

## ORIGINAL ARTICLE

# Genomic diversity and differentiation of a managed island wild boar population

L Iacolina<sup>1,2</sup>, M Scandura<sup>1</sup>, DJ Goedbloed<sup>3</sup>, P Alexandri<sup>4,5</sup>, RPMA Crooijmans<sup>4</sup>, G Larson<sup>6</sup>, A Archibald<sup>7</sup>, M Apollonio<sup>1</sup>, LB Schook<sup>8</sup>, MAM Groenen<sup>4</sup> and H-J Megens<sup>4</sup>

The evolution of island populations in natural systems is driven by local adaptation and genetic drift. However, evolutionary pathways may be altered by humans in several ways. The wild boar (WB) (*Sus scrofa*) is an iconic game species occurring in several islands, where it has been strongly managed since prehistoric times. We examined genomic diversity at 49 803 single-nucleotide polymorphisms in 99 Sardinian WBs and compared them with 196 wild specimens from mainland Europe and 105 domestic pigs (DP; 11 breeds). High levels of genetic variation were observed in Sardinia (80.9% of the total number of polymorphisms), which can be only in part associated to recent genetic introgression. Both Principal Component Analysis and Bayesian clustering approach revealed that the Sardinian WB population is highly differentiated from the other European populations ( $F_{ST}=0.126\text{--}0.138$ ), and from DP ( $F_{ST}=0.169$ ). Such evidences were mostly unaffected by an uneven sample size, although clustering results in reference populations changed when the number of individuals was standardized. Runs of homozygosity (ROHs) pattern and distribution in Sardinian WB are consistent with a past expansion following a bottleneck (small ROHs) and recent population substructuring (highly homozygous individuals). The observed effect of a non-random selection of Sardinian individuals on diversity,  $F_{ST}$  and ROH estimates, stressed the importance of sampling design in the study of structured or introgressed populations. Our results support the heterogeneity and distinctiveness of the Sardinian population and prompt further investigations on its origins and conservation status.

*Heredity* (2016) **116**, 60–67; doi:10.1038/hdy.2015.70; published online 5 August 2015

## INTRODUCTION

Single-nucleotide polymorphisms (SNPs) represent the most widespread type of sequence variation in genomes and their use in animal population studies is increasing, thanks to innovative methods for SNP identification and automated screening (Van Zyl *et al.*, 2014). SNPs are attractive markers for population genetic inference as they are usually the most abundant type of variation in genomes, and SNP-genotyping assays have low error rate and do not require calibration between laboratories (see Brumfield *et al.*, 2003; Morin *et al.*, 2004). SNP-based assays have proven to be powerful tools for inferring genetic relationships, population history and population structure within and between populations in model animal species (for example, Ramos *et al.*, 2009). However, their use is increasing even in non-model organisms (for example, Gomez-Uchida *et al.*, 2011), especially through their transfer from model (often domestic) to closely related species (for example, VonHoldt *et al.*, 2011; Goedbloed *et al.*, 2013a). Unfortunately, this practise is known to arise an ascertainment bias that could possibly translate into errors when estimating population history and other demographic parameters (for example, Morin *et al.*, 2004) making population samples appear more similar than they truly are (Brumfield *et al.*, 2003).

However, this bias usually have a weaker effect on closely related species (VonHoldt *et al.*, 2011) and should not affect within-population comparisons, while it could ask for cautiousness to compare a distant species with the one used for the SNPchip development (Pertoldi *et al.*, 2010).

The wild boar (WB) is an iconic game species of prime interest among European hunters and is therefore highly managed throughout its distribution range. Animals of heterogeneous origin, including alien subspecies and captive-bred individuals, have been often released to restore overhunted populations (Apollonio *et al.*, 2014). Furthermore, in several cases farmed WBs were crosses between wild and domestic pigs (DP) (Scandura *et al.*, 2011a; Goedbloed *et al.*, 2013a; Canu *et al.*, 2014).

Human-induced alterations to WB populations have likely been ongoing since pre-Neolithic time, when humans started to manage and introduce them to islands as supplementary food supply (Vigne *et al.*, 2009) and continued during the domestication process and the following co-existence with DP. Before the 16th century, pigs were kept free-ranging, and crossbreeding between human-reared and wild pigs would have been possible throughout Europe (White, 2011). Although these practises have apparently weakly affected the historic

<sup>1</sup>Department of Science for Nature and Environmental Resources, University of Sassari, Sassari, Italy; <sup>2</sup>Department of Chemistry and Bioscience, Faculty of Engineering and Science, Section of Biology and Environmental Science, Aalborg East, Denmark; <sup>3</sup>Braunschweig, Zoological Institute, Braunschweig, Germany; <sup>4</sup>Wageningen University, Animal Breeding and Genomics Centre, Wageningen, the Netherlands; <sup>5</sup>Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece; <sup>6</sup>Durham Evolution and Ancient DNA, Durham University, Department of Archaeology, Durham, UK; <sup>7</sup>The Roslin Institute, R(D)SVS, University of Edinburgh, Division of Genetics and Genomics, Midlothian, UK and <sup>8</sup>University of Illinois, Laboratory of Comparative Genomics, Urbana, IL, USA  
Correspondence: Dr M Scandura, Department of Science for Nature and Environmental Resources, University of Sassari, via Muroni 25, I-07100 Sassari, Italy.  
E-mail: scandura@uniss.it

Received 4 October 2014; revised 19 March 2015; accepted 5 May 2015; published online 5 August 2015

phylogeographic pattern of the species (Larson *et al.*, 2005; Scandura *et al.*, 2008; Vilaça *et al.*, 2014), their nowadays occurrence can have an impact on the genetic makeup of modern populations (Scandura *et al.*, 2011a, b; Goedbloed *et al.*, 2013a).

Recent studies have revealed that the WB population inhabiting Sardinia island shows an admixed composition, resulting from a mix of ancient and recent introductions, and by the local hybridisation with DP (Scandura *et al.*, 2011b). As the earliest palaeontological evidence for WB presence on the island dates back to the early Neolithic and coincide with the first human settlements (Wilkins, 2003), it is generally assumed that humans have brought the species to Sardinia, either as wild or very early domesticated form (Albarella *et al.*, 2006). After escaping from human control, being isolated from continental populations, they have diverged from other WBs (Masetti and Mazza, 1996). As a consequence of phenotypic and biogeographic distinctness, Sardinian WBs were classified as a separate subspecies (*Sus scrofa meridionalis* Major, 1883; Apollonio *et al.*, 1988). But crossbreeding with DP in some areas of Sardinia, where outdoor pig farming is still practiced, and the uncontrolled introduction of continental WBs, have threatened and possibly compromised the genetic identity of the island population.

Although the genetic diversity and population structure of the Sardinian population have been previously described with different classes of molecular markers including allozymes, mitochondrial DNA and microsatellites (Randi *et al.*, 1989; Scandura *et al.*, 2011b), recently developed genome-wide high-density porcine SNP panels (Ramos *et al.*, 2009) represent a valuable source of diagnostic power for the investigation of genomic distinctiveness, phylogeographic patterns and population admixture. The use of the Porcine SNP60 Beadchip technology was very informative in exploring the patterns of population structure and genetic introgression in north-western European WB populations (Goedbloed *et al.*, 2013a), disclosing patterns that were not easily detected by other markers.

In this study we use a genome-wide SNP-genotyping assay to assess the current genomic diversity and distinctiveness of the Sardinian WB population. In doing so, we specifically looked for signs of recent introgression and evaluated the possible impact on the overall genetic diversity and differentiation from continental wild populations and from DP. Considering that an unbalanced sampling can introduce a bias when comparing levels of genomic diversity and differentiation across a panel of populations (Pilot *et al.*, 2014), we investigated the effect of sample size and sample composition on our estimated parameters.

## MATERIALS AND METHODS

### Sampling and genotyping

WB samples collected throughout Sardinia in the period 2003–2010 ( $N=99$ ) were combined, for comparison, with WB samples from other six European regions (Central Italy  $N=15$ , Iberia  $N=18$ , Balkans  $N=55$ , South-Western Europe  $N=54$ , North-Western Europe  $N=37$ , Central-Northern Europe  $N=17$ ), giving a total of 295 individuals from 17 countries. In addition, 105 DP (with a number of individuals per breed from 8 to 10) belonging to five international (Berkshire, Duroc, Large White, Yorkshire, Pietrain) and five Italian breeds (Casertana, Nera Siciliana, Mora Romagnola, Calabrese, Cinta Senese) from the PigBioDiv project (see Megens *et al.*, 2008) were used for comparison in addition to a sample of Sardinian free-ranging pigs (see Supplementary Table S1a, b in Supplementary Information). All individuals were opportunistically sampled and checked for relatedness after genetic analysis, as they all resulted to be unrelated no additional selection was performed. Animals were analysed with the Porcine SNP60 Beadchip (Ramos *et al.*, 2009), according to the manufacturer's instructions ([http://www.illumina.com/products/porcineSNP60\\_dna\\_analysis\\_kit.ilmn](http://www.illumina.com/products/porcineSNP60_dna_analysis_kit.ilmn)). Analyses were performed

on the 49 803 SNPs (referred to as 50 K) mapping on autosomes, out of 62 163 total assayed SNPs. All individual samples included in this study had a genotyping rate  $>0.9$ , with an average of 0.993.

### Statistical analysis

To examine genetic differentiation among the Sardinian population and the other wild and domestic populations, we performed Principal Component Analysis (PCA) using FLASHPCA (Abraham and Inouye, 2014). To avoid the possible confounding effect related to the presence of loci in linkage disequilibrium the calculation was performed on a subset of 30 127 SNPs (referred to as 30 K) obtained by removing SNPs in linkage disequilibrium ( $r^2 > 0.5$ ) with PLINK 1.09 (Purcell *et al.*, 2007). On the basis of the PCA results and in concordance with previous findings (Scandura *et al.*, 2008), we grouped the populations into four main groups (domestic breeds—DP, Sardinian WB—WSar, Italian WB—WIta and the remaining European wild populations all together—WEur). The 50 K data set was used to assess variability levels of the populations by calculating minor allele frequencies (which indicates the abundance of rare alleles through the genome), as well as expected ( $H_E$ ) and observed ( $H_O$ ) heterozygosity within populations in PLINK.  $F_{IS}$  and Hardy–Weinberg equilibrium were tested in the Sardinian population using GENEPOP 4.1 (Rousset, 2008), adjusting significance levels by the sequential Bonferroni procedure (Rice, 1989). Genetic differentiation among populations was estimated by calculating pairwise  $F_{ST}$  values in GENEPOP using the 30 K subset to avoid a bias owing to physical linkage between loci (Helyar *et al.*, 2011).

With the aim to explore the genetic structure across populations and to assess the source of putative immigrants or introgressed individuals in the Sardinian population, a Bayesian clustering assignment was performed using the 30 K data set in STRUCTURE 2.2 (Pritchard *et al.*, 2000), accessing to the Biportal cluster at the University of Oslo (Kumar *et al.*, 2009). Ten independent runs of 100 000 iterations, following a burn-in of 80 000 iterations, were performed for each value of  $K$  between 1 and 20, neglecting prior population information, assuming independence among loci, and allowing admixture. The Evanno method (Evanno *et al.*, 2005) was applied to infer the most reliable number of genetic clusters ( $K$ ). Within the identified  $K$ -value, among the 10 runs, we selected the one with the highest log-likelihood value and calculated the proportion of membership to each of the  $K$  clusters for any single individual ( $q_i$ ).

To evaluate the extent and nature of genetic introgression in the Sardinian WB, we first identified at each value of  $K$ , which of the  $K$  clusters was associated to the Sardinian sample, that is, the cluster to which the WSar population showed the highest average membership ( $Q_{Sar}$ ). Then we studied the variation of  $Q_{Sar}$  at increasing values of  $K$  in the cumulative sample including all other populations in the data set, and selected the value of  $K$  at which this contribution levelled off (that is,  $K=12$ ). In doing so, we minimised the possible noise due to an incomplete partitioning of Sardinian vs non-Sardinian populations. We subsequently investigated the distribution of  $q_{Sar}$  among WSar individuals at the selected  $K$  ( $K=12$ ) and ordered individuals by  $q_{Sar}$  (from the purest to the most introgressed).

As both diversity measures and detected levels of differentiation can be affected by the number and nature of the individuals under study, we recalculated minor allele frequencies,  $H_E$ ,  $H_O$  and pairwise  $F_{ST}$  and repeated the PCA using a subset of 15 individuals (equalling the smallest population) for each group. In so doing, we used three different subsamples of the Sardinian population: a random subsample of 15 individuals (Random), the 15 individuals at the upper extreme (Top) and the 15 at the lower extreme (Bottom) of the distribution of  $q_{Sar}$ . Similarly, a possible bias in the Bayesian analysis due to unequal sample size among populations was addressed by re-running STRUCTURE with the same parameters described above ( $K=1-10$ ) on the random subset of 15 individuals per population.

Runs of homozygosity (ROH) were identified across the genome of Sardinian WB and compared with reference populations. Sliding windows of 10 Kbp (20 SNPs), allowing for one heterozygous call for each window, were checked across the genome using PLINK. To avoid overestimation of ROHs owing to rare allele removal, no additional filtering for low allele frequencies was applied. As ROH analysis is sensitive to sample size and within-population heterogeneity, we repeated it with the same parameters considering only

homogeneous clusters on the basis of *STRUCTURE* results at  $K=12$  and an even sample size (equalling  $N=8$ , that is, the smallest homogenous cluster in our data set), achieved by random sampling of individuals within clusters. For the WSar population, we also calculated ROH for the eight individuals at the two extremes of the  $q_{Sar}$  distribution (Top and Bottom). Differences between the normalised distributions of ROH values from the complete WSar data set and the reduced ones were tested using a two-sample Kolmogorov–Smirnov tests in R 3.0.2 (R Core Team, 2012).

To assess population-specific differences in the spatial distribution of ROHs across the genome, ‘ROH hotspots’ were identified as locations of the genome for which the proportion of individuals showing a ROH ( $f_{ROH}$ ) exceeded the 99th percentile of the distribution of  $f_{ROH}$  in the population (Pemberton *et al.*, 2012). In so doing, we grouped individuals into the four previously used populations (that is, WSar, WIta, WEur and DP). Then we checked the amount of ROH hotspots that were shared by different populations. Finally, for each individual, the number and cumulative size of ROHs in the genome was calculated and results averaged by population. As the chip methodology is likely to underestimate the amount of small ROHs in the genome (Bosse *et al.*, 2012), only ROHs > 10 Mbp were considered in this calculation.

## RESULTS

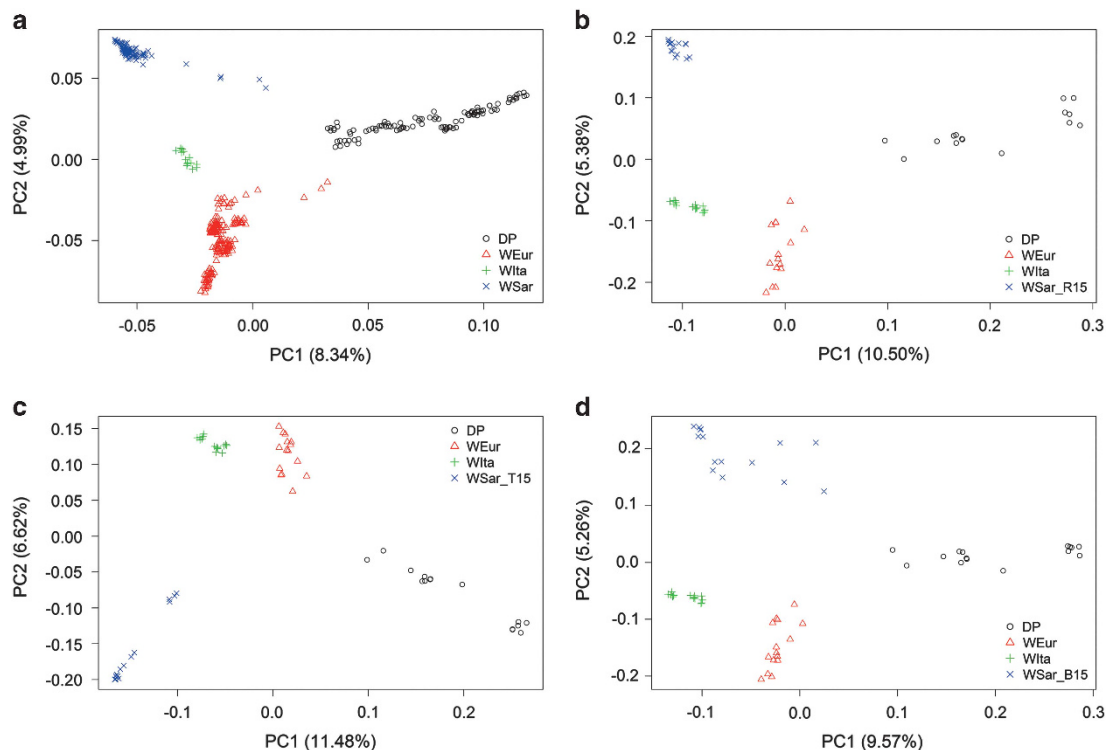
The PCA plot of the 400 individuals analysed with the 30 K SNPs clearly separates WB from DP and delineates a structuring of the WB sample into three major groups: WSar, WIta and WEur (Figure 1a). The first principal component (8.34% of variance) sharply distinguished between DP and WB, except for a small number of possible hybrid individuals. The second component (4.99% of variance) reflects differences among the three wild populations. Thus, the subdivision of our data set into four populations was confirmed.

Out of 49 803 SNPs analysed on the whole available sample, a total of 47 287 (95%) were polymorphic across populations, with average

minor allele frequencies of 0.206. Within the four populations, the number of polymorphic loci ranged from 47 099 (94.6%) in DP to 28 600 (57.4%) in WIta, while the average minor allele frequencies ranged from 0.251 in DP to 0.141 in WSar and WIta (Table 1a). Overall,  $H_E$  ranged from 0.190 in the WIta population to 0.387 in the DP (Table 1a), whereas  $H_O$  was lower than  $H_E$ , ranging from 0.161 (WSar) to 0.251 (DP). For WSar, the heterozygosity deficit implies a significant deviation from Hardy–Weinberg equilibrium ( $P < 0.001$ ), that could possibly be related to a moderately high-inbreeding coefficient ( $F_{IS} = 0.162$ ), but also to a substantial degree of population substructure (see Scandura *et al.*, 2011b).

Pairwise  $F_{ST}$  values, calculated from the 30 K SNPs data set, ranged from 0.091 between WEur and WIta to 0.169 between WSar and DP (Table 2a).  $F_{ST}$  between WSar and WEur amounted to 0.126, whereas it was 0.138 between WSar and WIta.

Concordantly with PCA, Bayesian ancestry inference using the programme *STRUCTURE* highlighted the differentiation of the WSar population (Supplementary Figure S1). Already at  $K=2$  WSar was separated from all other populations, whereas further differentiation among the wild populations, on a geographic base, and among DP breeds was observed at higher  $K$  values. The Evanno method identified in  $K=3$  the uppermost level of population structure separating the WSar population ( $Q_I = 0.956$ ) from both the DP ( $Q_{II} = 0.842$ ) and the other continental WB populations as a whole ( $Q_{III} = 0.814$ ). WIta resulted intermediate between WEur and WSar until  $K=12$ , when it was first recognised as a separate cluster (Figure 2b and Supplementary Figure S1, Supplementary Information). As at this value of  $K$ , the mean  $Q_{Sar}$  in non-Sardinian populations started to stabilise (Figure 2a),  $q_{Sar}$  values at  $K=12$  were adopted to evaluate



**Figure 1** Principal component analysis (PCA) for the 30 K SNPs data set: (a) based on all the available samples; (b) based on a random subset of 15 individuals for each population; (c) based on the 15 Sardinian individuals with the highest  $Q_{Sar}$  and the same random subset of 15 individuals for the other populations used in Figure 1b; (d) based on the 15 Sardinian individuals with the lowest  $Q_{Sar}$  and the same random subset of 15 individuals for the other populations used in Figure 1b. WSar—WB Sardinia, WSar\_R15—Random WB Sardinia, WSar\_T15—WB Sardinia with the highest  $Q_{Sar}$  values, WSar\_B15—WB Sardinia with the lowest  $Q_{Sar}$  values, WIta—WB Italy, WEur—WB Europe, DP—domestic pig.

**Table 1** Indexes of genetic diversity in different WB populations and in domestic pigs, calculated on the complete data set (50 K)

Population	Sample size	N of polymorphic loci	$H_E$	$H_D$	MAF ( $\pm$ s.d.)
<i>(a)</i>					
WB Sardinia	99	38 236	0.1903	0.1605	0.141 ( $\pm$ 0.160)
WB Italy	15	28 600	0.1898	0.1757	0.141 ( $\pm$ 0.163)
WB Europe	183	40 504	0.2367	0.1934	0.178 ( $\pm$ 0.168)
Domestic Pigs	105	47 099	0.3872	0.2513	0.251 ( $\pm$ 0.147)
<i>(b)</i>					
WB Sardinia (Top)	15	24 014	0.1571	0.1406	0.117 ( $\pm$ 0.157)
WB Sardinia (Bottom)	15	35 553	0.2098	0.1812	0.157 ( $\pm$ 0.159)
WB Sardinia (Random)	15	28 403	0.1811	0.1604	0.134 ( $\pm$ 0.160)
WB Italy	15	28 600	0.1898	0.1757	0.141 ( $\pm$ 0.163)
WB Europe	15	32 345	0.2242	0.1877	0.169 ( $\pm$ 0.168)
Domestic Pigs	15	45 297	0.3228	0.2544	0.243 ( $\pm$ 0.150)

Abbreviations: MAF, minor allele frequency; WB, wild boar. (a) Results for all the sampled individuals. (b) Results for a random subset of 15 individuals for each population. For the Sardinian population, values are referred to a random sample (Random) and to the upper (Top) and lower (Bottom) extremes of the  $q$ -value distribution (that is, the most pure and the most introgressed individuals).

**Table 2** Pairwise  $F_{ST}$  values calculated between populations, based on the 30 K data set

	WSar	WIta	WEur	DP
<i>(a)</i>				
WSar	0.000			
WIta	0.138	0.000		
WEur	0.126	0.091	0.000	
DP	0.169	0.138	0.103	0.000
<i>(b)</i>				
WSar	0.000			
WIta	0.138	0.000		
WEur	0.137	0.106	0.000	
DP	0.165	0.153	0.098	0.000

(a) Results for all the sampled individuals. (b) Results for a random subset of 15 individuals for each population.

introgression in the Sardinian sample (Figure 3). Around two-thirds (68%) of individuals in the island had  $q_{Sar} > 0.95$ , but the eight individuals (8%) at the lower extreme of the distribution showed a mean equal to 0.716, suggesting a recent introgression (all values  $< 0.90$ ). The  $q_{Sar}$  in the 15 most upper individuals of the distribution (Top) averaged 0.9997 ( $\pm 0.0005$  s.d.), whereas average  $q_{Sar}$  in the 15 lowest individuals was 0.8064 ( $\pm 0.1512$  s.d.).

In the WB population the autosomes showed an average of  $36.24 \pm 13.16$  (s.d.) ROHs per individual, with an average size of  $11.596 \pm 3.646$  Mbp. ROH size varied from 2.76 to 269.53 Mbp. WSar population had on average  $40.58 \pm 11.65$  ROHs per individual (mean size  $11.509 \pm 13.947$ ). Both average number and size of ROHs in Sardinia were intermediate with respect to other populations, despite having the highest frequency of short fragments (73% of ROHs  $< 10$  Mbp, Figure 4a) and a high number of long runs (0.56% of ROHs  $> 100$  Mbp against a mean of 0.31% in the other WB

populations). Most ROHs were present in a single individual and only a small number could be observed in several animals. ROHs hotspots (falling above the 99th percentile of  $f_{ROH}$ ) in the Sardinian WB were spread across the genome and 62.07% of them were shared by at least another population. The proportion of shared ROHs hotspots varied from 22.83% in WEur to 92.31% in DP (Supplementary Figure S3 in Supplementary Information). As to the number and cumulative size of medium and large-sized ROHs ( $> 10$  Mbp), Sardinian WB genomes showed a very similar distribution to that observed for the other mainland European populations (Supplementary Table S2 in Supplementary Information), with a maximum of 43 ROHs in one individual and a maximum cumulative size of around 1 Gbp. Much higher values were instead obtained for DP samples (up to 1.6 Gbp).

Taking an even sample size for the four populations by random selection of 15 individuals, the observed number of polymorphic sites decreased in all populations but more strongly in the Sardinian one (WSar—26%, WEur—20%, DP—4%). As expected, a general—though smaller—reduction was observed in the other indexes of genetic diversity (Table 1b). The genetic diversity of the Sardinian population changed a lot when the two compositionally extreme groups of individuals were taken into account. The use of subsamples of 15 individuals per population did not significantly alter the clustering into four groups in the PCA plot (Figure 1b–d). However, pairwise distances between the WSar population and the other reference populations changed (especially with WEur; Table 2b).

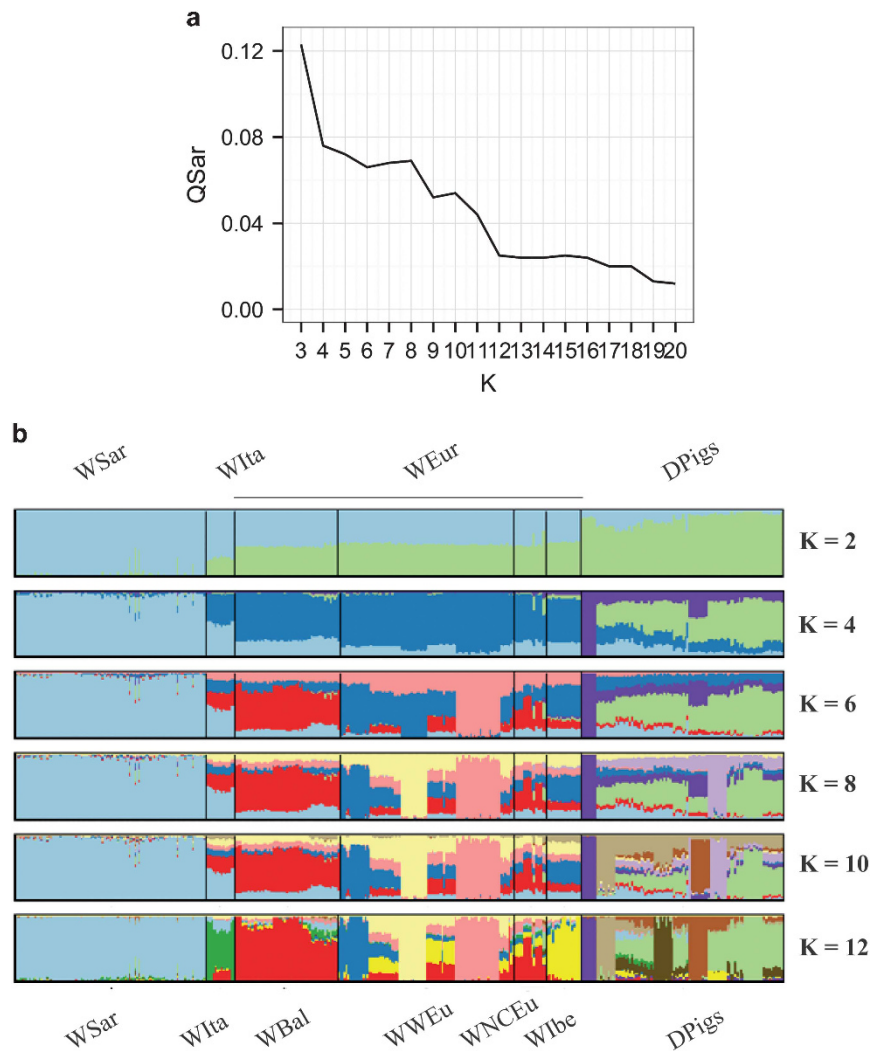
A more substantial change was observed when sample size was reduced to 15 random individuals in the STRUCTURE analysis (Supplementary Figure S2, Supplementary Information). The Evanno method identified  $K=2$  as the most reliable subdivision, separating the DP ( $Q_{II}=0.910$ ) from the WSar and WIta populations, which clustered into one group ( $Q_I=0.970$ ). The WEur was intermediate between them ( $Q_I=0.575$ ). Increasing the number of  $K$  had the immediate effect to distinguish the WSar and WIta populations (from  $K=3$  on) and to split the non-Sardinian populations into more and more clusters roughly corresponding to geographic areas or domestic lines.

As concerns the ROH analysis, among the three WSar subsets, the number of segments per individual increased from admixed individuals (WSar\_B8;  $35.38 \pm 14.99$  s.d.) to those at the opposite extreme of the  $q_{Sar}$  distribution (WSar\_T8;  $49.88 \pm 3.14$  s.d., Figure 4b). An opposite trend was observed for ROH size, that was on average larger in the WSar\_B8 subsample ( $16.943 \pm 20.234$  Mbp) and shorter in the WSar\_T8 subsample ( $11.405 \pm 15.346$  Mbp). The subset composed by admixed individuals was also the only one with a distribution differing from the initial WSar sample ( $D=0.429$ ,  $P=0.042$ ). When sample size was standardized, WSar showed intermediate number and size of ROHs ( $38.25 \pm 8.38$  and  $11.41 \pm 15.35$ , respectively), where as in the reference populations the number of ROHs ranged between 16.88 (WFra) and 47.63 (WIta) and their size ranged between 9.92 (Wibe) and 19.28 (WGer; Figure 4c). A similar number of homozygous segments was observed in DP populations (observed range between 19.38 for Calabrese and 53.63 for Duroc), whereas the length of stretches was seemingly higher (observed range between 12.95 for Duroc and 34.53 for Mora Romagnola Figure 4d).

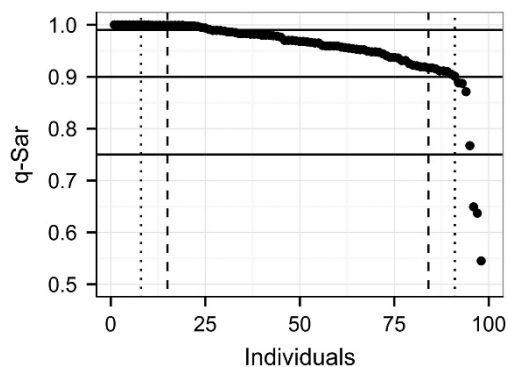
## DISCUSSION

### Genomic distinctiveness and origin of the Sardinian WB

The goal of this study was to assess the genomic differentiation of the Sardinian WB population from its continental counterparts and from DP. Our analyses unambiguously show that, within the genomic



**Figure 2** (a) Distribution of average  $Q_{Sar}$  in non-Sardinian individuals at increasing  $K$ ; (b) plot of individual assignments ( $q_{Sar}$ ) inferred by the software STRUCTURE to the  $K$  clusters ( $K=2-12$ ).

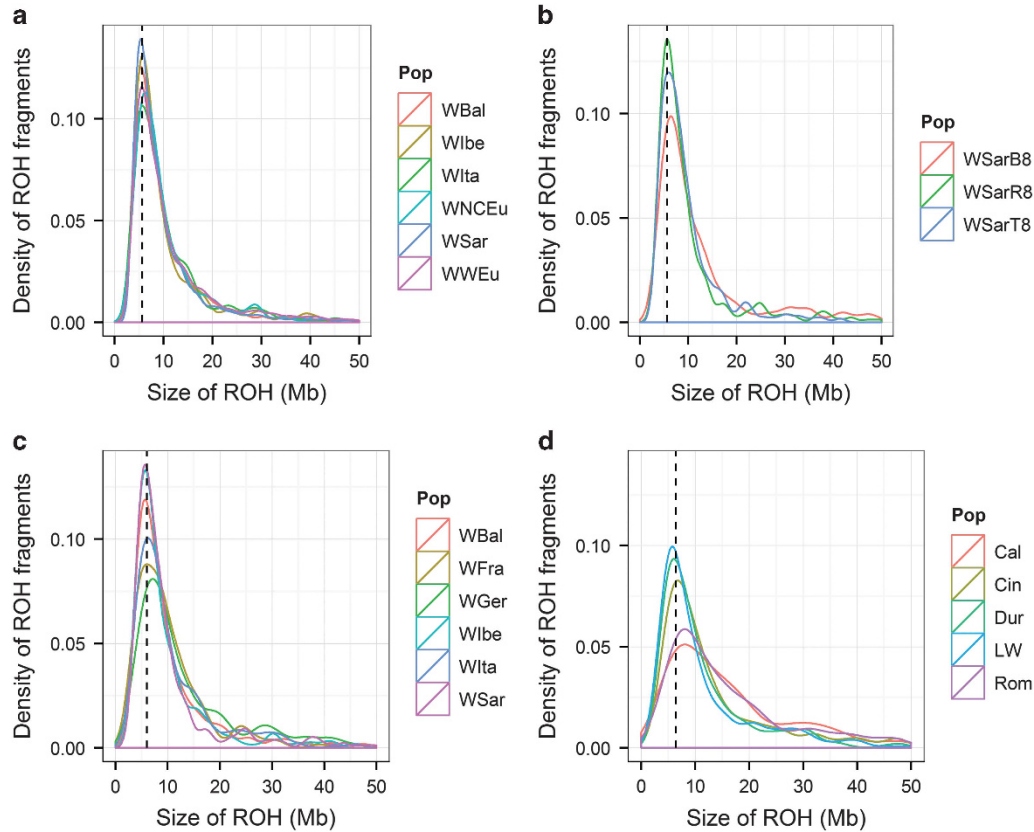


**Figure 3** Decreasing values of  $q_{Sar}$  averaged across the 10 runs at  $K=12$  for the WSar population. The dashed lines indicate the limits of the 15 individuals with highest and lowest  $q_{Sar}$  values. The dotted lines indicate the limits of the eight individuals with highest and lowest  $q_{Sar}$  values.

variation observed in the European porcine population as a whole, the Sardinian wild population forms a highly divergent cluster.

Although a possible ascertainment bias derived from the ascertainment panel of the Illumina porcine SNP60 genotyping beadchip can

lead to an overestimation of the actual divergence between WB and DP, it is considered to have no effect on the inference of the relative divergence among WB populations (see also Goedbloed *et al.*, 2013b). The number of polymorphic sites in the WSar population resulted to be 80.9% of the whole *S. scrofa* sample and increased to 88.5% when only WB were considered, suggesting a marginal influence of the ascertainment bias in the assessment of genetic variation. Therefore, the position of the Sardinian wild population in the PCA plot and the estimate of pairwise  $F_{ST}$  values testify to its significant genomic differentiation from all other European WB. These observations were not influenced by the differences of sample size among populations, as a random selection of a uniform number of individuals, yet implying a strong reduction of sample size in the target populations, produced little changes on diversity parameters (except number of polymorphic loci; Figure 1 and Table 1). However, a non-random selection within the WSar population produced remarkable changes in diversity statistics and  $F_{ST}$  values. It is worth to notice that the strongest deviations (decrease in diversity and increase in genetic divergence) were observed when using the ‘purest’ individuals, suggesting that present levels of admixture could have partially masked the original genetic features of the insular population.



**Figure 4** Density distribution of runs of homozygosity (ROH) in: (a) the complete WB data set; (b) a subsample of WSar individuals ( $N = 8$ ), the two extremes of the  $q_{Sar}$  distribution and a random subsample are shown; (c) a random subsample ( $N = 8$ ) of six genetically homogeneous WB populations; (d) a random subsample ( $N = 8$ ) of five genetically homogeneous DP breeds. WSarB8—WB Sardinia with the lowest  $q_{Sar}$  values, WSarT8—WB Sardinia with the highest  $q_{Sar}$  values, WSarR8—random WB Sardinian individuals, WBal—WB Balkans, WFra—WB France, WGer—WB Germany, Wibe—WB Iberia, WIta—WB Italy, WSar—WB Sardinia, Cal—Calabrese, Cin—Cinta Senese, Dur—Duroc, LW—Large White, Rom—Mora Romagnola.

Bayesian clustering resulting from 30K SNPs strongly supported the distinctiveness of the Sardinian population, which was associated to a private cluster already from the lowest values of  $K$  up to  $K = 17$  (Supplementary Figure S2, Supplementary Information). Furthermore, the Sardinian population appears more differentiated from DP than from other continental wild populations, independently from the sample size of the analysed populations (Table 2). This contrasts with the expectation of extensive hybridisation with local free-ranging DP and could suggest recent introgression from exotic WB, which could have disguised the ‘historical’ pattern of divergence of the island population. Actually, recent introgression from DP was evident only in a limited number of wild individuals that were readily discernible by cluster analyses (Figure 3). Thus, its uniqueness does not seem to be systematically affected by introgression from modern DP. This was true even after the random reduction of sample size, with the WSar population emerging already at  $K = 3$ . However, the effect of sample size was stronger in other populations, like WIta whose genetic structure was underestimated when the full data set was used (Supplementary Figures S1 and S2 in Supplementary Information).

Distinct features of Sardinian WB were initially detected at polymorphic enzymes (Randi *et al.*, 1989). Scandura *et al.*, (2008; 2011b) using 10 autosomal microsatellites, found the Sardinian WB clustering separately in a Bayesian analysis, with a minor overlap to other sampled populations. Mitochondrial data, in addition, showed a high percentage of private sequences in the island (Randi, 1995; Scandura *et al.*, 2008). Overall, the detected molecular and

morphological distinctiveness of the Sardinian WB (Apollonio *et al.*, 1988) is congruent with its classification as a different subspecies (*S. scrofa meridionalis*). This taxon designation is shared by WBs occurring in Corsica, but the present status of the Corsican population is uncertain as no updated genetic data are currently available.

Of all continental WB populations, SNP data show highest genetic affinity of Sardinian WB with the present-day populations of the Italian peninsula (Figure 1 and Supplementary Figure S1 in Supplementary Information). Previous analyses on mtDNA of ancient and modern samples from mainland Italy and Sardinia revealed the occurrence of a private clade (D4 in Larson *et al.*, 2005; E2 in Scandura *et al.*, 2008; 2011a), which showed different haplotypes in the two populations (Scandura *et al.*, 2008). Considering these data jointly, our study strongly supports the hypothesis that the current island population arose from an ancient colonisation from peninsular Italy, probably mediated by humans. Yet, multiple colonisation events from different source populations might have occurred. Furthermore, this hypothesis is supported by the lack of archaeological records before the early Neolithic (Wilkens, 2003; Albarella *et al.*, 2006), whereas a natural colonisation would have most likely occurred during the sea lowering at the last Pleniglacial. However, there is no clear evidence on the status (wild or domestic) of the introduced individuals, as size differences in early specimens with respect to contemporary mainland populations are controversial (Albarella *et al.*, 2006). A possible founder effect, followed by genetic drift under isolation, could explain the present divergence of the Sardinian WB.

But an important role may have been played also by local adaptation, possibly promoted by limited food supply and intra-specific competition, as suggested by an early size reduction during the Neolithic (Albarella *et al.*, 2006).

### High genomic variation in the insular population

The observed SNP variation is within the range estimated for SNP data in other wild species, including bison (*Bison* spp.—Pertoldi *et al.*, 2010), wolf-like canids (*Canis* spp.—VonHoldt *et al.*, 2011) and the Alaskan salmon (*Oncorhynchus nerka*—Gomez-Uchida *et al.*, 2011).

As stated above, even though we cannot compare diversity statistics between DP and WB in absolute values, we can confidently compare those among WB populations (see Bosse *et al.*, 2012; Goedbloed *et al.*, 2013b). The fraction of polymorphic SNPs was relatively high for an island population (76.8% of the total amount found across populations). By comparison, in the whole sample of continental WB (excluding Italy) this proportion amounted to 81.3% (see Table 1). The observed variability was still comparable with that reported for the non-isolated WIta population when a random subset of individuals was analysed (Table 1b). We suggest four possible explanations for such an unexpectedly high variation in this island population: (1) Sardinia was colonized by a large number of individuals; (2) repeated introductions took place from multiple sources; (3) since its origin the island population has maintained a relatively high population size; (4) population substructuring caused by landscape features is present. Actually, the Sardinian population has not undergone heavy demographic fluctuations in the last century, and WBs were abundant on the island even when they had almost disappeared across most of the Italian peninsula (Ghigi, 1950).

Patterns of ROHs help to elucidate which factors could have left a major signature in the genome of the island WB. Interestingly, the WSar population showed the highest number of short (<10 Mbp, Figure 4a) and a high number of very long ROHs (>100 Mbp). A random reduction of the sample size did not affect these results; however, levels of autozygosity assessed by ROHs differed when the individuals with lowest  $q_{Sar}$  were considered, producing a lower number of segments per individual, as expected in presence of recent hybridisation events. Short ROHs may derive from ancient bottlenecks (like in case of a founding event) and can be maintained through time by a low  $N_e$ . The evidence that many high-frequency ROHs in WSar were shared by other WB and DP populations might suggest an ancestral origin and a possible signature of positive selection on these homozygous regions (Pemberton *et al.*, 2012), although a role of introgression cannot be excluded. Conversely, long ROH are sensitive to recent population changes (Bosse *et al.*, 2012) and their presence suggests that groups of inbred individuals are likely to be present in the island. In fact, although the fraction of the genome occupied by ROHs was similar to continental populations, a few individuals showed an exceedingly high number and size of ROHs (Supplementary Figure S4 in Supplementary Information). As some of these animals either belong to a previously identified isolated subpopulation or show relatively low  $q_{Sar}$  values, their ROHs may derive from low  $N_e$  in local demes or from the release/escape of introgressed individuals from inbred captive stocks (see also Canu *et al.*, 2014). Overall, this pattern is suggestive of a combination of past demographic events (bottlenecks) and a more recent natural or artificial genetic substructuring in the Sardinian population (see Scandura *et al.*, 2011b).

Concluding, despite a certain level of recent introgression from both wild and domestic populations, the Sardinian WB still shows significant divergence and distinctiveness at both mitochondrial and

nuclear loci. Accordingly, genetic data would support its, representing an evolutionarily significant unit, although field studies are needed to test its ecological exchangeability (Crandall *et al.*, 2000). