Original article

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

Sibelle T. Vilaça[1], Daniela Biosa[2], Frank Zachos[3,4], Laura Iacolina[2], Julia Kirschning[4], Paulo C. Alves[5,6], Ladislav Paule[7], Christian Gortazar[8], Zizzis Mamuris[9], Bogumiła Jędrzejewska[10], Tomasz Borowik[10], Vadim E. Sidorovich[11], Josip Kusak[12], Stefano Costa[13], Laurent Schley[14], Günther B. Hartl[4], Marco Apollonio[2], Giorgio Bertorelle[1†], Massimo Scandura[2†]

[†] These authors contributed equally to this work.

[1] *Dipartimento di Biologia ed Evoluzione, Università di Ferrara, via L. Borsari, 46, 44121 Ferrara, Italy*

[2] *Dipartimento di Scienze della Natura e del Territorio, Università di Sassari, via Muroni 25, 07100 Sassari, Italy*

[3] *Natural History Museum Vienna, A-1010 Vienna, Austria*

[4] *Zoological Institute, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany*

[5] *CIBIO, Faculdade de Ciências, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal*

[6] *Wildlife Biology Program, University of Montana, College of Forestry and Conservation, 32 Campus Drive, Missoula, MT59812, USA*

[7] *Faculty of Forestry, Technical University in Zvolen, T. G. Masaryka 24, 960 53 Zvolen, Slovakia*

[8] *SaBio IREC, CSIC-UCLM-JCCM, Ronda de Toledo s.n. 13071 Ciudad Real, Spain*

[9] *Department of Biochemistry and Biotech, University of Thessaly, 26 Ploutonos St., 41221, Larissa, Greece*

[10] *Mammal Research Institute, Polish Academy of Sciences, ul. Waszkiewicza 1c, 17-230 Białowieża, Poland*

[11] *Institute of Zoology, National Academy of Sciences of Belarus, ul. Akademicheskaya 27, 220072 Minsk, Belarus*

[12] *Department of Biology, Faculty of Veterinary Medicine, University of Zagreb, Heinzelova 55, 10000 Zagreb, Croatia*

[13] *Laboratorio Chimico della Camera di Commercio di Torino, Via Ventimiglia 165, 10127 Torino, Italy*

[14] *Administration de la Nature et des Forêts, 16 rue Eugène Ruppert, L-2453, Luxembourg*

Corresponding author: Massimo Scandura, DipNeT, via Muroni 25, 07100 Sassari, Italy, Tel. +39-079228628, Fax +39-079228665, E-mail: scandura@uniss.it

Running title: Mitochondrial phylogeography of the European wild boar

Word count (inclusive of Abstract, text and References): 6906
Word count (Abstract only): 300
Estimated number of pages for Tables and Figures: 3.5

ABSTRACT

Aim Climate changes in the past had a deep impact on the evolutionary history of many species and left a genetic signature which is often still detectable today. We investigated the geographical pattern of mitochondrial DNA diversity in the European wild boar. 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 geographic 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 an 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 haplotype was 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). Extensive ice coverage of northern regions during ice ages temporarily confined thermophilous species to lower latitudes, whereas northward recolonizations took place during the milder interglacials. In particular, most of the current phylogeographical discontinuities likely arose during the last glaciation (~110,000 - 12,000 yr BP), as suggested also by ancient DNA studies (e.g., Hofreiter *et al.*, 2004). During the last glacial maximum (LGM, 23,000 - 16,000 yr BP), the ice cap 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 played a major role 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 gradually decrease moving from refugia to the 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 represents one of the most important game species (Apollonio *et al.*, 2010). Apart from human exploitation, climate is the main limiting factor for it, 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). Secondly, highly energetic food, like acorns or agricultural crops, has to be consumed through the year to overcome winter (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 during the LGM the species survived in Iberia, in southern France, in the Italian peninsula and in the Balkans (Sommer & Nadachowski, 2006). Genetic data, based on a numerically and geographically restricted sampling, support the view that Italy, 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 a likely different contribution, depending on the effective size of relict populations and on local environmental conditions.

In the present study 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: *i)* how is genetic diversity geographically partitioned in Europe? *ii)* does genetic diversity show a latitudinal cline, being higher in southern glacial refuges? *iii)* does genetic diversity correlate with present and past climatic suitability? *iv)* does genetic divergence increase with spatial distance? *v)* 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 and Herzegovina, Serbia, Slovakia, Romania, Bulgaria, Poland, Belarus, Ukraine, and Russia). Areas 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), overlapped only partially 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 or Qiagen, Hilden, Germany). Laboratory analyses consisted in amplification of 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) after purification with Exo/Sap (ThermoFicher Scientific, Waltham, MA). Detailed methodological procedures are described in Scandura *et al.* (2008).

Electropherograms were visually inspected, and base calls edited in FINCHTV 1.2. By virtue of the high quality of 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 a nucleotide position.

Phylogenetic analysis of D-loop sequences

The new 467 sequences were aligned using the CLUSTALX algorithm implemented in MEGA 4.0 (Tamura *et al.*, 2007) together with 632 published 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 the newly produced CR sequences (411-bp region) and corresponded to 1,099 wild boars from three continents. Haplotypes were collapsed in Collapse 1.2 (Posada, 2011). New haplotypes were deposited in GenBank (accession codes KC608827-KC608847). Then a median-joining (MJ) network of haplotypes (Bandelt *et al.*, 1999) was built in Network 4.6 (Fluxus Technologies Ltd.).

