Research Article

Southern Italian wild boar population, hotspot of genetic diversity

Valeria Maselli1, Daniela Rippa1, Adriana De Luca2, Greger Larson3, Barbara Wilkens4, Anna Linderholm5, Marco Masseti5, Domenico Fulgione1,∗

1 Department of Biology, University of Naples Federico II, Naples, Italy
2 Department of Veterinary Medicine and Animal Production, University of Naples Federico II, Naples, Italy
3 Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY
4 Department of Science for Nature and Environmental Resources (DepNeT), University of Sassari, Sassari, Italy
5 Department of Biology, University of Florence, Florence, Italy

Keywords:
genetic variability
Italian peninsula
MC1R
mitochondrial DNA
Sus scrofa

Abstract

The wild boar, Sus scrofa, is an important game species widely distributed in Eurasia. Whereas the genetic variability of most European wild boar populations is well known, the status of wild boar living in Southern Italy is not as clear. We evaluated the present and past genetic diversity (D-loop, mtDNA) of the Southern Italian population, comparing it with that observed in other Mediterranean wild boar populations. The mitochondrial DNA diversity of the Southern Italian wild boar population was likewise the result of post-glacial re-colonization of Northern Europe from the Southern European peninsulas (Italy, Iberia and the Balkans), as well as extra-Mediterranean refugia (Scandura et al., 2008; Alves et al., 2010; Alexandri et al., 2012; Schmitt and Varga, 2012). Like many other species, the wild boar was involved in natural and human-mediated dispersal during the Holocene. The early example in the introduction of wild boar to island of Cyprus by at least 11,700-11,400 BP (Vigne et al., 2009, 2011). In fact, Sus scrofa is common in both archaeo-

Introduction

The wild boar Sus scrofa (Linnaeus, 1758) split from the Asian suid (either in Sundaland or mainland Southeast Asia; Larson et al., 2005) and appeared in Western Eurasia about 1 million years ago (Rook and Martínez-Navarro, 2010; Frantz et al., 2013a). Quaternary climatic change shaped geographical patterns of genetic diversity in this species, resulting in two major Old World clades – one in Asia and the other in Europe (Giuffra et al., 2000; Okumura et al., 2001; Alves et al., 2003; Larson et al., 2005; Alexandri et al., 2012). Within Western Eurasia, three well-supported mitochondrial wild boar clades diversified during the Pleistocene period (Larson et al., 2005). One of these clades is present in the Near East (Fertile Crescent), one in continental Europe, and the last one is confined to the Italian peninsula (Larson et al., 2005, 2007b, 2010; Ottoni et al., 2012; Krause-Kyora et al., 2013). In general terms, Quaternary climatic oscillations in Europe provoked species movement along a North-South axis, with the Iberian, Italian and Balkan peninsulas acting as Southern refugia, in the face of climatic deterioration (Bennett et al., 1991; Hewitt, 2004). Phylogeographic studies confirmed that a variety of taxa, including plants (Taberlet et al., 1998), and vertebrates (Hewitt, 1996, 1999, 2000) recolonized Europe from the refugia upon glaciers retreat, and especially so after the Last Glacial Maximum (LGM). It has been proposed that the modern genetic structure of North West European S. scrofa was likewise the result of such post-glacial re-colonization of Northern Europe from the Southern European peninsulas (Italy, Iberia and the Balkans), as well as extra-Mediterranean refugia (Scandura et al., 2008; Alves et al., 2010; Alexandri et al., 2012; Schmitt and Varga, 2012). Like many other species, the wild boar was involved in natural and human-mediated dispersal during the Holocene with the earliest example in the introduction of wild boar to island of Cyprus by at least 11,700-11,400 BP (Vigne et al., 2009, 2011). In fact, Sus scrofa is common in both archaeologi-

Acknowledgements

Research activities were supported by the Glieno/Ufficio di Diano and Alburni National Park and Ambito Territoriale di Caccia Salerno 1 and 2. We would like to thanks for ancient and museum samples the Soprintendenza of Sassary, Nuoro Provinces, the University of Naples, Scandura et al., 2008). Thisspecies was particularly important in the Roman economy and its bones were always found in Roman sites, with higher propor-

©cbe

doi: 10.4404/hystrix–11489

Published by Associazione Teriologica Italiana Volume 27 (2): 137–144, 2016
boars are believed to descend from escaped animals during the Neolithic (Albarella et al., 2006; Scandura et al., 2011b). Thence, they experienced a multi-millennium history of independent evolution under geographic isolation (Randi, 1995; Larson et al., 2005; Iacolina et al., 2016) followed by introgressive hybridisation and gene flow with continental wild boars transplanted to the island in recent years (Scandura et al., 2011b). Instead, free-living pigs currently live in some areas of Sardinia under open-air conditions (Maselli et al., 2014b). Although they are domestic pigs, their feral habits exposes them to some similar selective pressures as wild boars.

We designed our study to provide more details on the South Italian wild boar population, in the light of present and past genetic variability, comparing it with that observed in other Mediterranean glacial refugia.

Materials and methods

We collected both modern and ancient specimens from across Italy (Fig. 1). Modern samples were collected in 2008–2012 during legal hunts in accordance with Italian national laws on wild boar management (prot n° 24581 20/07/2014). Ancient samples came from archaeological sites and museum collections.

Modern samples

We analysed a total of 157 modern samples of which 151 wild boars and six free-living pigs collected from six Italian regions (Tuscany, Latium, Campania, Basilicata, Apulia, and Sardinia) (Fig. 1 and Tab. S2).

DNA extraction, PCR amplification and sequencing

Fresh muscle samples were either stored in plastic tubes (15 ml) filled with 96% alcohol or kept frozen at −20 °C temperature. Total genomic DNA was isolated using a commercial DNA isolation kit (QIAamp DNA Mini Kit, Qiagen, Hilden, Germany). A fragment (652 bp) of the mtDNA D-loop control region (sites 15435–16088 of the reference mtDNA sequence coded AJ002189, Ursing and Arnason, 1998) of the mtDNA D-loop control region (sites 15435–16088 of the reference mtDNA sequence coded AJ002189, Ursing and Arnason, 1998) was amplified using the primers H16108 and L15387 (Watanobe et al., 2001; Larson et al., 2005). This region was selected in order to maximize the size of possible alignments including already published GenBank sequences. PCR reactions were performed in a total volume of 25 µl, containing 50 ng of genomic DNA, 1 unit of Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Hudson, NH, USA), 2.5 µl of 10x GC-rich PCR buffer (Fermentas GmbH, Germany), 2.5 µl of 10x reaction buffer, with MgCl2, 0.25 mM of each dNTP and 12.5 pmol of each primer. The amplifications were carried out in an Eppendorf Mastercycler (Eppendorf AG, Hamburg, Germany) and involved one initial denaturation step at 94 °C for 4 min, followed by 40 cycles of 94 °C for 30 sec, 58 °C for 40 sec, 72 °C for 40 sec and a single final extension step of 72 °C for 5 min.

PCR products were purified by Exo/SAP (Fermentas) digestion, according to the manufacturer’s instructions, and were sequenced using the BigDye® Terminator kit version 3.1 (Applied Biosystems, Foster City, CA). Fragments were finally purified and run in an ABI PRISM 3100 Avant automatic sequencer (Applied Biosystems). All modern samples were sequenced with the forward primer and ambiguous positions were verified by re-sequencing with the reverse primer. Sequences were visually edited using Geneious v. 5.5 software (Biomatters Ltd; Drummond et al., 2011) and multiple sequences from the same sample were compiled into consensus sequence. Multiple alignments of the consensus sequences and published Sus scrofa mtDNA reference sequences were finally created for phylogenetic analysis.

Pig introgression analysis

We analysed the polymorphic melanocortin receptor 1 (MC1R) coat colour gene because it is a useful marker for detecting natural hybridisation between pigs and wild boars (Frantz et al., 2012, 2013b; Krause-Kyora et al., 2013). Various studies support the use of this gene as a tool for distinguishing breeds in a range of animal species (Klungland et al., 1995; Crepaldi et al., 2010). In particular, MC1R allelic variants are closely associated to particular breeds and colour morphs in S. scrofa (Marklund et al., 1996; Möller et al., 1996; Kijas et al., 1998, 2001; Giuffra et al., 2002; Pielpberg et al., 2002; Fang et al., 2009; Koutsogiannouli et al., 2010). For each putative hybrid of modern samples (N=14), we amplified the entire MC1R coding sequence (963 bp) with different pairs of primers to analyse all the polymorphisms in the coding region (MERL1, EPIG1, EPIG3R (Kijas et al., 1998); MF1 and MR1 (Li et al., 2009)). PCR reactions were performed in a total volume of 25 µl, containing 50 ng of genomic DNA, 1 unit of Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Hudson, NH, USA), 2.5 µl HF-GCrich PCR buffer (10x), 200 mM of each dNTP, 20 pmol of each primer. The amplifications were carried out in a Eppendorf Mastercycler (Eppendorf AG, Hamburg, Germany) and involved one initial denaturation step at 94 °C for 4 min, followed by 40 cycles of 94 °C for 30 sec, 58 °C for 40 sec, 72 °C for 40 sec and a single final extension step of 72 °C for 5 min. The sequences of MC1R were aligned and edited by Geneious v. 5.5 software (Biomatters Ltd; Drummond et al., 2011) analysing nucleotide composition and variable sites, compared with GenBank sequences and known alleles (Marklund et al., 1996; Möller et al., 1996; Kijas et al., 1998, 2001; Giuffra et al., 2002; Pielpberg et al., 2002; Fang et al., 2009; Koutsogiannouli et al., 2010).

Ancient Samples

We collected 19 Sus scrofa ancient samples (Tab. S1), from Italian Peninsula (N=9) and Sardinia (N=10). In the former we analysed four specimens dated to about 1,800 AD (one from Tuscany and three from Campania), three specimens dated 79 BC (Pompeii, Campania) and two specimens collected in a Neolithic site dated to about 6,000 BC (Latronico village, Basilicata). In Sardinia we obtained ten samples ranging from 3,500 BC to 1,800 AD. The samples were collected from secure stratigraphic contexts associated with well-defined ceramic assemblages, which provide secure relative chronology (i.e., period and/or sub-period; Wilkens, 1987, 2012).

Prevention of Contamination

Throughout the extraction procedure, we took care to minimize contamination with modern DNA, especially problematic for domestic species (Leonard et al., 2007), and multiple measures were undertaken to exclude contamination by exogenous DNA. Bone powders, DNA extraction and PCR processing were prepared in a facility dedicated to the analysis of ancient specimens, using separate equipment and reagents, following stringent laboratory protocols tailored on ancient DNA (aDNA) processing (Gilbert et al., 2005). Non-disposable equipment (e.g. drill bits) was decontaminated between consecutive treatments when handling the bones. Standard contamination precautions, such as working in a regularly UV-irradiated laboratory exclusively used for pre-PCR work on highly degraded DNA, were employed. Worktops for DNA extraction and PCR setup were separated and cleaned frequently with bleach, distilled water, and 95% eth-

Figure 1 – Sampling locations of modern wild boars and free-living pigs (circles), ancient Sus scrofa specimens (square), or both (circle bullet). See Tab. S1 and S2.
anol. Laboratory personnel wore disposable face masks, full zip suits, protective shoe covers, and sterile latex gloves. The facility was not entered if PCR products had been handled the same day. All equipment, disposables, and nonorganic buffers, except for the tubes used for the aDNA extracts (Eppendorf® RNA/DNA LoBind microcentrifuge tubes, RNase and DNase free), were autoclaved and/or UV irradiated at 254 nm in a crosslinker for 30’. A separate laboratory was used for preparation and addition of the modern standard DNA to PCR reactions.

**DNA Extraction**

Ancient samples were surface sterilized by washing with 0.5% sodium hypochlorite and then rinsed with running deionized distilled water (ddH2O) for 5 min, then air-dried and exposed to UV irradiation for 1 h. One-mm diameter drill-bit (Velleman® electric drill and engraving set. VHMD21B-DC 18V) run at 1,000 rpm generated the bone powder. About 5 g of powder were taken from the internal part of each bone, depending on the age and state of preservation. The DNA extraction was performed by using the QIAquick PCR Purification Kit (Qiagen), which is based on a modified procedure of the protocol described by Yang et al. (1998). Bone powders were dissolved in 8 mL extraction buffer (0.5 M EDTA pH 8.0, 0.5% sodium dodecyl sulfate, and 100 µg/mL proteinase K) and incubated in a shaking waterbath at 55 °C overnight. The extraction solution was centrifuged at 2,000 g for 5 minutes, and 1.75 mL aliquots of the supernatant were transferred to 2.0 mL centrifuge tubes and spun in a micro-centrifuge (12,800 g) for 5 minutes. The supernatant was then transferred to a 10-mL tube and mixed with 5 volumes of QiAquick PB buffer. Using a sterile disposable pipette, 750 µl was loaded directly into a QiAquick column and centrifuged at 12,800 g for 1 min. The flowthrough was discarded and the process was repeated until all of the extract had been passed through the column. The DNA was washed by adding 750 µl of QiAquick PE buffer and centrifuging for 1 min. The flowthrough was discarded and the DNA was then eluted from the column by loading 100 µl TE buffer and centrifuging for 1 min. The isolated DNA was quantified with the NanoDrop 2000 (Thermo Scientific). The duplicate samples were extracted separately, and extraction controls followed each extraction batch with a frequency of one control to every two to five extracts.

**PCR amplification**

Mitochondrial genes are present in up to a thousand times higher concentration than nuclear genes in bone tissues (Smith et al., 2003) and are, for this reason, extensively used in studies on aDNA (Handl et al., 1996; Kringset al., 1997; Stone and Stoneking, 1998; Caramelli et al., 2003; Gillbert et al., 2007b). A highly variable 80-bp fragment (that included numerous indels), identified within a 652-bp alignment (Larson et al., 2007a) was amplified using the primers AncV1 (5’ TTCGTATGCAAACCAAAAG 3’) and AncV2 (5’ TGCATGGGGACTAGCAATTA 3’). PCR reactions were performed in a total volume of 50 µl, containing 100 ng of DNA template, 2.5–3.5 µl of AmpliTaq Gold LD (Applied Biosystems), 50 mM KCl, 10 mM Tris–HCl, 2.5 mM MgCl2, 0.2 mM each of deoxyriboNucleotides (dNTP), 1.0 mg/mL BSA, and 0.3 µM each primer. The amplifications were carried out in an Eppendorf Mastercycler (Eppendorf AG, Hamburg, Germany) and involved one initial denaturation step at 95 °C for 10 min followed by 55 cycles of 94 °C for 20 sec, 60 °C for 30 sec, 72 °C for 50 sec and a single final extension step of 72 °C for 5 min. Each PCR was performed independently twice for each extract. All ancient samples were sequenced with forward and reverse primers as described above.

**Authenticity of Ancient DNA Results**

The authenticity of the ancient sequences and confidence in haplotype assignments are supported by the precautionary measures implemented in the experimental protocol and the coherence of the results obtained in this study: (i) dedicated ancient DNA facilities were used; (ii) robust decontamination protocol of the bone samples was employed before DNA extraction; (iii) all blank extracts were included in PCRs and each PCR was performed with negative and positive controls (iv) no PCR amplification was observed in blank extracts and PCR negative controls; (v) multiple haplotypes were obtained as a whole, as well as within most extraction batches; (vi) the same ancient haplotypes were obtained in two separate laboratories, using different extraction and amplification protocols; and (vii) all sequences belong to Sus scrofa. Two repeated extractions and amplifications were conducted for all samples. Consistent results were obtained for all replications. Although some differences, likely attributable to post-mortem genetic damage, were detected, these did not influence the designation of the specimens to different haplotypes. Ancient DNA experiments were performed in dedicated facilities in the University of Naples Federico II. In order to verify the authenticity of the data obtained in Naples, seven samples were further analysed for blind replication in the independent laboratory at Durham University. Both laboratories followed the standard criteria for ancient DNA handling (Gillbert et al., 2005).

**Data analysis**

A first round of analysis was conducted in order to assess the genetic relationship between the identified new haplotypes and additional 132 sequences of wild boar sequences sampled worldwide, chosen as to represent the current genetic diversity of Western Eurasia (GenBank, Tab. S3) (Watanobe et al., 1999; Giuffra et al., 2000; Okumura et al., 2001; Watanobe et al., 2001; Kim et al., 2002; Randi et al., 2002; Alves et al., 2003; Gongora et al., 2005, 2007a; Scandura et al., 2008; Lattuada et al., 2009; Alves et al., 2010; Gongora et al., 2011). A second round of analysis was conducted on sequences (n=501) grouped according to specific refugia (Balkan, Italian and Iberian) across eight European and Mediterranean countries (Tab. S4), comparing shorter sequences of 610 bp. Finally we assessed the genetic variability within the Italian refugium (335 sequences) considering populations grouped by regions (Tab. S5). In this dataset the newly generated sequences were trimmed in a shorter segment of 410 bp in order to compare them with the most of Italian sequences deposited in GenBank (Scandura et al., 2008).

In order to deeper analyse the genetic variability of Italian refugium, we aligned the newly 19 ancient sequences (Tab. S1) with sequences of 60 different ancient haplotypes obtained from GenBank, truncated to make them comparable ( Larson et al., 2007a) (Tab. S6). All data-sets were performed using Geneious v. 5.5.5 software (Biomatters Ltd; Drummond et al., 2011) and final adjustments were performed by eye. The alignments were collapsed in distinct haplotypes using DnaSP v5 (Librado and Rozas, 2009).

To detect genetic structure within Italy and to reveal the phylogeography and the phylogeography of Italian wild boar within a European framework, Bayesian phylogenetic analyses were carried out, using reference sequences of modern (Tab. S3) and ancient European samples (Tab. S1). The best fit model of nucleotide substitution was estimated using a hierarchical likelihood ratio test (hLRT) as implemented in jMODELTEST 3.7 (Posada and Crandall, 1998; Guindon and Gascuel, 2003). The best model resulted the HKY model with inv-gamma distributed (HKY+G+I) rate variation across the sites, based on the corrected Akaike Information Criterion (-lnL=-3535.78). AICc weight=0.792) The best fit parameters were included as priors for MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and eight Markov Chain Monte Carlo (MCMC) searches were run for 30 million generations. For each MCMC, a tree was sampled every 100 generations. Tracer 1.6 (Rambaut et al., 2014) was used to summarize Bayesian analyses, to inspect the validity of the burn-in fraction applied and to evaluate the convergence. The first 10% of samples were discarded as burn-in (Rambaut et al., 2014) and convergence was assessed by checking the log likelihoods, the average standard deviation of split frequencies (<0.01), and the potential scale reduction factor in MrBayes. We also assessed convergence by visual inspection of the trace and the estimate of the effective sample size (ESS>200). A consensus tree from the retained trees was computed with MrBayes. Sulawesi Warty Pig (Sus celebensis) D-loop sequence was included as an outgroup (GenBank accession number: GQ338955). Sometime consistent trees are diffi-
cult to obtain under high rates of homoplasy, this is the reason why we used a median joining (MJ) network (Bandelt et al., 1999). Therefore, in addition to phylogenetic trees, a MJ network of haplotypes was created for the European two alignment sets (modern and ancient, Tab. S3 and Tab. S1), by using the software Network 4.1.0.9 (Fluxus Technology). A maximum parsimony calculation was used to eliminate non-parsimonious links (Bandelt et al., 1999). To obtain better insights into the genetic diversity and demographic history of wild boar among Mediterranean refugia and within Italian refugium we measured the levels of molecular diversity. Number of polymorphic sites (S) and haplotypes (H), average number of pairwise differences (k), Theta (Θ) (Nei, 1987) and nucleotide diversity (π) were computed with DnaSp v5 (Librado and Rozas, 2009). To gain an insight into the possible demographic history over time, a mismatch analysis was carried out (Rogers and Harpending, 1992) for all mtDNA sequences from Italy (Tab. S5) and Mediterranean refugia (Tab. S4). The mismatch analysis does not consider the Asian samples, because attributed to local introgression with domestic pigs. Significance was assessed after 10,000 replicates, using a parametric bootstrap approach, under the null hypothesis that the observed data fit the sudden expansion model (Schneider and Excoffier, 1999). The mismatch analysis and the raggedness index analysis were performed in DnaSp v5 (Librado and Rozas, 2009).

Results

Modern samples

In modern samples, we identified 39 different haplotypes (named from VM_H1 to VM_H39; [Genbank: HM-051193-201, HM448432, KF873473-501]), of which 32 had been never described before. We identified 17 distinct haplotypes exclusive to Sardinia, 20 exclusive to the Italian peninsula and only two haplotypes were shared by Southern Italy and Sardinia. Bayesian phylogenetic analysis (using the HKY85+G+I model) from living specimens pointed out three distinct clades: European E1, Italian E2, and East Asian A (Larson et al., 2005; Scandura et al., 2008) (Fig. S7). The European E1 clade accounted for 80.9% of sequences, 10.2% for the Italian E2 clade, and 8.9% for the Asian clade. Private Italian haplotypes (clade E2) were found both in Sardinia and in Italian peninsula, including the most common BG_I-Italy described in Larson et al. (2005). We observed also some differentiation among haplotypes assigned to the Italian clade. Sardinian samples carried two distinct haplotypes, basal to other E2 haplotypes detected in peninsular samples, possibly reflecting the long geographic isolation experienced (Fig. 2).

Inside the European clade only 29% of the 157 sequences were ascribed to the most common European haplotypes (W, BK, AF, E and M in Larson et al., 2005). Among Southern Italian samples (Fig. 2a), one free-living domestic pig and 13 wild boars, belonged to the East Asian haplogroup (D2 in Larson et al., 2005). These samples were genotyped at MC1R gene, revealing that just the free-living pig (from Castelnuovo Cilento) showed a typical domestic pig sequence (accountable to genotype (501), whereas all wild boars were wild-type (genotype 0101; Fang et al., 2009). In addition to the phylogenetic tree, we performed a network analysis (Fig. 2a) in which we identified three distinct clades (European, Italian, and Asian) corroborating the Bayesian analyses results. The Italian haplotypes showed a scattered pattern highly dispersed according to the private mutations and new haplotypes. Common haplotypes were surrounded by a starlike pattern, suggesting a population expansion (Fig. 2a). In the map showing the frequency of the three haplogroups (E1, E2, and A) in each of the eight Italian populations (Fig. 2b) we observed a high degree of variability in the Southern populations, sampled in Campania and Apulia. Resolving the European clade (Fig. 3) in consideration of C-side and A-side assignment (Larson et al., 2005) we observed that the Southern Italian haplotypes occurred within both European core lineages, with similar proportions (46.6% A-side and 53.4% C-side). Genetic variability in Italian wild boar populations can be inferred also comparing nucleotide characteristics of sequences among three refugia.

Southern Italian samples thus had higher nucleotide diversity overall than other refugia (Tab. 1; π=0.00391 for Balkans, π=0.00396 for Iberia, and π=0.00994 for South Italy). We observed a similar trend for the average number of nucleotide differences (k) (Tab. 1; k=2.3017 for Balkans, k=2.3888 for Iberia, and k=5.9668 for South Italy). The number of haplotypes was quite similar for South Italy and Balkans (38 and 40 respectively), but they showed a high frequency of private haplotypes (89.5% and 90.0% respectively) and only two haplotypes were shared between these two refugia. The Iberian refugium showed a lower number of haplotypes (16) and percentage of private haplotypes (68.8%). Focusing on samples from Italy, owing to the reduction in the total length of sequences used for the analyses (410 bp), several Italian haplotypes collapsed together (Tab. 2). Among the different Italian populations, it was clear that Campania and Sardinia showed the highest diversity in the peninsula, with higher number of haplotypes (18 and 15 respectively) and private haplotypes (66.7% and 77.8% respectively). Moreover mtDNA allowed us to make further inference on the demographic characteristic of population. The pattern of mismatch distribution suggested that the Iberian and Balkan populations experienced a sudden expansion, which was absent in Italian populations (Fig. 4). High peaks in pairwise difference were in accordance with a demographic history including different haplogroups (secondary contacts), especially in Southern Italy and Sardinia (Fig. 4).
Table 1 – Genetic diversity observed in modern wild boar populations of Mediterranean refugia (D-loop, 610 bp). Pop, population; n, sample size of each population; S, number of polymorphic sites; H, number of haplotypes; AU, number of private haplotypes; U/k, percentage of private haplotypes; Θ, Theta; π, nucleotide diversity; k, average number of pairwise differences.

<table>
<thead>
<tr>
<th>Pop</th>
<th>n</th>
<th>S</th>
<th>H</th>
<th>AU</th>
<th>U/k</th>
<th>Θ</th>
<th>π</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balkans</td>
<td>176</td>
<td>38</td>
<td>34</td>
<td>89.5</td>
<td>0.01155</td>
<td>0.00391</td>
<td>2.3017</td>
<td></td>
</tr>
<tr>
<td>Iberia</td>
<td>159</td>
<td>15</td>
<td>11</td>
<td>68.8</td>
<td>0.00441</td>
<td>0.00396</td>
<td>2.3888</td>
<td></td>
</tr>
<tr>
<td>South Italy</td>
<td>166</td>
<td>30</td>
<td>40</td>
<td>36</td>
<td>90.0</td>
<td>0.00881</td>
<td>0.00994</td>
<td>5.9668</td>
</tr>
</tbody>
</table>

Table 2 – Genetic diversity observed in modern Italian wild boar populations (D-loop, 410bp). Pop, population; n, sample size of each population; S, number of polymorphic sites; H, number of haplotypes; AU, number of private haplotypes; U/k, percentage of private haplotypes; Θ, Theta; π, nucleotide diversity; k, average number of pairwise differences.

<table>
<thead>
<tr>
<th>Pop</th>
<th>n</th>
<th>S</th>
<th>H</th>
<th>AU</th>
<th>U/k</th>
<th>Θ</th>
<th>π</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piedmont</td>
<td>19</td>
<td>12</td>
<td>7</td>
<td>1</td>
<td>14.3</td>
<td>0.00846</td>
<td>0.01089</td>
<td>4.4211</td>
</tr>
<tr>
<td>Emilia Romagna</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0.0</td>
<td>0.00435</td>
<td>0.00867</td>
<td>2.6667</td>
</tr>
<tr>
<td>Tuscany</td>
<td>57</td>
<td>13</td>
<td>6</td>
<td>1</td>
<td>16.7</td>
<td>0.00694</td>
<td>0.01248</td>
<td>5.0664</td>
</tr>
<tr>
<td>Latium</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>50.0</td>
<td>0.00079</td>
<td>0.00126</td>
<td>0.5128</td>
</tr>
<tr>
<td>Campania</td>
<td>107</td>
<td>19</td>
<td>15</td>
<td>10</td>
<td>66.7</td>
<td>0.00897</td>
<td>0.01153</td>
<td>4.6563</td>
</tr>
<tr>
<td>Basilicata</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td>3</td>
<td>60.0</td>
<td>0.01219</td>
<td>0.01806</td>
<td>7.3333</td>
</tr>
<tr>
<td>Apulia</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>50.0</td>
<td>0.01970</td>
<td>0.01970</td>
<td>8.0000</td>
</tr>
<tr>
<td>Sardinia</td>
<td>117</td>
<td>20</td>
<td>18</td>
<td>14</td>
<td>77.8</td>
<td>0.00926</td>
<td>0.00867</td>
<td>3.5111</td>
</tr>
</tbody>
</table>

Figure 3 – Median-joining network of European clade (E3). Detail of the main European clade depicting the relationships among haplotypes detected in the different Italian populations. The A and C haplotypes correspond to the two core lineages. Size of circles is proportional to the number of sampled individuals; if more than one mutation occurs between two haplotypes, they will be connected by dashed line, each dash representing one mutational step.

Figure 4 – Mismatch distributions based on pairwise site differences between sequences, performed in DNAsp (Librado and Rozas, 2009) for different populations: (a) Balkans; (b) Iberia; (c) South Italy; (d) Sardinia. The expected data were computed under the model of demographic expansion. On the horizontal axis there is the number of nucleotide differences between pairs of sequences and on the vertical axis there is the relative frequencies of pairwise comparisons. Grey and black bars represent expected and observed frequencies, respectively.

Ancient samples
We successfully extracted and sequenced DNA from 19 ancient *Sus scrofa* bones sampled from archaeological sites or museum. Nucleotide sequences from each replicate were identical for 18 samples. Just in one sequence from Pompeii (PR_VM03) we found a misincorporation (A instead of T, at position 20), but this did not influence the designation of the haplotypes. In ancient samples we identified five different haplotypes (named from VM_AH3 to VM_AH7; Additional file 1 [GenBank: KF873503-06, HM051202]). Phylogenetic Bayesian consensus tree (Fig. S7b) drawn using only the 80-bp partial sequence of mtDNA D-loop was useful to identify four distinct clades (European, Italian, Asian, and Near East Asian). Surprisingly, we found a Near Eastern haplotype (NE2-Y1) in three museum samples from Campania, (about 1,800 AD) and one museum sample from Tuscany (about 1,800 AD), suggesting that *Sus scrofa* with this genetic trait was present in Italy at least until 200 BP, before disappearing from modern populations (Fig. 2, Fig. 5, Tab. S1, Tab. S6, Fig. S7) (Zinser et al., 2003).

Two haplotypes belonging to the endemic Italian clade came from ancient samples from Sardinia (Alghero, Neolithic) and Basilicata (Mesolithic). The Sardinian haplotype was consistent with those observed in modern samples. All other haplotypes from Sardinia (Carbonia, Sassari, Alghero) and all samples from Pompeii (79 AD) belonged to the European clade (ANC-C side), indicating the presence of this clade on Sardinia during Neolithic age. The alignment with ancient samples was used to draw a median-joining network (Fig. 5a) to further elucidate the differences between the major clades (European, Italian, Asian, and Near East Asian). The time series of maps (Fig. 5b) depicting the shifting geographic positions for the Italian, European and Near Eastern haplotypes over the past, is consistent with those previously reported Larson et al. (2007a) and expanded the range and timing of the Italian clade, observed in Sardinia.

Discussion
The role of Italy as a refugium in Europe for a lot of species during the LGM is already well known (Hewitt, 1999, 2004) and there is evidence that the wild boar distribution followed the general pattern exhibited by the majority of mammal species (Scandura et al., 2008; Alexandri et al., 2012). We analysed the genetic structure and variability of wild boar populations from Sardinia and South Italy, previously suggested as
one of the most important glacial refugium (Scandura et al., 2008; Vilaça et al., 2014). We provided further evidence supporting the role of this refugium as hotspot of genetic variability for wild boar. Our findings were based on data from Mesolithic age to recent times, elucidating how both glaciation and human movement affected the wild boar variability pattern. Data on mtDNA from Southern Italian wild boar populations, principally due to the presence of the three major mtDNA lineages (E1, E2, and A) and private haplotypes, corroborated the general pattern of decreasing northward genetic diversity (Vilaça et al., 2014). The most frequent wild boar haplotypes in Italy belonged to E1 haplogroup, which was also shared among the different refugia (Italian, Balkans and Iberian) as well as by all other European populations analysed so far (Kuszsa et al., 2014; Vilaça et al., 2014). The existence of a large number of private haplotypes in South Italy suggested that a large fraction of the pre-glacial wild boar diversity was maintained in this area during the last glacial period. Similar cases were found in other species (Sommer et al., 2005). These findings supported the scenario of a post-glacial northward expansion from Italy to Central Europe and beyond, with the gene pool from North Italy contributing more to the recolonization process than the gene pool from Southern parts. The genetic variability observed in Central Italy (Tuscany) was mainly due to the presence of a high proportion of samples belonging to the private Italian clade (E2). Finally, only one private Italian haplotype was found in North Italy (Piedmont and Emilia Romagna).

A minority of Southern Italian haplotypes clustered with East Asian wild boar and pigs. Indeed, the Asian A haplotypes made up as much as 29% of European domestic pig breeds (Fang and Andersson, 2006), due to historical introgression with Chinese pigs (18th–19th century) (Giuffra et al., 2000). The presence of A haplotypes in wild individuals thus proved the occurrence of a domestic pig ancestor in their maternal line. The occurrence of free-ranging pigs in our sampled areas and to the practice of restocking in southern regions, possibly using captive hybrids, can explain this presence. In the Southern Italy wild boar population the contact between wild and domestic pigs is well known and it is due to local farming practices still existing in few regions of the Mediterranean basin (Randi, 2005; Maselli et al., 2014a,b). When pigs escape from farms or are deliberately allowed to graze freely, their chance to get into contact with wild boars is high. In Italy, another source of genetic variability came from the Sardinian population. It shared the major mtDNA haplotypes with the Italian Peninsula, mostly European E1, but a lot of haplotypes were nonetheless endemic. In Sardinia, the Italian lineage (E2=6%) confined on the island was characterised by a nucleotide divergence generating a sub-clade within the Italian one. High level of genetic variability in wild boar of Southern regions (Latium, Campania, Basilicata, Apulia and Sardinia) can be explained by the long-term persistence of favourable environmental conditions within the refugium. They allowed for the maintenance of stable and genetically variable populations (Canestrelli et al., 2010), and probably, a constantly high effective population size. This condition was probably maintained by a series of secondary contacts as suggested by mtDNA analysis. In fact in Italy multiple peaks can be explained by the presence of E1 and E2 haplotypes as well as to wild boar introductions occurred in the past. Whereas Balkan and Iberian sequences showed an unimodal distribution, that likely reflected a long-term demographic expansion (Fig. 4).

Ancient DNA from Sus contributed to provide information about the complex temporal and geographical pattern. In South Italy, the private Italian haplotype (Anc-Italy) was originally present, as shown in the remains of Latronico (Basilicata). Similarly, Sardinian samples, dating back to the Neolithic age, belonged to the Italian clade, reflecting contacts between Peninsula and island cultures (Fig. 5). With our contribution we expanded the range and timing of some of the haplotypes (Larson et al., 2007a), e.g. samples belonging to the Italian clade observed in Sardinia during Neolithic age, and samples assignable to the Near Eastern lineage occurring in Tuscany and Campania in 1,800 AD (Fig. 5b). As plainly indicated also by our ancient mtDNA data (see Pompeii [79 AD], Sassari [27 BC - 476 AD] and Corbara [VI–VII sec. BC] samples), the European E1 haplotypes were well represented in the past in South Italy, either as autochthonous populations previously settled or as a consequence of repeated, ancient introductions. In Italy the Near Eastern haplotype NE2-Y1 was found in museum samples 1,800 AD from Tuscany and Campania, previously recorded in archaeological remains from Romania to France and in modern populations on Corsica, where it still exists (Larson et al., 2007a; Larson, 2012). The first domestic pigs introduced into Europe possessed mitochondriald signatures acquired from Anatolian wild boar, but this signature was replaced by mtDNA haplotypes possessed by European wild boar about 4,000 years ago (Larson et al., 2007a). However recent studies typed the presence of the Near Eastern haplogroup NE2-Y2 in North Italy in some pre-Neolithic Sus remains (Vai et al., 2015), so our samples can descend from early domesticated pigs arrived into Europe during the Neolithization of the continent or from a legacy of previous contacts with some northern populations. The difference between the two haplotypes observed in North and Central-South Italy (NE2-Y2 and NE2-Y1, respectively) is in a single mutation (an indel), and together with the large temporal difference from those remains, made this interpretation of results very difficult. Both hypotheses should be verified, considering more samples from Mesolithic to recent times. Finally, considering data on modern and ancient samples altogether, we can speculate about a high genetic diversity in Italian wild boar as a result of several additive factors including glaciation cycles, natural expansion and human translocation of animals, since Neolithic. 

References


Supplemental information

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

Table S1 A list of ancient Sus scrofa specimens obtained for this study, extracted, and sequenced and deposited on GenBank.

Table S2 A list of modern Sus scrofa specimens analyzed in this study. Corresponding sequences were deposited on Genbank.

Table S3 A list of modern Sus scrofa sequences deposited on GenBank used to build Bayesian tree.

Table S4 Sequences used to define the genetic variability among different populations each assigned to specific refugia (Balkans, Iberia, and Italy).

Table S5 Sequences used to define the genetic variability within the Italian refugium, considering populations divided by regions (Piedmont, Emilia Romagna, Tuscany, Latium, Campania, Basilicata, Apulia, Sardinia).

Table S6 A list of ancient Sus scrofa specimens available on Genbank and used in this study.

Figure S7 Bayesian trees.