

ORIGINAL RESEARCH

Maternal genomic variability of the wild boar (*Sus scrofa*) reveals the uniqueness of East-Caucasian and Central Italian populations

Saber Khederzadeh^{1,2,3} | Szilvia Kusza⁴ | Cui-Ping Huang^{1,2,3} | Nickolay Markov⁵ | Massimo Scandura⁶ | Elmar Babaev⁷ | Nikica Šprem⁸ | Ivan V. Seryodkin^{9,10} | Ladislav Paule¹¹ | Ali Esmailizadeh¹²  | Hai-Bing Xie¹ | Ya-Ping Zhang¹ 

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China

³University of Chinese Academy of Sciences, Beijing, China

⁴Animal Genetics Laboratory, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Debrecen, Hungary

⁵Department of Game Animals' Ecology, Institute of Plant and Animal Ecology, Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia

⁶Department of Veterinary Medicine, University of Sassari, Sassari, Italy

⁷Caspian Institute of Biological Resources, Makhachkala, Russia

⁸Department of Fisheries, Beekeeping, Game Management and Special Zoology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

⁹Pacific Geographical Institute FEB RAS, Vladivostok, Russia

¹⁰Far Eastern Federal University, Vladivostok, Russia

¹¹Faculty of Forestry, Technical University in Zvolen, Zvolen, Slovakia

¹²Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

Correspondence

Ya-Ping Zhang and Hai-Bing Xie, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China.

Emails: zhangyp@mail.kiz.ac.cn (Y.-P. Z.); xiehb@mail.kiz.ac.cn (H.-B. X.)

Ali Esmailizadeh, Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman PB 76169-133, Iran.

Email: aliesmaili@uk.ac.ir and

Nickolay Markov, Department of Game Animals' Ecology, Institute of Plant and Animal Ecology, Ural Branch of Russian Academy of Sciences, 8 Marta 202a, 620144 Yekaterinburg, Russia.

Email: nimarkov@mail.ru

Abstract

The phylogeography of the European wild boar was mainly determined by postglacial recolonization patterns from Mediterranean refugia after the last ice age. Here we present the first analysis of SNP polymorphism within the complete mtDNA genome of West Russian ($n = 8$), European ($n = 64$), and North African ($n = 5$) wild boar. Our analyses provided evidence of unique lineages in the East-Caucasian (Dagestan) region and in Central Italy. A phylogenetic analysis revealed that these lineages are basal to the other European mtDNA sequences. We also show close connection between the Western Siberian and Eastern European populations. Also, the North African samples were clustered with the Iberian population. Phylogenetic trees and migration modeling revealed a high proximity of Dagestan sequences to those of Central Italy and suggested possible gene flow between Western Asia and Southern Europe which was not directly related to Northern and Central European lineages. Our results support

Saber Khederzadeh, Szilvia Kusza, and Nickolay Markov contributed equally to the work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Funding information

National Natural Science Foundation of China, Grant/Award Number: 31472000; Chinese Academy of Sciences, Grant/Award Number: XDA08010304-1; Ministry of Agriculture of China, Grant/Award Number: 2016ZX08009003-006-005; the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences

the presence of old maternal lineages in two Southern glacial refugia (i.e., Caucasus and the Italian peninsula), as a legacy of an ancient wave of colonization of Southern Europe from an Eastern origin.

KEYWORDS

Caucasus, migration modeling, phylogeny, *Sus scrofa*, whole mtDNA

1 | INTRODUCTION

Wild boar (*Sus scrofa*) is widely distributed all over the world, and their population is almost four million, hence, considered the second most abundant ungulate species in Europe and a pest in several regions (Apollonio, Andersen, & Putman, 2010; Bobek, 2018). However, some authors have attributed their range expansion to climate change (mild winters with no or low snow cover, triggering the necessity for higher food intake; Jędrzejewska, Jędrzejewski, Bunevich, Miłkowski, & Krasieński, 1997; Root et al., 2003; Vetter, Ruf, Bieber, & Arnold, 2015). Others argue that this population increase has been caused by intensive crop production, changes in agricultural practices, reduced number of predator species, declined hunting pressure, human activities such as supplementary feeding, intentional releases for hunting purposes, and hybridization with domestic pigs (Apollonio, Scandura, & Šprem, 2014; Baskin & Danell, 2003; Bieber & Ruf, 2005; Geisser & Reyer, 2004, 2006; Iacolina, Corlatti, Buzan, Safner, & Šprem, 2019; Jędrzejewska et al., 1997; Leaper, Masei, Gorman, & Aspinall, 1999; Masei & Genov, 2004; Masei et al., 2015; Sáez-Royuela & Tellería, 1986).

Genetic studies can be used to discover more diversity patterns as well as in explaining the changes observed in the species' geographical range. Throughout its history, the wild boar has been strongly influenced by human practices such as hunting, pig domestication, and animal translocation (Larson et al., 2005; Scandura, Iacolina, & Apollonio, 2011). Thus, analysis of the changes in the geographical range of *S. scrofa* could aid our understanding of not only the global patterns of evolution and ecosystem change (Anijalg et al., 2018). Previous studies revealed the complex genetic structure of wild boar populations in Eurasia, including multiple domestication events and gene flow between wild boar and domestic pig breeds (Iacolina et al., 2016; Larson et al., 2005; Ramírez et al., 2009; Ribani et al., 2019; Šprem et al., 2014).

Since the most of these studies had been focused on Europe or Eastern Asia (Kusza et al., 2014; Larson et al., 2005; Scandura et al., 2011; Veličković et al., 2015; Vilaça et al., 2014), the specific status and phylogenetic position of West Asian pigs have only recently been reported. A recent investigation by Khalilzadeh et al. (2016) revealed the presence of Middle Eastern, European, and East Asian haplotypes in Iranian wild boar and proposed that there had been contact between the European and East Asian wild boar populations. Thus, analysis of samples from a large number of geographical localities, across different regions, could provide a more comprehensive description of the evolutionary history of wild boar.

A factor that could affect the quality of analysis is the type of genetic marker used in any study. Most genetic studies on the diversity or phylogenetics of wild boar were based on partial D-loop sequences of mitochondrial DNA (mtDNA; less than 7% of the whole mtDNA genome), or cytochrome b sequences (Fang & Andersson, 2006; Kusza et al., 2014; Larson et al., 2005; Ramírez et al., 2009; Scandura et al., 2008; Veličković et al., 2015, 2016; Vilaça et al., 2014), which occupies less than 7% of the whole mtDNA genome, sometimes in combination with another region (e.g., cytochrome b). To the best of our knowledge, there are only a few previous examples of using a complete mtDNA genome to assess the phylogeographical relationships of *S. scrofa*. Ni et al. (2018) addressed their global phylogeography, based on sequences data from domestic pigs (72%), and a comparatively low number of wild boar samples. The authors showed patterns similar to those obtained by use of partial mtDNA genome sequences, thus demonstrating a clear separation of European and Asian clades. Chen et al. (2018) addressed the relationships between North Asian and South Asian wild boars, based on the effect population size (N_e) and climate on the non-synonymous/synonymous mutation ratio (K_a/K_s). Thus, there is a gap in genetic studies of wild boar, due to the lack of complete mtDNA on the Western part of the species' geographical range.

In this study, we present an analysis of SNPs variability in the whole mtDNA genome of wild boar from North Africa and continental Eurasia (various regions from Western Europe to Western Siberia). Firstly, we aimed to determine the phylogenetic position of animals from Eastern Caucasus—a region not previously included in any phylogeographical study. Secondly, we addressed the question of how a phylogenetic tree of wild boar, based on SNPs of whole mtDNA genome data, would differ from a phylogenetic tree based on shorter fragments (D-loop and cytochrome b markers).