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 AJ314540). The best model, was the HKY model (Hasegawa *et al.*, 1985) with gamma-distributed (G) rate variation across sites, based on both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

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 (1 cold and 3 heated) with 1,000,000 generations and sampling every 100 generations. The first 25% of the sampling trees and estimated parameters were discarded as burn-in. Results of log-likelihood scores were plotted against generation times 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 boars, 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 from 77 sites were considered (see Appendices S2 and S3a), corresponding to 19 countries. Some sites with sample size smaller than ten 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 was highly different from the nearest populations. After pooling, the final number of populations used in the statistical analyses was 39 (see Appendix S3b).

The genetic structure in Europe was analysed with the Spatial Factor Analysis (spFA, Frichot *et al*., 2012) in R (R 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 the geographical distance when the genetic structure is inferred from a principal component 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 global 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 (Φ_{st} /(1- Φ_{st})). 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 (> 2,000 m a.s.l.) as 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 size 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, haplotype diversity values (*Hk*) for 38 out of 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 which creates a smooth surface suitable when sampling is uneven across space. The Russian sample, which is 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 and all having 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 precautionary excluded Russia from this analysis.

Fu's F_S (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 as evidence of deviation from demographic stability.

Modelling present and past (LGM) species' range

To evaluate the influence of past range variation on the present wild boar genetic diversity in Europe, we assessed 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 v.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 ones used in Melis *et al.* (2006), and those available from the GBIF (Global Biodiversity Information Facility) database. Since 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 have a good representation of the different environmental contexts. A total of 215 locations, fully covering the portion of Europe interested by our study, were used.

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 (Bio01), temperature seasonality (Bio04), annual temperature range (Bio07), 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 in binary presence/absence maps we used the 10% of training presence as a threshold value. We ran 10 replicates, and the average among all runs was calculated. To evaluate model performance, we used the Area Under the Curve (AUC) of the Receiver Operated Characteristics (ROC), 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 predictive 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 the multivariate similarity surface (MESS) analysis when projecting in MAXENT, according to Elith *et al.* (2010). The MESS analysis allows to distinguish novel climate regions, by predicting values that fall outside the training range of variables.

To validate the palaeoclimatic map we used fossil records compiled by William Daves (available on the Stage Three Project website) and also by Sommer & Nadachowski (2006). We considered only archaeological sites for which the presence of *S. scrofa* was reported. Two time intervals were considered: LGM (23,000-16,000 yr BP) and older than 23,000 yr BP. 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 refuge.

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 comparatively checked the effect of five predictive variables (latitude, longitude, present suitability, LGM suitability according to MIROC model, and LGM suitability according to CCSM model) on *Hk* and *AR* by a multiple linear regression. This analysis was performed using the software R (R 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 more variables were tested and the most parsimonious model was selected on the basis of the Akaike Information Criterion (AIC), using the corrected formula for small sample sizes (AICc; 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 1,099 individuals and 87 different haplotypes, Appendix S2) confirmed three major groups (Fig. 1): the Asian clade (corresponding to clade A from Giuffra *et al*., 2000), a pan-European (E1) clade and one clade exclusively found in Italy (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 the Asian clades were separated by six fixed mutations (at positions 15545, 15568, 15573, 15583, 15732 in the reference mtDNA sequence, Ursing & Arnason, 1998).

The analysis of 763 European sequences resulted in 50 haplotypes, corresponding to two A, 41 E1 and seven E2 (see Appendices S2). The Bayesian tree restricted to these haplotypes (Fig. 2a) gave high (≥ 95%) posterior probabilities to the three clades, also suggesting the existence of an additional sub-group, 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 sub-clade showed high frequencies in Italy, France, Germany, Austria and North Adriatic (Fig. 2b), and had H022 and H023 as the 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 (roughly matching the previously reported C-side group, Larson *et al.*, 2007 and Scandura *et al.*, 2011) prevail (Fig. 2b).

Besides the private haplogroup E2, Italian wild boars 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, sub-clade E1a), which differed by a single mutation from H023 (Fig. 1). Greece showed a large proportion of private sequences (45%, H128), matching haplotype G in Alexandri *et al*. (2012), which was reported as common in this region. France and Germany had low numbers of haplotypes when compared to the rest of European populations, whereas Sardinia had the largest number of haplotypes (14), most of which were private (64%).

In the spFA on wild boar populations (Fig. 3), factor 1 separated three geographical regions: Italy/Central Europe, Iberia, and 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

Balkans/Eastern Europe). Sardinian wild boar (population 27) appeared intermediate between populations from mainland Italy and Austria.

No significant correlation was detected between linearized genetic distances and geographical distances (Euclidean distance, $R^2 = 0.0003$, $P = 0.74$; least cost distance, $R^2 = -0.033$, $P = 0.78$).

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 *Fs* and Tajima's *D* neutrality tests were non-significant for the majority of the populations (see Appendix S3b).

Present and past (LGM) species' 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 as 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, like 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 when compared to 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, located at lower latitudes. Among the considered models including as factors geographical coordinates and present and past (LGM) climatic suitability for wild boar, the best fitting model to explain haplotype diversity was that represented by 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). Yet, it is 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). On the contrary, 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

MtDNA sequences from throughout Europe allowed us to reconstruct the wild boar’s phylogeographical pattern and to identify the relevant factors that likely generated it. Our most evident findings can be summarized as follows: 1) as a general pattern, mtDNA genetic diversity decreases moving toward northern areas; 2) genetic diversity is better explained by climatic suitability during the last glacial maximum than during modern times; 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; 6) genetic introgression from domestic breeds into the wild boar lineages appears 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 contrary, 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 single haplotypes from Iberia to the Balkans and western Russia shows some specific features not observed in other species. Excluding few Asian sequences likely 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. Also, the spFA remarked such similarity showing a few reciprocal mismatches between Iberian and East European populations and, more in general, supporting a genetic proximity between eastern and western areas, which is not compatible with a simple IBD pattern.

This global pattern also seems incompatible with what we know about recent translocations, which were common but have probably left only 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, see Appendix S3c), point to the following refuge areas: Balkans, Italy, Southern France and Iberia. These areas 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 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 (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 compared to 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 refugia, and leading wild boars into central Europe, and another starting from the Balkans with wild boars colonizing the north-eastern regions. On the contrary, 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 and leading-edge pulse. Rapid expansions were sustained by high density 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 (< 2,000 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 the Balkan refugia did not contributed 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 admittedly 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 in contradiction with the absence of the E2 clade north of the Alps. E2 haplotypes were (Larson *et al.*, 2007) and are common nowadays only in mainland Italy and Sardinia, though they were also detected in ancient specimens from Croatia (dated around 11,000 yr BP, Larson *et al.*, 2007). Their current absence in 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 related to recent dispersal or translocation events, since the wild boar was extinct in this region at the turn of the XIX and XX 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 more northerly regions (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 main effect of 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). DB had a PhD grant financed by the European Social Fund, Operational Programme 2007/2013 – Axis IV - Human Capital Investment. STV has been supported by a PhD Fellowship for International students of the University of Ferrara. BJ and TB 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. Chirichella, 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 STV.

REFERENCES

Adams, R. & Hadly, E. (2013) Genetic diversity within vertebrate species is greater at lower latitudes. *Evolutionary Ecology*, **27**, 133-143.

Alexandri, P., Triantafyllidis, A., Papakostas, S., Chatzinikos, E., Platis, P., Papageorgiou, N., Larson, G., Abatzopoulos, T. J. & Triantaphyllidis, C. (2012) The Balkans and the colonization of Europe: the post-glacial range expansion of the wild boar, *Sus scrofa*. *Journal of Biogeography*, **39**, 713-723.

Alves, P. C., Pinheiro, I., Godinho, R., Vicente, J., Gortazar, C. & Scandura, M. (2010) Genetic diversity of wild boar populations and domestic pig breeds (*Sus scrofa*) in South-western Europe. *Biological Journal of the Linnean Society*, **101**, 797-822.

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., Andersen, R. & Putman R. (eds) (2010) *European Ungulates and their Management in the 21st Century*. Cambridge University Press, Cambridge, UK. 604 pp.

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

Banks, W. E., d'Errico, F., Peterson, A. T., Kageyama, M. & Colombeau, G. (2008) Reconstructing ecological niches and geographic distributions of caribou (*Rangifer tarandus*) and red deer (*Cervus elaphus*) during the Last Glacial Maximum. *Quaternary Science Reviews*, **27**, 2568-2575.

Berg, F., Gustafson, U. & Andersson, L. (2006) The Uncoupling Protein 1 Gene *(UCP1)* Is Disrupted in the Pig Lineage: A Genetic Explanation for Poor Thermoregulation in Piglets. *PLoS Genetics*, **2**, e129.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330-342.

Excoffier, L. & Lischer, H. E. L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Frichot E, Schoville S, Bouchard G, François O. (2012) Correcting Principal Component Maps for Effects of Spatial Autocorrelation in Population Genetic Data. *Frontiers in Genetics*, **3**, 254.

Fu, Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and backgroud selection. *Genetics*, **147**, 915-925.

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*, **267**, 89-96.

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, O., Jeon, J. T. & Andersson, L. (2000) The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics*, **154**, 1785-1791.

Guo, Q. (2012) Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology*, **21**, 5396-5403.

Hajji, G. E. & Zachos, F. E. (2011) Mitochondrial and nuclear DNA analyses reveal pronounced genetic structuring in Tunisian wild boar *Sus scrofa*. *European Journal of Wildlife Research*, **57**, 449-456.

Hasegawa, M., Kishino, H. & Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160-174.

Herrero, J., Couto, S., Rosell, C. & Arias, P. (2004) Preliminary data on the diet of wild boar living in a Mediterranean coastal wetland. *Galemys*, **16**, 115-123.

Hewitt, G. M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 183-195.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hofreiter, M., Serre, D., Rohland, N., Rabeder, G., Nagel, D., Conard, N., Munzel, S. & Paabo, S. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 12963-12968.

Hosmer DW, Lemeshow S (2000) Applied Logistic Regression, second edition. New York, USA: John Wiley & Sons.

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

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J. D., Edwards, C. J., Schlumbaum, A., Dinu, A., Balacsescu, A., Dolman, G., Tagliacozzo, A., Manaseryan, N., Miracle, P., Van Wijngaarden-Bakker, L., Masseti, M., Bradley, D. G. & Cooper, A. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 15276-15281.

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

Melis, C., Szafranska, P. A., Jedrzejewska, B. & Barton, K. (2006) Biogeographical variation in the population density of wild boar (*Sus scrofa*) in western Eurasia. *Journal of Biogeography*, **33**, 803-811.

Morueta-Holme, N., Flojgaard, C. & Svenning, J. C. (2010) Climate Change Risks and Conservation Implications for a Threatened Small-Range Mammal Species. *PLoS ONE*, **5**, e10360.

Nei, M. (1987) *Molecular evolutionary genetics*, Columbia University Press, New York.

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

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

Posada, D. (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.

Posada, D. (2011) *Collapse: Describing haplotypes from sequence alignments. [Online]. Website last modified on May 28, 2011 (accessed on August 11, 2011).*

R Core Team (2012) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ronquist, F. & Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

Scandura, M., Iacolina, L. & Apollonio, M. (2011) Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal Review*, **41**, 125-137.

Scandura, M., Iacolina, L., Crestanello, B., Pecchioli, E., Di Benedetto, M. F., Russo, V., Davoli, R., Apollonio, M. & Bertorelle, G. (2008) Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? *Molecular Ecology*, **17**, 1745-1762.

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

Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Sommer, R. S. & Nadachowski, A. (2006) Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review*, **36**, 251-265.

Sommer, R. S. & Zachos, F. E. (2009) Fossil evidence and phylogeography of temperate species: 'glacial refugia' and post-glacial recolonization. *Journal of Biogeography*, **36**, 2013-2020.

Sommer, R. S., Fahlke, J. M., Schmolcke, U., Benecke, N. & Zachos, F. E. (2009) Quaternary history of the European roe deer *Capreolus capreolus*. *Mammal Review*, **39**, 1-16.

Sommer, R. S., Zachos, F. E., Street, M., Joris, O., Skog, A. & Benecke, N. (2008) Late Quaternary distribution dynamics and phylogeography of the red deer (*Cervus elaphus*) in Europe. *Quaternary Science Reviews*, **27**, 714-733.

Stockwell, D. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

Symonds, M. & Moussalli, A. (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, **65**, 13-21.

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.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Ursing, B. M. & Arnason, U. (1998) The Complete Mitochondrial DNA Sequence of the Pig (*Sus scrofa*). *Journal of Molecular Evolution*, **47**, 302-306.

van Andel, T. H. (2002) The climate and landscape of the middle part of the Weichselian glaciation in Europe: The Stage 3 Project. *Quaternary Research*, **57**, 2-8.

Vandenberghe, J., Renssen, H., Roche, D. M., Goosse, H., Velichko, A. A., Gorbunov, A. & Levavasseur, G. (2012) Eurasian permafrost instability constrained by reduced sea-ice cover. *Quaternary Science Reviews*, **34**, 16-23.

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

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 XX century.

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

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

BIOSKETCHES

S.T.V. was a PhD student at the University of Ferrara in G.B.'s team, and she is mainly interested in conservation genetics and vertebrate phylogeography. M.S. is a tenured researcher at the University of Sassari, engaged in studies on population genetics and management of wild mammal populations.

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.K. 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 have contributed to sampling and revised a preliminary version of the manuscript.

Editor: Peter Linder

Table 1

Comparison of the performance of multiple linear regression models explaining gene diversity in European wild boar populations on the basis of their geographical position (latitude, Lat, and longitude, Long), and of present (Current) and past (at LGM, MIROC and CSSM) environmental suitability. The Russian population was excluded for the reasons explained in the text. Environmental suitability was assessed in MaxEnt on the basis of presence data and eleven climatic variables. Climatic data at LGM were estimated on the basis of the MIROC and CSSM global climatic models. AICc, AIC corrected for small sample sizes; Δi, AICc difference between model *i* and the best model; Wi, weight of model *i*; cumW*i*, cumulative weight of model *i* and the upper models; adjR^2, adjusted R^2.

Model predictors	AICc	Δi	Wi	cumW*i*	adjR^2
Lat	-12.091	0	0.154	0.154	0.172
MIROC	-11.574	0.517	0.119	0.273	0.160
Lat+Long+MIROC	-11.531	0.559	0.116	0.390	0.223
Lat+MIROC	-11.440	0.651	0.111	0.501	0.189
Lat+Long	-11.026	1.065	0.090	0.591	0.180
MIROC+Current	-10.681	1.410	0.076	0.668	0.172
Lat+Long+MIROC+Current	-10.319	1.772	0.064	0.731	0.233
Long+MIROC	-9.991	2.100	0.054	0.915	0.157
Lat +Current	-9.645	2.446	0.045	0.830	0.149
Lat +CCSM	-9.602	2.489	0.044	0.875	0.148
Lat+MIROC+Current	-9.429	2.662	0.041	0.915	0.179
Lat+Long+Current	-8.518	3.573	0.026	0.941	0.159

Figure legends

Figure 1. Median-joining network of mitochondrial D-loop haplotypes (411 bp) observed in 1,099 wild boars from Europe, Asia and North Africa. Circle size is proportional to haplotype frequencies, border colour refers to haplogroups (see Fig. 2), fill colour to geographic distribution. Numbers indicate haplotype codes (abbreviated for space limitations, i.e. 29 = H029). Connections with more than one nucleotide change are identified by transversal bars. Major clades (A, E1 and E2) are delimited by dashed lines.

Figure 2. Mitochondrial DNA haplogroups observed in 763 European wild boars. (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 at 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).

Figure 3. Bidimensional plot of the European wild boar populations sampled in this study, obtained by the Spatial Factor Analysis (spFA). Black dots represent population located in Italy and central Europe, grey stars represent populations in Iberia, and black triangles populations in the Balkans and eastern Europe. Numbers indicate population codes.

Figure 4. Estimated distribution of mitochondrial (D-loop) haplotype diversity (*Hk*) in wild boar across Europe, resulting from the interpolation of 38 different sampling populations (Russia

excluded) using the Kriging method in ArcGIS 10. Bluish colours correspond to low values of *Hk*, while reddish colours represent high values.

Figure 5. Distribution range of wild boar in Europe, as predicted on the basis of climatic suitability data estimated for the present time and for the last glacial maximum (LGM). (a) Current predicted distribution, (b) LGM distribution based on the MIROC model, (c) LGM distribution based on the CCSM model.

Figure 1

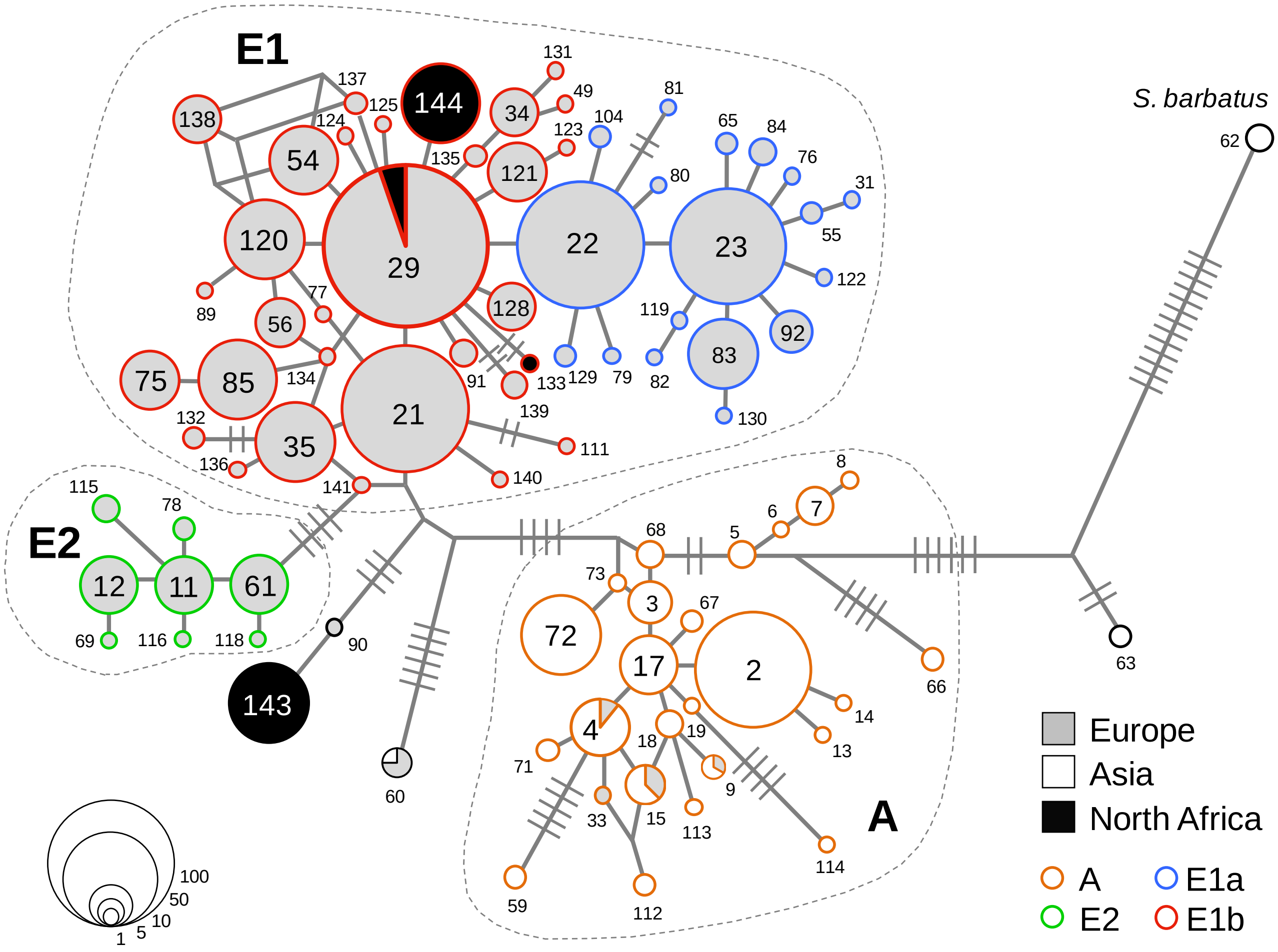
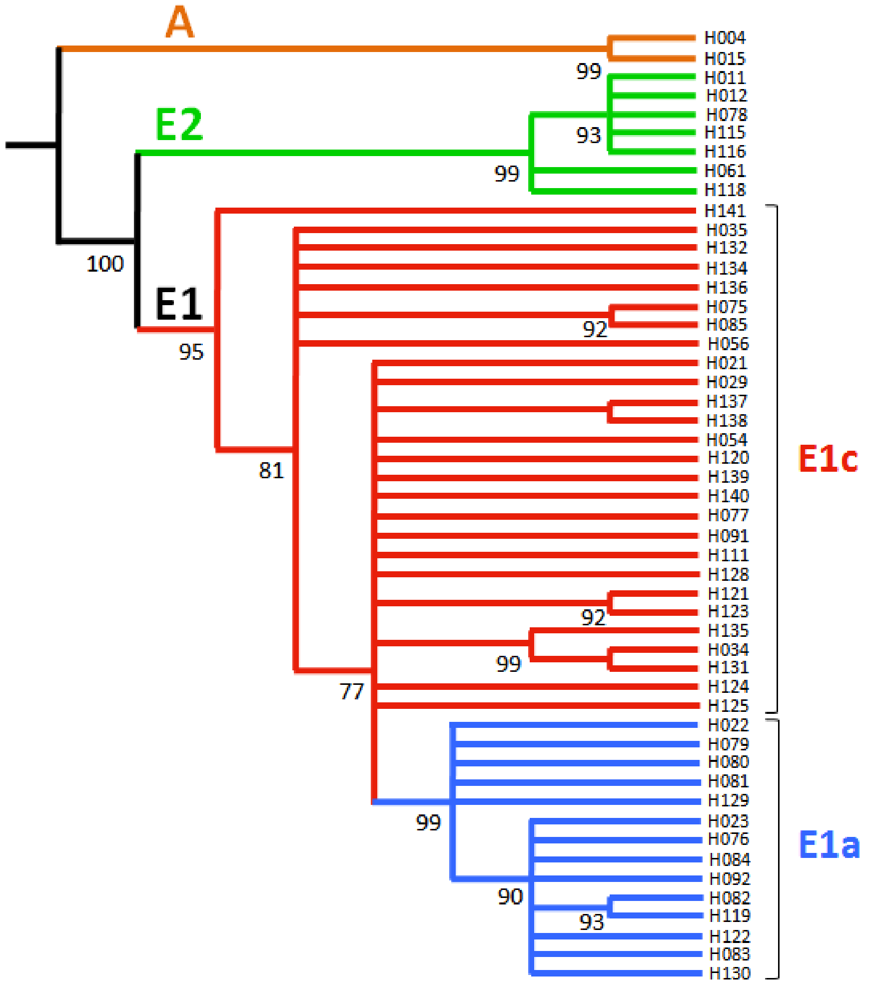

E1
E2
A
S. barbatus
62
63
137
125
144
131
49
34
81
104
123
121
135
65
84
76
31
55
122
138
124
54
120
29
22
23
89
77
56
128
119
92
83
82
79
129
133
91
139
75
85
134
21
132
35
136
141
111
140
130
115
78
12
11
61
69
116
118
90
143
60
8
7
6
5
68
73
3
67
72
17
2
14
13
66
4
18
19
9
71
33
15
113
114
59
112
100
50
10
5
1
Europe
Asia
North Africa
A
E1a
E2
E1b

Figure 2

(a)


A
E2
E1
100
95
81
77
99
93
99
92
92
99
99
90
93
H004
H015
H011
H012
H078
H115
H116
H061
H118
H141
H035
H132
H134
H136
H075
H085
H056
H021
H029
H137
H138
H054
H120
H139
H140
H077
H091
H111
H128
H121
H123
H135
H034
H131
H124
H125
H022
H079
H080
H081
H129
H023
H076
H084
H092
H082
H119
H122
H083
H130
E1c
E1a

(b)

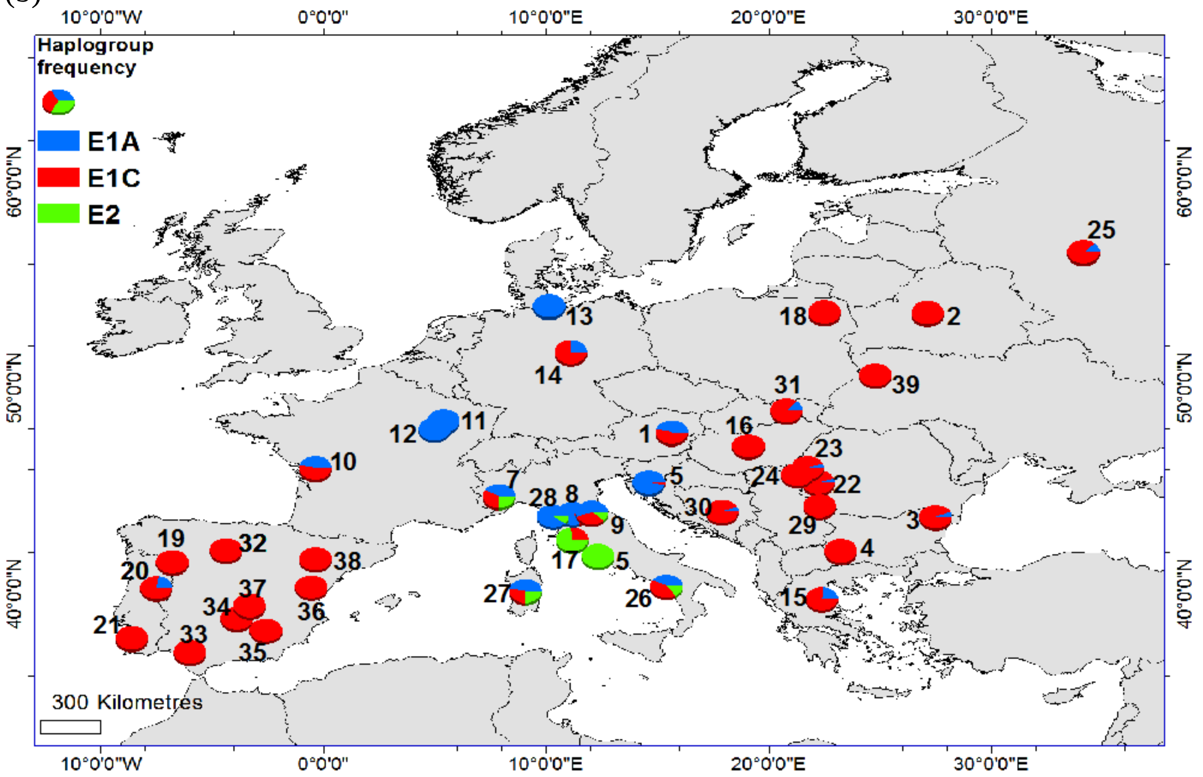

10°0'0"W
0°0'0"
10°0'0"E
20°0'0"E
30°0'0"E
60°0'0"N
50°0'0"N
40°0'0"N
Haplogroup frequency
E1A
E1C
E2
300 Kilometres
1
2
3
4
5
5
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

Figure 3

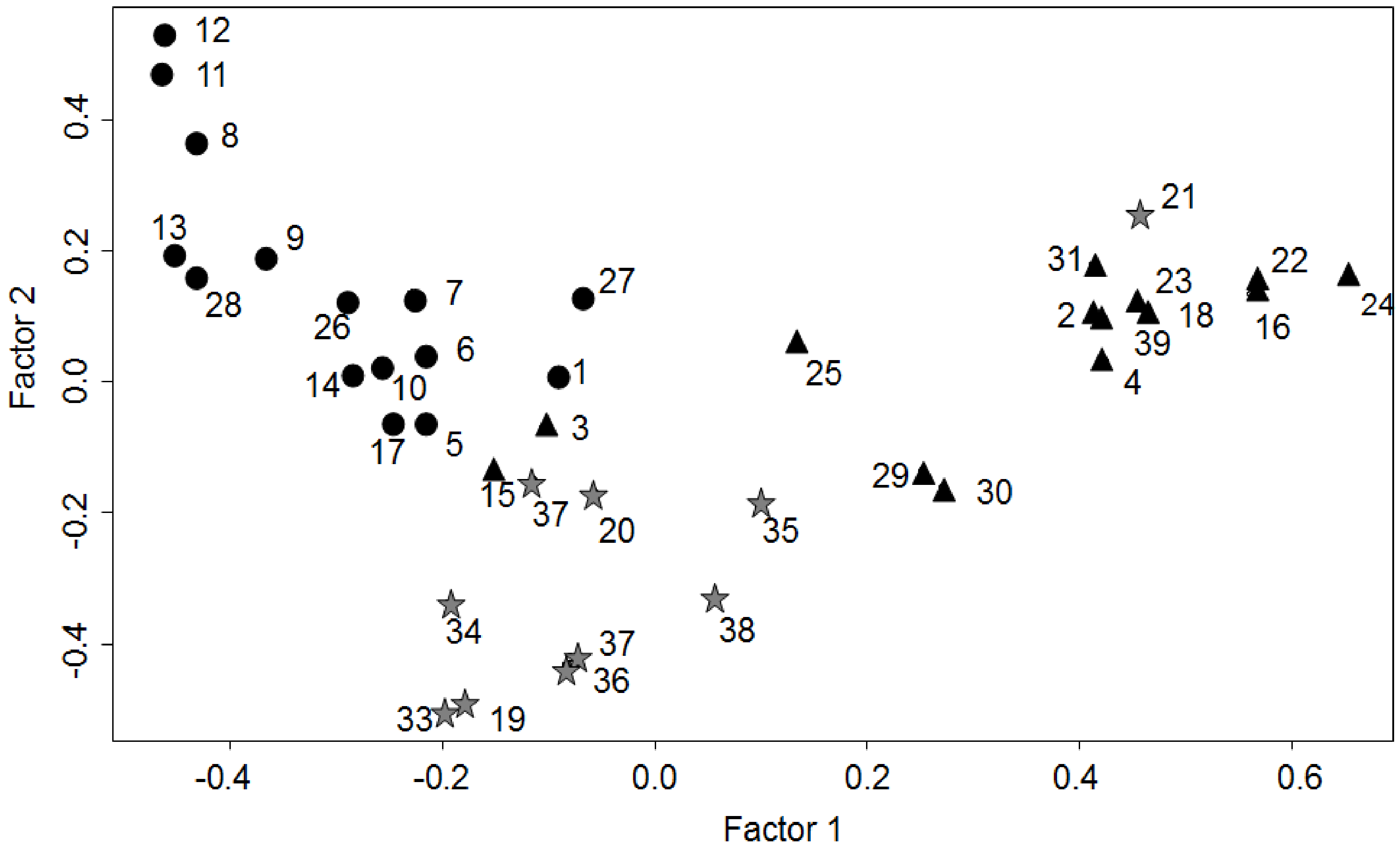

Factor 1
Factor 2
-0.4
-0.2
0.0
0.2
0.4
0.6
-0.4
-0.2
0.0
0.2
0.4
12
11
8
13
9
28
7
27
26
6
14
10
1
17
5
3
15
37
20
35
25
29
30
34
37
36
38
33
19
21
31
22
23
24
2
18
16
39
4

Figure 4

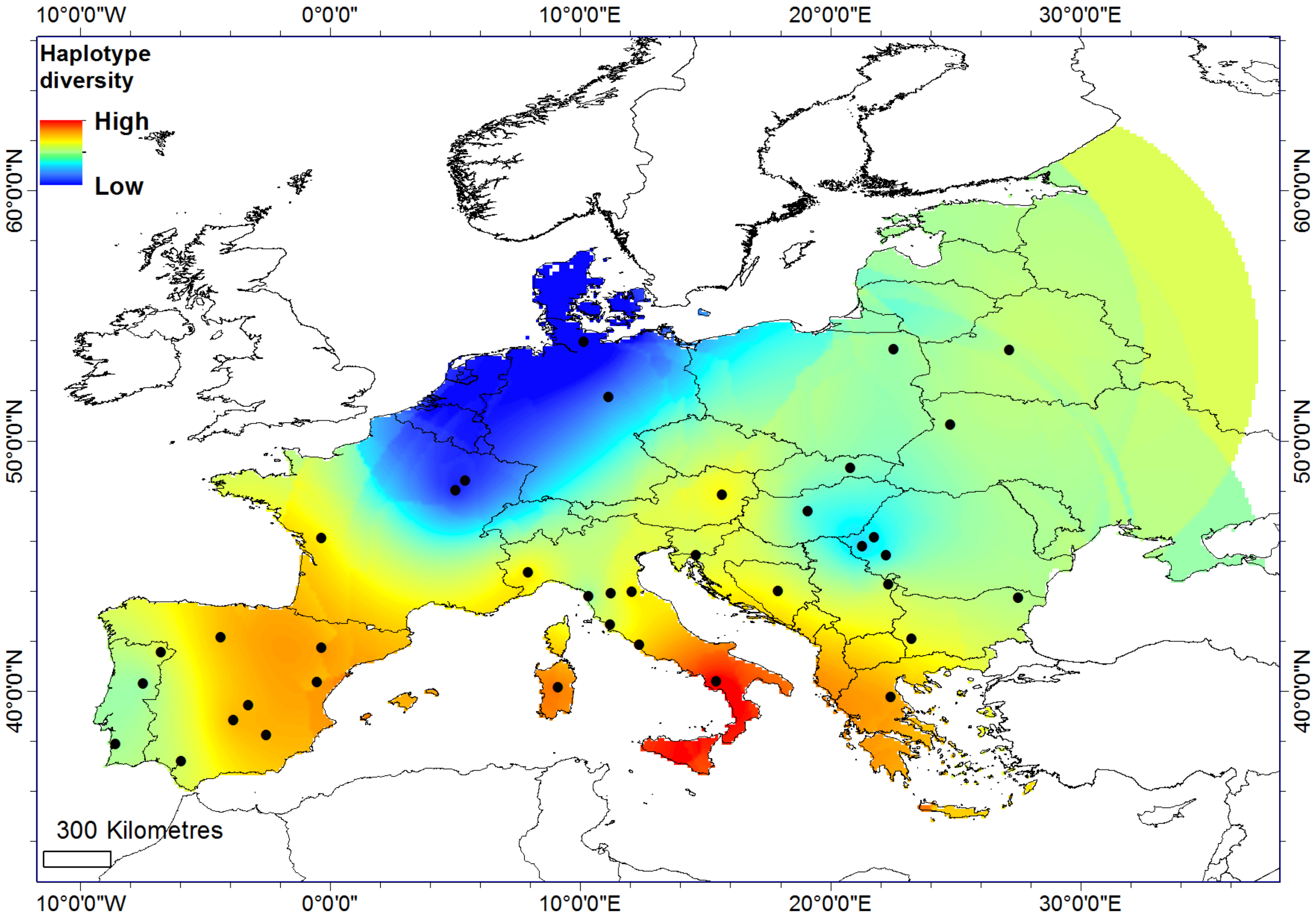

10°0'0"W
0°0'0"
10°0'0"E
20°0'0"E
30°0'0"E
60°0'0"N
50°0'0"N
40°0'0"N
Haplotype diversity
High
Low
300 Kilometres

Figure 5