe remained isolated, although they still share some of their pre-LGM diversity. Due to the leading-edge process, southern populations within the Italian and Balkan refugia did not contribute much to the recolonization of new suitable regions in the north, as suggested by their high diversity and genetic distinctiveness (for Greece, see also Alexandri et al., 2012). Such a complex demographic and historical model can also explain the absence of isolation by distance, but it certainly needs further analyses and validation.

The suggested role of Italian refugial populations to the post-glacial recovery of wild boar in Europe may appear to be contradicted by the absence of the E2 clade north of the Alps. E2 haplotypes were (Larson et al., 2007) and remain common only in mainland Italy and Sardinia, although they were also detected in ancient specimens from Croatia (dated to around 11,000 years ago; Larson et al., 2007). Their current absence from the rest of Europe can be attributed to a low frequency in leading-edge populations at the time of the post-glacial population expansion, and to successive drift events, due to demographic oscillations. The moderate frequency of E2 we observed in northern Italy is more likely to be related to recent dispersal or translocation events, because the wild boar was extinct in this region at the beginning of the 20th century (Apollonio et al., 1988).

CONCLUSIONS

As recently noted, a latitudinal gradient of intraspecific genetic diversity seems to be the rule in mammals (Adams & Hadly, 2013). In addition to greater species richness, low latitudes tend to show a higher genetic variation within species than regions closer to the pole (Guo, 2012). This pattern can be attributed to periodic global climate changes which led to repeated extinctions toward the poles, followed by natural recolonizations (Hewitt, 2004).

This general trend is observed also in the European wild boar, despite the fact that translocation, restocking activities and extensive hunting have been common in this species in recent times. Our results are consistent with a prominent role for climatic and habitat oscillation during the Quaternary. Southern areas acted as genetic reservoirs in glacial times, and northern areas were mainly recolonized from Italian and French refugia in central Europe, and from the Balkans in Eastern Europe. Leading-edge expansion and density-dependent migration processes are also required to explain the complex mtDNA phylogeographical pattern we observed. Further studies on additional (nuclear) markers are needed to test our biogeographical reconstruction.

ACKNOWLEDGEMENTS

Sequencing activities were funded by the Sardinian Regional Government (L.R. 7/2007 Promozione della Ricerca Scientifica e dell’Innovazione Tecnologica in Sardegna). D.B. had a PhD grant financed by the European Social Fund, Operational Programme 2007/2013 – Axis IV – Human Capital Investment. S.T.V. was supported by a PhD Fellowship for International students of the University of Ferrara. B.J. and T.B. were supported by grants N N304 253935 from the Polish Ministry of Science and Higher Education and Research Potential in Conservation and Sustainable Management of Biodiversity – BIOCONSUS from the EC’s 7th Framework Programme (contract no. 245737). Mobility between labs in Sassari and Kiel was financed by the Italy–Germany Vigoni cooperation program 2010–2011. We are grateful to R. Chi-richella, A. Canu, G. Falceri, the Population and Conservation Genetics group at the University of Ferrara, and especially to A. Benazzo and S. Hoban, for their help in statistical analysis, English review, and insightful discussions. Thanks are due to CASPUR standard HPC grant (2011) and CINECA-ISCRA HPC class C grant (2010) for computational resources awarded to S.T.V.

REFERENCES

Mitochondrial phylogeography of the European wild boar


shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? Molecular Ecology, 17, 1745–1762.


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Current status of wild boar in different European countries and comparison with the early 20th century.

Appendix S2 Information on samples and sequences used for phylogenetic and phylogeographical analyses.

Appendix S3 Additional information and calculations on wild boar distribution and mtDNA diversity in Europe.

BIOSKETCHES

Sibelle T. Vilaça was a PhD student at the University of Ferrara in Giorgio Bertorelle’s team, and she is mainly interested in conservation genetics and vertebrate phylogeography.

Massimo Scandura is a tenured researcher at the University of Sassari, engaged in studies on population genetics, ecology and management of wild mammals.

Author contributions: M.S. and G.B. conceived the idea of this investigation; M.S., M.A. and G.B.H. established the network of collaborations; D.B., L.I. and J.Ki. performed the laboratory work. S.T.V., M.S. and D.B. performed the statistical analyses; S.T.V., M.S., G.B. and F.Z. wrote the manuscript. All other authors contributed to sampling and revised a preliminary version of the manuscript. All authors have approved the final version of the manuscript.

Editor: Peter Linder