2 | MATERIALS AND METHODS**2.1 | Ethics statement**

We declare that we have no financial or personal relationships with other people or organizations that can inappropriately influence our work, and that we have no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing our research.

No wild boars were culled solely for the purpose of the present study. Tissue samples from each different country were obtained from collaborators and hunters. All samples were collected

in compliance with each county's national regulations on wild boar management plans.

2.2 | Material

A total of 77 samples (fresh muscle $n = 72$, blood $n = 5$) were collected in 2014–2017 in Europe (Germany $n = 7$, Poland $n = 7$, Italy $n = 6$, Bulgaria $n = 5$, Hungary $n = 5$, Ukraine $n = 5$, Netherlands $n = 4$, Croatia $n = 4$, Estonia $n = 4$, Romania $n = 4$, Slovakia $n = 4$, Czech Republic $n = 4$, Spain $n = 3$, France $n = 1$ and Switzerland $n = 1$), Asian (Western Siberia $n = 4$), and East-Caucasus (Dagestan $n = 4$) parts of Russia and North Africa (Morocco $n = 4$ and Tunisia $n = 1$).

Tissue samples were stored in plastic tubes (5–30 ml) filled with 96% ethanol. Blood samples were kept frozen in EDTA tubes at -20°C until further analysis. Laboratory experiments were performed at the laboratory of Molecular Evolution and Genome Diversity of State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences. Available whole mtDNA sequence data of for five suids warthog (*Phacochoerus africanus*), Javan warty pig (*Sus verrucosus*), Visayan warty pig (*Sus cebifrons*), Celebes warty pig (*Sus celebensis*), and Bornean bearded pig (*Sus barbatus*) were downloaded, to be used as outgroups, from the NCBI database (accession numbers: ERR173209, ERR173210, ERR173177, ERR173211, ERR173203). These five genomes were combined with the 77 mtDNA genomes obtained in this study for further statistical analyses.

2.3 | Laboratory methods

In this research, some samples were extracted with QIAamp DNA Blood Spin kits (Qiagen), and most of which were extracted using the phenol–chloroform method (Sambrook, Fritsch, & Maniatis, 1989). The quality and quantity of the extracted DNA were checked using a NanoDropTM 8000 spectrophotometer (NanoDrop Technologies Inc.). All experiments were performed according to the ethical regulations of the Chinese Academy of Sciences (Approval ID: SYDW-2015012).

2.4 | Sequence analysis

Library preparation for the Illumina sequencing platform required fragmentation of our DNA (1–3 μg of genomic DNA), followed by repair of 3' and 5' ends to form blunt-ended, phosphorylated molecules, and the addition of a non-templated dA-tail before ligation to an adaptor. DNA libraries were prepared according to the standard Illumina library preparation protocol, with a short insert size range of 300–500 bp, and then sequenced on an Illumina HiSeq 2000 platform with the 150 bp paired-end sequencing kits. Pair-end reads (150 bp) were sequenced to about $10\times$ sequencing depth and $\geq 99\%$ coverage for each individual.

Quality control procedure was used to remove reads with low sequencing quality. Particularly, reads were trimmed for minimum Phred quality >20 over three consecutive base pairs and discarded if shorter than 45 bp. Clean reads were trimmed from raw reads that

were preprocessed to remove index adaptors and low-quality reads. Quality control for removing the low-quality reads was done based on the following criteria: up to 10% of the read bases include “N” content of each sequenced reads, up to 50% of the read bases include low-quality ($Q \leq 5$) base content in any sequenced reads, and finally, by removing duplicate reads, using Picard tools v.2.12.1.

After quality trimming, clean reads of each sample were aligned against the mitochondrial genome from the domestic pig reference genome (*S. scrofa* 10.2, which was downloaded from Ensembl genome browser) using Burrows-Wheeler Aligner (BWA) software (Li & Durbin, 2010). After extracting whole mitochondrial genome from our raw data by VCFtools 1.13 (Danecek et al., 2011), variants were called using Genome Analysis Toolkit (GATK) (Nekrutenko & Taylor, 2012). To avoid potential bias between our extracted data and the publicly available data, we called SNPs by comparing the mitogenome sequence of each individual to the mitochondrial reference genome and then merged the called SNPs to form a common set of SNP data for the 82 individuals (including 77 samples of wild boar and five sequences for outgroups). Then, several filtering steps were applied before using candidate SNPs for further analyses to minimize the number of false positive calls. Then, we removed all of the missing sites and also SNPs with minor allele frequency of 0.01, leaving 933 SNPs for next step. Afterward, the Integrative Genomics Viewer 2.5 (IGV) (Robinson, Thorvaldsdóttir, Wenger, Zehir, & Mesirov, 2017) was used as a high-performance visualization tool for interactive exploration of large, integrated genomic datasets and to confirm reliability of high-quality polymorphic sites (269 SNPs) with no any linkage makers based on reference genome for further analyses.

2.5 | Statistical analyses

Single nucleotide polymorphism sequences from the whole mtDNA were aligned using the Clustal W software (Thompson, Gibson, Plewniak, Jeanmougin, & Higgins, 1997) implemented in MEGA v.6 software (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Levels of molecular diversity such as haplotype number (h), haplotype diversity (H_d), nucleotide diversity (π), Fu's F_s , and Tajima's D were independently computed with DnaSP v.5.10 (Librado & Rozas, 2009) for 77 samples of *S. scrofa*.

The Bayesian phylogenetic tree was constructed using 78 sequences (adding 1 sequence of warthog [*P. africanus*] as an outgroup) using MrBayes 3.2.2 software (Ronquist & Huelsenbeck, 2003). The evolutionary parameters were given by jModeltest2.1 (Posada, 2008). Two independent Monte Carlo Markov Chains (MCMC) were run for 50,000,000 generations, with sampling every 100 generations and discarding the first 10% as burn-in. The resulting trees were visualized using FigTree 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Further phylogenetic analyses were performed using neighbor-joining method, based on the genetic distance matrix in MEGA v.6 software (Tamura et al., 2013). Maximum likelihood tree (ML) was constructed with 1,000 bootstrap replicates using TREE-MIX v.1.1 when the migrations were prohibited. For both phylogenetic trees (NJ & ML), sequences of five suids were used as outgroups ($n = 82$).

For a large number of haplotypes, the complexity of a network can be solved with a median-joining (MJ) network algorithm. So, based on all haplotypes of the complete mtDNA, a MJ network was constructed using NETWORK, version 5.0.0.3 (Bandelt, Forster, & Röhl, 1999). We also performed a principal component analysis (PCA) with population-scale SNPs, as implemented in the smartPCA program in the EIGENSOFT package v.6.1.4 (Galinsky et al., 2016; Patterson, Price, & Reich, 2006; Price et al., 2006).

TREEMIX v.1.1 (Pickrell & Pritchard, 2012) was used to model demographic scenarios in the form of a bifurcating tree based on the maternal genome, allowing for inferring migration events between individuals and populations to provide insights into hidden demographic events of the past and testing for the effect and presence of gene flow between divergent populations. TREEMIX software first builds a maximum likelihood tree of sampled populations by approximation of genetic drift. Subsequently, it identifies populations whose genetic covariance is underestimated by the model. Three migration events were selected on the basis of a different number of bootstraps to improve fitness of the model. To compensate for linkage disequilibrium, we applied strict filters to exclude all positions with genetic linkage. As the use of our dataset would require input file to be in to Ped/Map formats, we avoided any interference of linked markers by applying --indep-pairwise parameter in PLINK v.1.07 (Purcell et al., 2007) for windowed pruning of the data to obtain a pruned file which we then used to get second Ped/Map files, the final input for subsequent analyses. Therefore, our analysis was based on non-link markers. Together, the nature of our dataset, Biallelic SNP VCF of the whole mtDNA, and data filtering steps made it possible for us to reliably use TREEMIX.

3 | RESULTS

3.1 | Mitogenome variability of wild boars

A total of 77 complete wild boar mtDNA genomes from Europe ($n = 64$), North Africa ($n = 5$), and Russia ($n = 8$) were sequenced for this study. After the quality control steps, 269 polymorphic sites were retrieved for the analyses. A total of 33 haplotypes were detected in our dataset, of which 27, 4 and 3 were found in Europe, North Africa and Russia, respectively (Figure 1, Table 1). Seventy-nine percent of haplotypes were found in just one region, namely Europe. The most widely distributed haplotype was Hap12, which was found in Western, Central, and Eastern Europe. The proportion of unique haplotypes was 0.54. The haplotypes shared by the highest number of individuals were Hap 19, found in 10 animals in Central and Eastern Europe and Hap9 found in 7 animals in Western and Central Europe. The haplotype diversity was generally high (over 0.9) in North Africa and Europe, while for Western Siberia and Eastern Caucasus it was relatively low (Table 1). The highest values of both haplotype and nucleotide diversity were found in Western Europe. We did not detect a significant departure from the neutrality expectation, as calculated by both Fu's F_S and Tajima's D (Table 1).

3.2 | Phylogenetic analysis

Phylogenetic (NJ, Maximum likelihood, Bayesian) trees and median-joining networks (MJ) were constructed using SNPs of sequenced mtDNA genomes from the present study and mitochondrial genomes of five *Suidae* species across the World. Both NJ, ML (Appendix 1) and Bayesian (Figure 2) trees showed similar topologies. In the

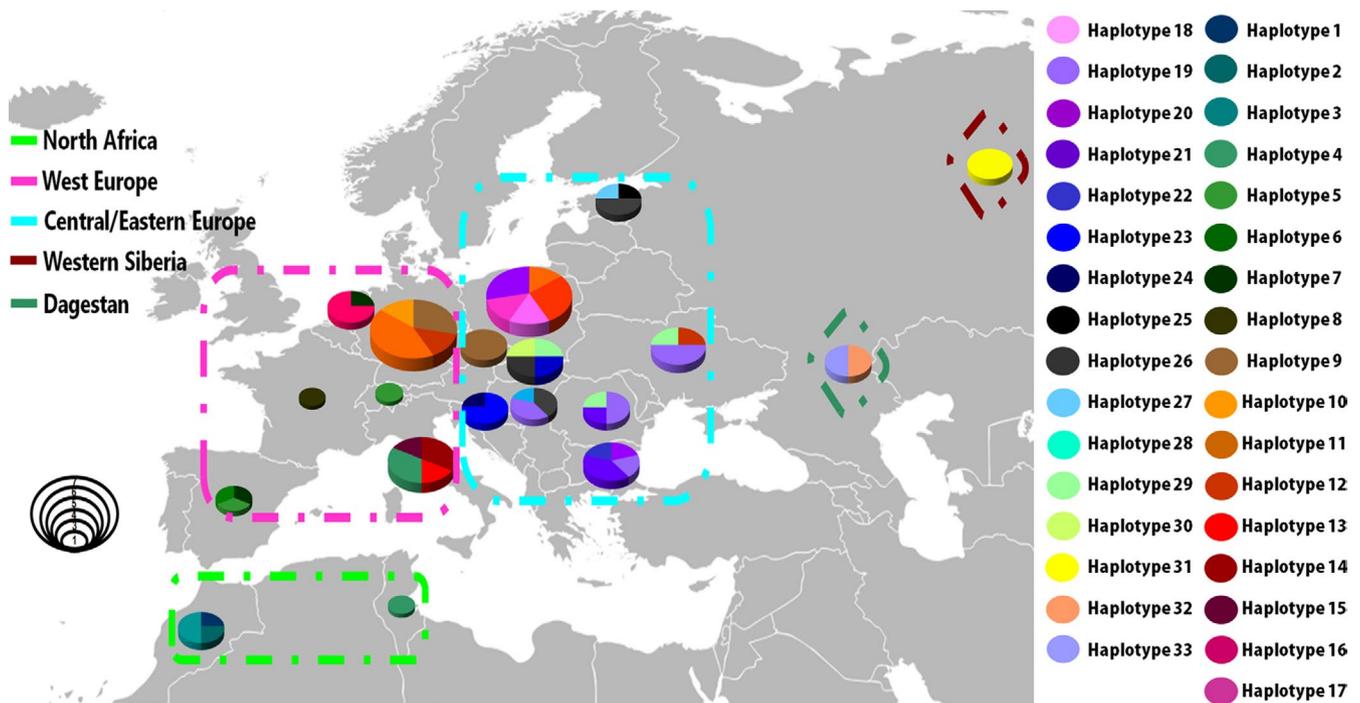


FIGURE 1 Geographical distribution of the sequenced haplotypes. Areas of circles on the map are proportional to the number of sampled individuals. Dotted lines represent geographical regions

TABLE 1 Basic parameters for genetic (mtDNA) variability within wild boar populations

| Populations | Number of samples | Number of different haplotypes | Detected haplotypes (Hap) | Haplotype diversity \pm SD | Variance in haplotype diversity | Nucleotide diversity \pm SD | Tajima's D | Fu's Fs | Number of polymorphic sites |
|-------------------------------------|-------------------|--------------------------------|---------------------------|------------------------------|---------------------------------|-------------------------------|----------------------|----------------------|-----------------------------|
| North Africa | 5 | 4 | Hap1/Hap2/Hap3/Hap4 | 0.900 \pm 0.161 | 0.025 | 0.002 \pm 0.000 | -0.894 ^{ns} | 0.212 ^{ns} | 10 |
| West Europe | 22 | 13 | Hap4-16 | 0.944 \pm 0.026 | 0 | 0.015 \pm 0.004 | -0.843 ^{ns} | 3.987 ^{ns} | 137 |
| Central and Eastern Europe | 42 | 17 | Hap9/Hap11/Hap12/Hap17-30 | 0.912 \pm 0.026 | 0 | 0.004 \pm 0.000 | -1.491 ^{ns} | -0.931 ^{ns} | 58 |
| Western Siberia (Russia) | 4 | 1 | Hap31 | N/a | N/a | N/a | 0 | N/a | N/a |
| Eastern Caucasus (Dagestan, Russia) | 4 | 2 | Hap32/Hap33 | 0.667 \pm 0.204 | 0.041 | 0.001 \pm 0.000 | 2.080 ^{ns} | 2.719 ^{ns} | 4 |

Note: The details of five main populations are in italics font.

Bayesian tree, the samples from Dagestan and some samples from Central Italy (Italy 1-3, Hap13,14) were rooted directly in the basal branch and clearly separated from all other populations. All the other samples formed a single cluster where all haplotypes originated from the same root. West Siberian samples showed a phylogenetic position close to Eastern European (Ukrainian, Romanian) populations, while North African samples were close to the Spanish population.

The MJ network (Figure 3) also shows the distal position of the East-Caucasian and Central Italian samples to other haplotypes. This analysis also revealed the geographical structure within the group of European and North African samples. They form two star-shaped structures. One consists of haplotypes found in Western Europe and Northern Africa, while the other includes haplotypes from Central and Eastern Europe and from Western Siberia. The "central" haplotypes in the Western European and the Eastern European clusters are Haplotype 4 (found in Italy and Switzerland) and Haplotype 19 (found in Romania, Poland, Hungary, and Ukraine).

3.3 | Principal component analysis and gene flow

In the principal component analysis, the variation along the first principal component (PC1) explained 51.62% of the variance and separated the East-Caucasian samples from all the other geographical regions. The variation along the second principal component (PC2), explained 22.93% of the variance and showed the specificity of the Italian (Latium) population in comparison with all the other samples (Figure 4a). If the PCA was run excluding all Dagestan and samples from Latium, Italy, it did not demonstrate any clear geographical structure (Figure 4b), except the distal position of French samples to the position of other European samples.

We also studied population splits and gene flow between wild boar populations. Using TREEMIX, we firstly constructed a phylogenetic tree where no migration event was allowed. Here, outgroup and Dagestan were separated from other populations. Then, up to ten migration events were added to the tree. This phylogenetic tree showed that the Dagestan population was clearly separated from other populations (Figure 5). TREEMIX runs showed a high genetic affinity between Dagestan samples and the Italian samples (according to the model 79.14% of alleles in Central Italian clade were descended from Dagestan clade) and evidence of possible very weak gene flow not only between Dagestan and Western Siberia, but also between Dagestan and Germany (3.1% and 3%, respectively).

4 | DISCUSSION

Most published studies of the genetic diversity and phylogenetic relationships among wild boar mtDNA haplogroup lineages have been performed using the D-loop or the cytochrome b regions of the mtDNA genome (Alexander, Novembre, & Lange, 2009; Kusza et al., 2014; Larson et al., 2005; Scandura et al., 2008; Veličković et al., 2015). Wild boars have an intermediate level of mtDNA D-loop variability (Djan et al., 2013), and little information is available

FIGURE 2 Phylogenetic relationships of the studies samples basing on the Bayesian phylogenetic tree. Posterior probabilities ≥ 0.8 are given. The numbers on the branches are posterior probabilities from the Bayesian inference. Different colored lines represent the clusters of populations geographically close to each other (green indicates North Africa; pink indicates West Europe; light blue indicates Central Europe and Eastern Europe; dark red indicates Western Siberia; dark green indicates Dagestan)

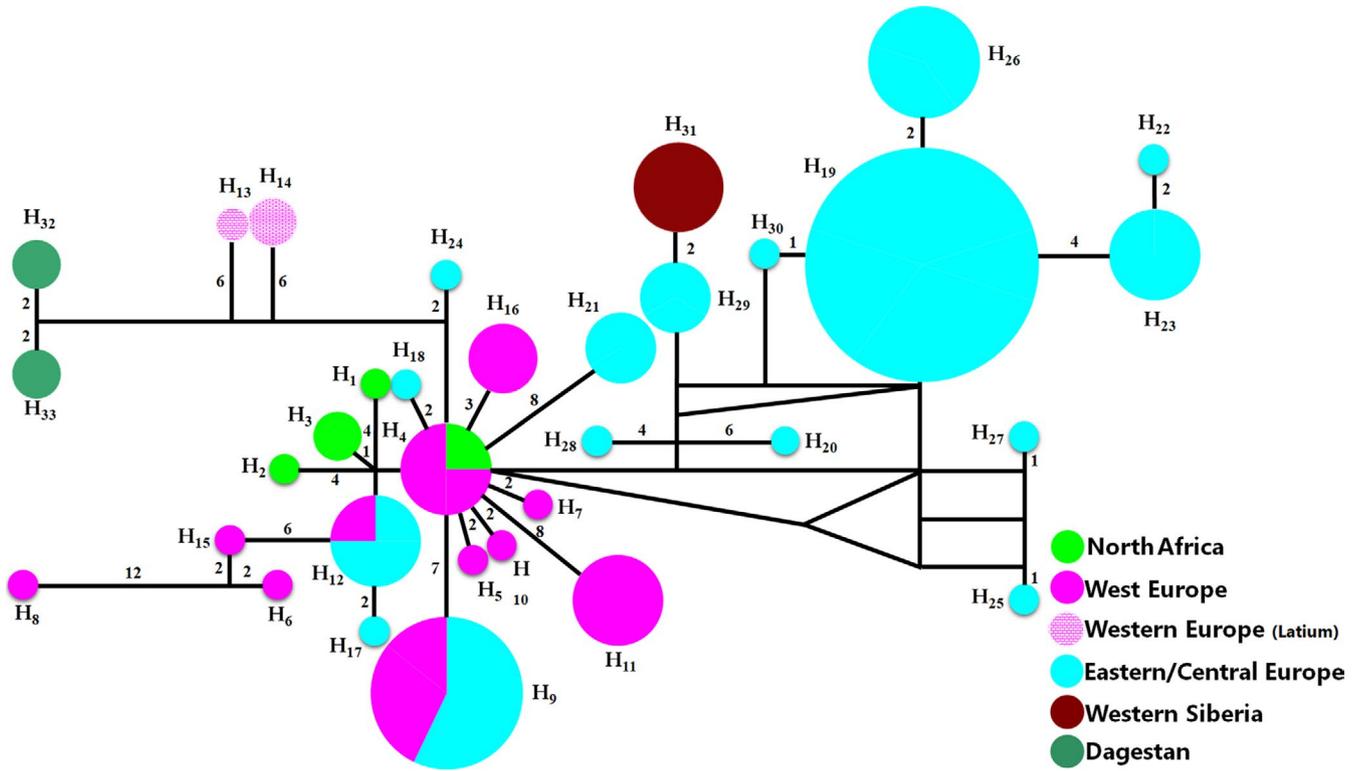


FIGURE 3 Median-joining networks based on SNPs of complete mtDNA. The network was constructed with 77 wild boar sequences. Mutations are represented by the numbers in which they occurred. Geographical locations of samples are represented by circles with color, with sizes that are in proportion to the number of sampled individuals

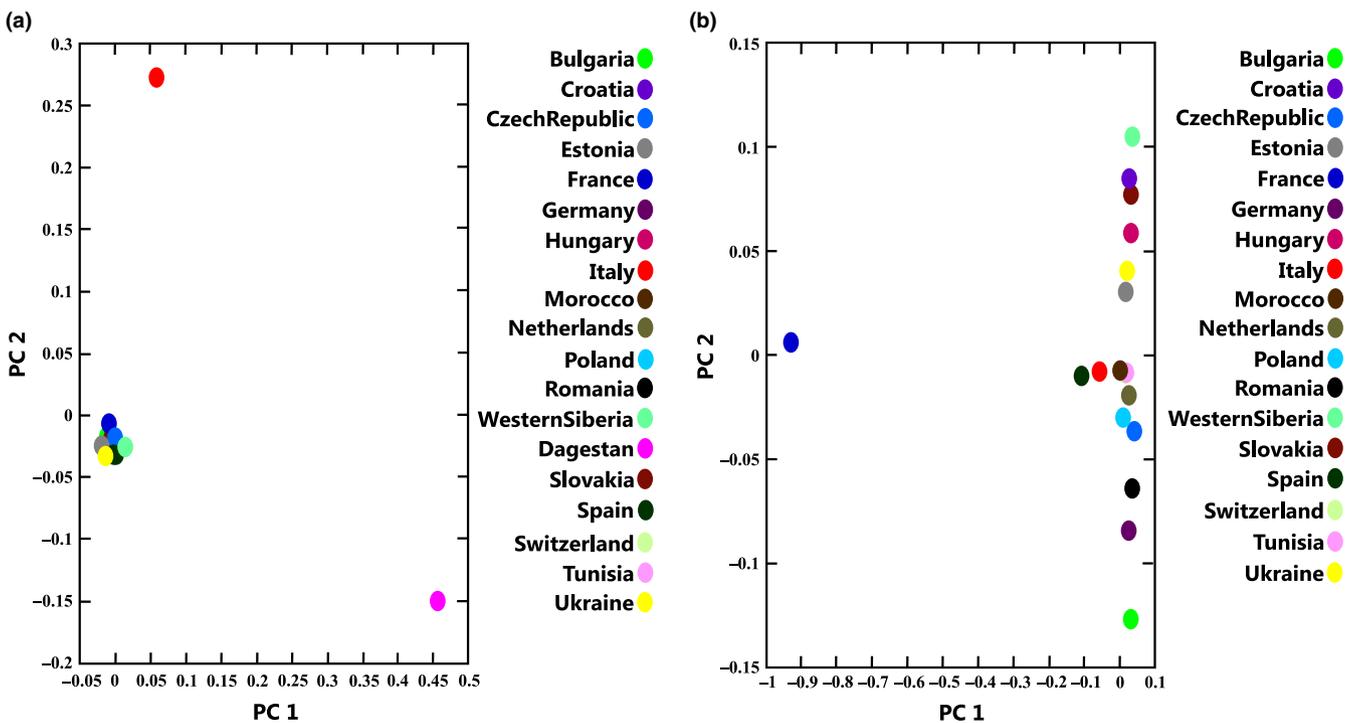


FIGURE 4 PCA for different subpopulations. (a) PCA result for 77 samples belonging to 19 subpopulations. (b) PCA result for 70 samples belonging to 18 subpopulations (4 samples from Dagestan and 3 from Italy were excluded). PC1, principal component 1; PC2, principal component 2

about their complete mtDNA genome diversity (Chen et al., 2018; Kijas & Andersson, 2001; Ni et al., 2018; Tan et al., 2018). However, the analysis of whole mtDNA genome could lead to new and important inferences about the origin and diversification of *S. scrofa* populations. Here, we studied the SNPs of complete mtDNA genome and variability of wild boars from West Siberia, Eastern Caucasus, Europe, and North Africa to provide new data on the species' genetic diversity and phylogenetic patterns. The results of our study revealed the peculiar haplotype composition of the wild boar inhabiting two specific areas: East-Caucasus (Dagestan) and Central Italy. For the first time, we have detected unique mtDNA haplotypes from Dagestan that could probably appear early in the wild boar phylogeny. The genetic uniqueness of Italian wild boars has already been shown in previous studies (Larson et al., 2005; Scandura et al., 2008), indicating that the Italian peninsula has served as a Southern refugia during the Pleistocene glaciations, whereas the Alps may have acted as a barrier to gene flow (Scandura et al., 2008; Veličković et al., 2015; Vilaça et al., 2014). Most interesting was the similarity of Central Italian and Caucasian haplotypes (as shown by TREEmix analysis, Figure 5). A similar trend involving the proximity of Caucasian and Mediterranean haplotypes has also been reported for common hamster (*Cricetus cricetus*), by Feoktistova et al. (2017). This could explain the presence of two genetic lineages in the territory of modern Italy, as shown by a number of previous studies (Larson et al., 2005; Scandura et al., 2008). High genetic diversity in Southern Europe, particularly in Italy, has been explained by the preservation of lineages during glaciation in several Pleistocene refugia. Another genetically distinct lineage (commonly known as E1 and found over the whole of Europe) could have arrived to Italy from Northern Europe during the contraction of range during the period of glaciation.

The hypothesis of migration of West Asian haplotypes to Europe and particularly to Italy is supported by the results of other researchers. Particularly, Alexander et al. (2009) have reported Near Eastern haplotype on the island Samos, while Veličković et al. (2015) have also reported Asian haplotype from the South of Balkan Peninsula.

According to Maselli et al. (2016), the proportion of "Asian" haplotypes in Southern Italy and Sardinia was about 9%. Manunza et al. (2013) based on the analysis of autosomal SNPs also suggested the scenario of migration from Trans-Caucasus region and Western Asia to Western Europe.

The lack of distinct clades (similar to E1 and E2) on the Balkan Peninsula (Alexander et al., 2009; Veličković et al., 2015) could be treated as contradicting the hypothesis of wild boar expansion from West Asia to Southern Europe through Asia Minor. Thus, the South Balkans haplogroup may represent a remnant of the pre-LGM gene pool of the Balkans, in the same way as the European E2 haplogroup is a remnant of pre-LGM diversity in Italy (Veličković et al., 2015). However, it is possible that on the Balkans the haplotypes that arrived from West Asia could be replaced by haplotypes described as the E1 group. In contrast to the Balkans, Italy is separated from Northern Europe by the Alps, which extend in an east-west direction. This could have reduced the expansion of northern wild boars to the Apennine peninsula and have helped preservation of lineages, which hypothetically could arrive from the South.

The genetic proximity of Caucasian and South European samples could also be related to direct translocations of pigs from West Asia to Southern Europe. Maselli et al. (2016) suggested that the Near Eastern haplotypes found in the ancient pigs of Italy could have descended from early domesticated pigs that arrived into Europe during the Neolithization of the continent or from a legacy of previous contacts with some northern populations. However, Vai et al. (2015) suggested reconsidering this hypothesis. Based on the analysis of ancient samples, they showed that the E2 clade was present in Italy before the NE haplotypes arrived to Southern Europe, and the NE haplotypes in turn were dated to the pre-Neolithic period, thus prior to domestication. Further studies of central Asian wild boar are needed to clarify this subject. It will also be important to study the genetic relationships between Caucasian and East Asian wild boars.

Weak genetic drift from Caucasus to Western Europe (Germany) indicates that Dagestan clade did not contribute significantly to

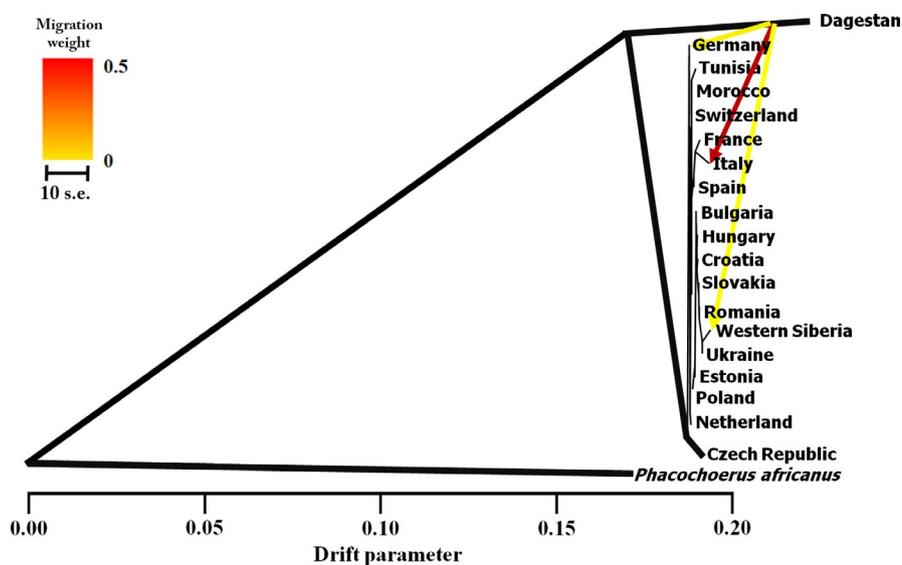


FIGURE 5 Maximum likelihood tree depicting the genetic relationships between 19 *Sus scrofa* populations. Results were based on TREEmix for all populations, allowing for three migration events. The weight of the migration component follows the key on the left. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix

North and West European genetic lineages. This result supports the hypothesis of expansion of West Asian haplotypes to Southern Europe through Asia Minor, but not through Northern Europe.

The genetic proximity of West Siberian wild boars to European groups can be accounted by the fact that West Siberian population was established as a result of intentional releases of animals presumably from the European part of Russia, Eastern Europe (Belarus), and the Northern Caucasus (Markov & Bolshakov, 1996). Our results indicate that European lineages are more common in newly established populations of Western Siberia. These results are consistent with those obtained by Zinovieva (2013) using polymorphism of genes associated with quantitative trait loci. They showed that West Siberian animals grouped with European and Caucasian wild boar rather than with populations from the Trans-Baikal region and the Far East part of Russia.

The topology of the Bayesian tree based on whole mtDNA genomes did not differ from trees created using the partial sequences on control region and cytochrome b, presented in previously published articles (Fang & Andersson, 2006; Kusza et al., 2014; Larson et al., 2005; Veličković et al., 2015). As in previous studies (e.g., Fang & Andersson, 2006; Larson et al., 2005; Soria-Boix, Donat-Torres, & Urios, 2017), this showed a lack of genetically specific groups within continental Western, Central, Eastern Europe, and Northern Africa (Morocco and Tunisia) and the specificity of the Italian population. On the other hand, the MJ network based on the whole mtDNA sequences suggests differences in haplotype composition between Western Europe and Central-Eastern Europe. The division of samples on the MJ network is however in general agreement with clusters E1-A and E-1C described in Scandura et al. (2011). Clearly, there was no isolation between Western and Central-Eastern Europe wild boar populations, but the proportion of different haplotypes in these two regions and star-shaped MJ networks raises the question of possible differences in origin of wild boars from different parts of the European continent.

We conclude that inclusion of the data from a previously unsampled region (Eastern Caucasus), using the complete mitochondrial genome as a basis for phylogeographical analysis and modeling animals' migrations, allowed us to confirm diversification of populations based on partial sequences of mitochondrial genome and hypothesize about possible ways of expansion in the western part of the *S. scrofa* geographical range. Further analysis of data collected over a larger area, and inclusion of ancient samples is needed to check these findings.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (31472000), Chinese Academy of Sciences (XDA08010304-1), the Ministry of Agriculture of China (2016ZX08009003-006-005) and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding). The collection of samples and data analysis were partly performed within the

framework of state contract with the Institute of Plant and Animal Ecology, Ural Branch Russian Academy of Sciences. The authors would like to acknowledge the support of Dr. Marcel Amills from Department of Animal Genetics and Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB) of Campus de la Universitat Autònoma de Barcelona in Spain and the colleagues Antonio Canu and Barbara Franzetti from Italy and to other institutions for helping us with sampling. We would heartily like to thank Dr. Fiona Worthy from York University, UK, who deeply revised and proofread our manuscript and also to the author of photograph Dr. Tibor Pataky.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

The study was designed by YPZ, AE, SK, NM. SK, and CPH performed the molecular analysis. SK and Sz.K performed the population genetic analyses. SK, NM, MS, EB, NS, IVS, LP, AE, and HBX provided the data. NM, Sz.K, and SK wrote the paper and all authors read, corrected and approved the manuscript.

DATA AVAILABILITY STATEMENT

Our novel data has an accession number of PRJEB30825 and is available in the European Bioinformatics Institute (EMBL-EBI).

ORCID

Ali Esmailizadeh  <https://orcid.org/0000-0003-0986-6639>

Ya-Ping Zhang  <https://orcid.org/0000-0002-5401-1114>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Anijal, P., Ho, S. Y. W., Davison, J., Keis, M., Tammeleht, E., Bobowik, K., ... Saarma, U. (2018). Large scale migrations of brown bears in Eurasia and to North America during the Late Pleistocene. *Journal of Biogeography*, 45, 394–405. <https://doi.org/10.1111/jbi.13126>
- Apollonio, M., Andersen, R., & Putman, R. (2010). *European ungulates and their management in the 21st century* (p. 618). New York, NY: Cambridge University Press.
- Apollonio, M., Scandura, M., & Šprem, N. (2014). Reintroductions as a management tool for European ungulates. In R. Putman, & M. Apollonio (Eds.), *Behavior and management of European ungulates* (pp. 46–77). Dunbeath, Caithness, UK: Whittles Publishing.
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Baskin, L. M., & Danell, K. (2003). *Ecology of ungulates: A handbook of species in Eastern Europe and Northern and Central Asia* (p. 434). Heidelberg, Germany: Springer Verlag.

- Bieber, C., & Ruf, T. (2005). Population dynamics in wild boar *Sus scrofa*: Ecology, elasticity of growth rate and implications for the management of pulsed resource consumers. *Journal of Applied Ecology*, 42, 1203–1213. <https://doi.org/10.1111/j.1365-2664.2005.01094.x>
- Bobek, B. (2018). Resolving conflicts between farmers and wild boar in Europe and Northern Asia. In M. Meletti, & E. Meijaard (Eds.), *Ecology, conservation and management of wild pigs and peccaries* (pp. 376–387). New York, NY: Cambridge University Press.
- Chen, J., Ni, P., Tran, T., Thuy, N., Varisovich, K. E., Lavrentyevich, P. V., ... Zhao, S. (2018). Selective constraints in cold-region wild boar may defuse the effects of small effective population size on molecular evolution of mitogenomes. *Ecology and Evolution*, 8, 8102–8114. <https://doi.org/10.1002/ece3.4221>
- Danecek, P., Auton, A., Abecasis, G., Albers, A. C., Banks, E., DePristo, A. M., ... 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Djan, M., Veličković, N., Obreht, D., Kočić Tubić, N., Marković, V., Stevanović, M., & Beuković, M. (2013). Mitochondrial DNA control region variability in wild boars from West Balkan. *Genetika*, 45(2), 515–526. <https://doi.org/10.2298/GENSR1302515D>
- Fang, M., & Andersson, L. (2006). Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B: Biological Sciences*, 273(1595), 1803–1810. <https://doi.org/10.1098/rspb.2006.3514>
- Feoktistova, N. Y., Meschersky, I. G., Bogomolov, P. L., Sayan, A. S., Poplavskaya, N. S., & Surov, A. V. (2017). Phylogeographic structure of the common hamster (*Cricetus cricetus* L.): Late Pleistocene connections between Caucasus and Western European populations. *PLoS ONE*, 12(11), e0187527. <https://doi.org/10.1371/journal.pone.0187527>
- Galinsky, K. J., Bhatia, G., Loh, P. R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast principal components analysis reveals independent evolution of ADH1B gene in Europe and East Asia. *American Journal of Human Genetics*, 98, 456–472. <https://doi.org/10.1016/j.ajhg.2015.12.022>
- Geisser, H., & Reyer, H. U. (2004). Efficacy of hunting, feeding, and fencing to reduce crop damage by wild boars. *Journal of Wildlife Management*, 68(4), 939–946. [https://doi.org/10.2193/0022-541X\(2004\)068\[0939:EOHFAF\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2004)068[0939:EOHFAF]2.0.CO;2)
- Geisser, H., & Reyer, H. U. (2006). The influence of food and temperature on population density of wild boar *Sus scrofa* in the Thurgau (Switzerland). *Journal of Zoology*, 267, 89–96. <https://doi.org/10.1017/S095283690500734X>
- Iacolina, L., Brajković, V., Canu, A., Sprem, N., Cubric-Curik, V., Fontanesi, L., ... Scandura, M. (2016). Novel Y-chromosome short tandem repeats in *Sus scrofa* and their variation in European wild boar and domestic pig populations. *Animal Genetics*, 47, 682–690. <https://doi.org/10.1111/age.12483>
- Iacolina, L., Corlatti, L., Buzan, E., Safner, T., & Šprem, N. (2019). Hybridization in European ungulates: An overview of the current status, causes and consequences. *Mammal Review*, 49, 45–59. <https://doi.org/10.1111/mam.12140>
- Jędrzejewska, B., Jędrzejewski, W., Bunevich, A. N., Miłkowski, L., & Krasieński, Z. A. (1997). Factors shaping population densities and increase rates of ungulates in Białowieża Primeval Forest (Poland and Belarus) in the 19th and 20th centuries. *Acta Theriologica*, 42, 399–451. <https://doi.org/10.1.1.918.1235&rep=rep1&type=pdf>
- Khalilzadeh, P., Rezaei, H. R., Fadarar, D., Serati, M., Aliabadian, M., Haile, J., & Goshtasb, H. (2016). Contact zone of Asian and European wild boar at North West of Iran. *PLoS ONE*, 11(7), e0159499. <https://doi.org/10.1371/journal.pone.0159499>
- Kijas, J. M., & Andersson, L. (2001). A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution*, 52, 302–308. <https://doi.org/10.1007/s002390010158>
- Kusza, S., Podgórski, T., Scandura, M., Borowik, T., Jávora, A., Sidorovich, V. E., ... Jędrzejewska, B. (2014). Contemporary genetic structure, phylogeography and past demographic processes of wild boar *Sus scrofa* population in Central and Eastern Europe. *PLoS ONE*, 9, e91401. <https://doi.org/10.1371/journal.pone.0091401>
- Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., ... Cooper, A. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, 307, 1618–1621. <https://doi.org/10.1126/science.1106927>
- Leaper, R., Massei, G., Gorman, M. L., & Aspinall, R. (1999). The feasibility of reintroducing wild boar (*Sus scrofa*) to Scotland. *Mammal Review*, 29(4), 239–259. <https://doi.org/10.1046/j.1365-2907.1999.2940239.x>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Manunza, A., Zidi, A., Yeghoyan, S., Balteanu, V. A., Carsai, C., Scherbakov, O., ... Amills, M. (2013). A high throughput genotyping approach reveals distinctive autosomal genetic signatures for European and Near Eastern wild boar. *PLoS ONE*, 8(2), e55891. <https://doi.org/10.1371/journal.pone.0055891>
- Markov, N. I., & Bolshakov, V. N. (1996). Wild boar in middle Urals: History and modern population status. (pp. 83–88). In *Proceedings of the wild boar symposium, 24–27 March, 1996, Sopron, Hungary*.
- Maselli, V., Rippa, D., Deluca, A., Larson, G., Wilkens, B., Linderholm, A., ... Fulgione, D. (2016). Southern Italian wild boar population, hotspot of genetic diversity. *Hystrix, the Italian Journal of Mammalogy*, 27(2), 1–8. <https://doi.org/10.4404/hystrix-27.2-11489>
- Massei, G., & Genov, P. V. (2004). The environmental impact of wild boar. *Galemys*, 16, 135–145.
- Massei, G., Kindberg, J., Licoppe, A., Gačić, D., Šprem, N., Kamler, J., ... Náhlik, A. (2015). Wild boar populations up, numbers of hunters down? A review of trends and implications for Europe. *Pest Management Science*, 71(4), 492–500. <https://doi.org/10.1002/ps.3965>
- Nekrutenko, A., & Taylor, J. (2012). Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. *Nature Reviews Genetics*, 13, 667–672. <https://doi.org/10.1038/nrg3305>
- Ni, P., Bhuiyan, A. A., Chen, J.-H., Li, J., Zhang, C., Zhao, S., ... Li, K. (2018). De novo assembly of mitochondrial genomes provides insights into genetic diversity and molecular evolution in wild boars and domestic pigs. *Genetika*, 146(3), 277–285. <https://doi.org/10.1007/s10709-018-0018-y>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigen analysis. *PLoS Genetics*, 2, e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, 25, 1253–1256. <https://doi.org/10.1093/molbev/msn083>
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909. <https://doi.org/10.1038/ng1847>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>

- Ramirez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L. S., Folch, J. M., ... Amills, M. (2009). Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Molecular Biology and Evolution*, 26, 2061–2072. <https://doi.org/10.1093/molbev/msp118>
- Ribani, A., Utzeri, V. J., Geraci, C., Tinarelli, S., Djan, M., Veličković, N., ... Fontanesi, L. (2019). Signatures of de-domestication in autochthonous pig breeds and of domestication in wild boar populations from MC1R and NR6A1 allele distribution. *Animal Genetics*, 50(2), 166–171. <https://doi.org/10.1111/age.12771>
- Robinson, T. J., Thorvaldsdóttir, H., Wenger, M. A., Zehir, A., & Mesirov, P. J. (2017). Variant review with the integrative genomics viewer (IGV). *Cancer Research*, 77(21), 31–34. <https://doi.org/10.1158/0008-5472.CAN-17-0337>
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Root, T. L., Price, J. T., Hall, K. R., Schneiders, S. H., Rosenzweig, C., & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants. *Nature*, 421, 57–60. <https://doi.org/10.1038/nature01333>
- Sáez-Royuela, C., & Tellería, J. L. (1986). The increased population of the wild boar (*Sus scrofa* L.) in Europe. *Mammal Review*, 16, 97–101. <https://doi.org/10.1111/j.1365-2907.1986.tb00027.x>
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual* (2nd ed., pp. xxxviii + 1546 pp). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- 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. <https://doi.org/10.1111/j.1365-2907.2010.00182.x>
- Scandura, M., Iacolina, L., Crestanello, B., Pecchioli, E., Di Benedetto, M. F., Russo, V., ... 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. <https://doi.org/10.1111/j.1365-294X.2008.03703.x>
- Soria-Boix, C., Donat-Torres, M. P., & Urios, V. (2017). Contacts in the last 90,000 years over the Strait of Gibraltar evidenced by genetic analysis of wild boar (*Sus scrofa*). *PLoS ONE*, 12(7), e0181929. <https://doi.org/10.1371/journal.pone.0181929>
- Šprem, N., Salajpal, K., Safner, T., Đikić, D., Jurić, J., Curik, I., ... Cubric-Curik, V. (2014). Genetic analysis of hybridization between domesticated endangered pig breeds and wild boar. *Livestock Science*, 162, 1–4. <https://doi.org/10.1016/j.livsci.2013.12.010>
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tan, Z. D., Shen, L. Y., Cheng, X., Gan, M. L., Zhang, S. H., & Zhu, L. (2018). The complete mitochondrial genome sequence of Changbai Mountains wild boar (*Cetartiodactyla: Suidae*). *Conservation Genetics Resources*, 10, 99–102. <https://doi.org/10.1007/s12686-017-0773-6DO>
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The Clustal X Windows interface: Flexible strategies for multiple sequence alignment aided by quality tools. *Nucleic Acids Research*, 24, 4876–4882. <https://doi.org/10.1093/nar/25.24.4876>
- Vai, S., Vilaça, S. T., Romandini, M., Benazzo, A., Visentini, P., Modolo, M., ... Bertorelle, G. (2015). The Biarzo case in northern Italy: Is the temporal dynamic of swine mitochondrial DNA lineages in Europe related to domestication? *Scientific Reports*, 5, 16514. <https://doi.org/10.1038/srep16514>
- Veličković, N., Djan, M., Ferreira, E., Stergar, M., Obrecht, D., Maletic, V., & Fonseca, C. (2015). From north to south and back: The role of the Balkans and other southern peninsulas in the recolonization of Europe by wild boar. *Journal of Biogeography*, 42, 716–728. <https://doi.org/10.1111/jbi.12458>
- Veličković, N., Ferreira, E., Djan, M., Ernst, M., Vidaković, D. O., Monaco, A., & Fonseca, C. (2016). Demographic history, current expansion and future management challenges of wild boar populations in the Balkans and Europe. *Heredity*, 117, 348–357. <https://doi.org/10.1038/hdy.2016.53>
- Vetter, S. G., Ruf, T., Bieber, C., & Arnold, W. (2015). What is a mild winter? Regional differences in within-species responses to climate change. *PLoS ONE*, 10, e0132178. <https://doi.org/10.1371/journal.pone.0132178>
- Vilaça, S. T., Biosa, D., Zachos, F., Iacolina, L., Kirschning, J., Alves, P. C., ... Scandura, M. (2014). Mitochondrial phylogeography of the European wild boar: The effect of climate on genetic diversity and spatial lineage sorting across Europe. *Journal of Biogeography*, 41, 987–998. <https://doi.org/10.1111/jbi.12268>
- Zinovieva, N. A. (2013). Polymorphism of genes associated with the quantitative trait loci in wild boar (*Sus scrofa* L.) in Russia. *Selskokhozyaistvennaya Biologiya (Agricultural Biology)*, 2, 77–82.

How to cite this article: Khederzadeh S, Kusza S, Huang C-P, et al. Maternal genomic variability of the wild boar (*Sus scrofa*) reveals the uniqueness of East-Caucasian and Central Italian populations. *Ecol Evol*. 2019;9:9467–9478. <https://doi.org/10.1002/ece3.5415>

APPENDIX 1

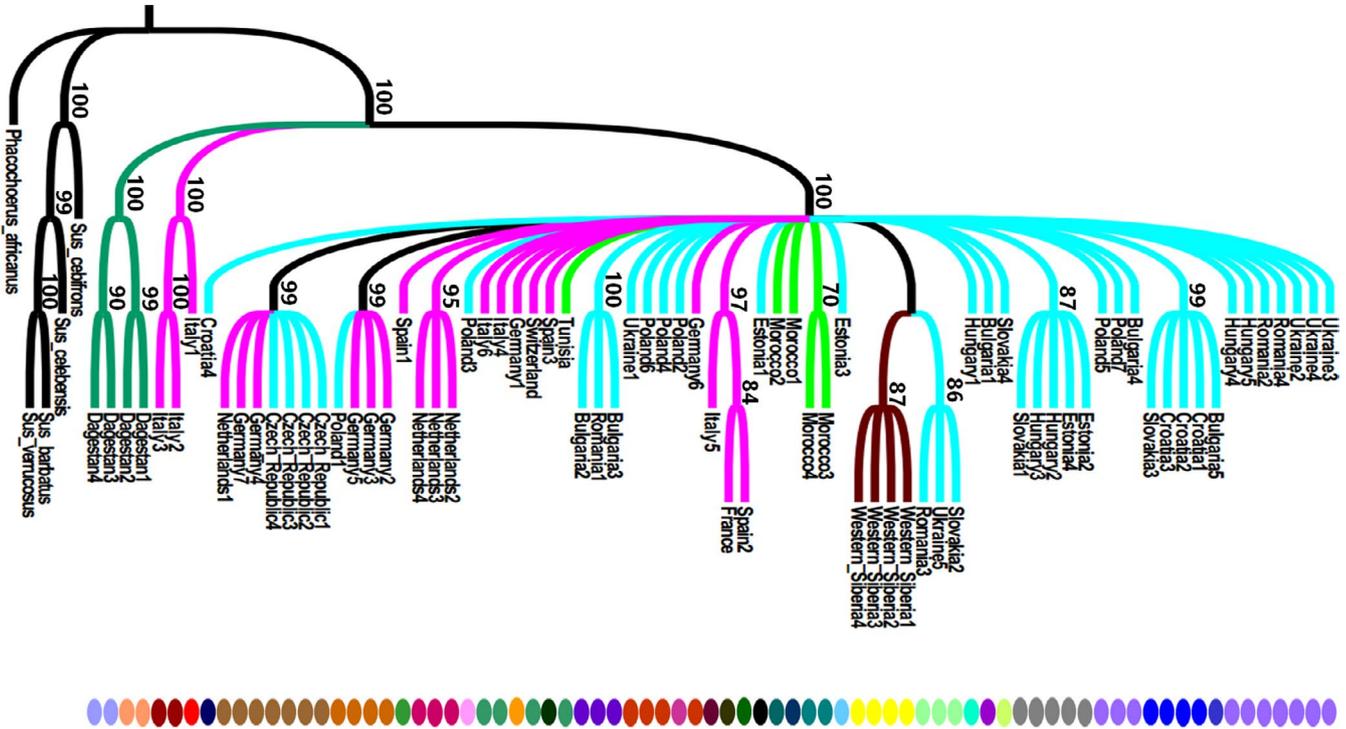


FIGURE A1 Neighbor-joining tree (Tamura-Nei model, 1,000 bootstrap replicates). Different colored lines represent the clusters of populations geographically close to each other (green indicates North Africa; pink indicates West Europe; yellow indicates Central Europe; light blue indicates Eastern Europe; orange indicates Estonia; purple indicates West Russia). Different colored circles represent each of 33 haplotypes

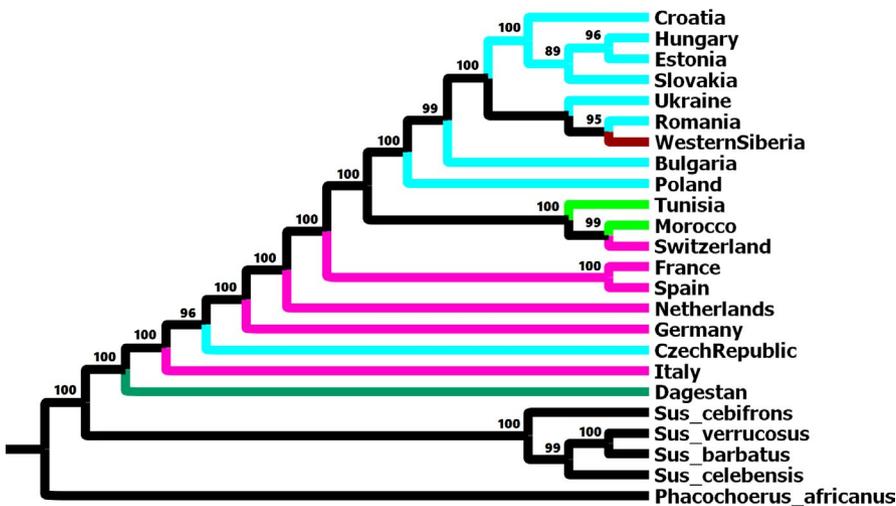


FIGURE A2 Maximum Likelihood tree, depicting the relationship between the SNPs of whole mtDNA genome sequences used in this study (1,000 bootstrap replicates). The colors of the branches represent the clusters of subpopulations (green indicates North Africa; pink indicates Western Europe; yellow indicates Central Europe; light blue indicates Eastern Europe; orange indicates Estonia; purple indicates Western Russia)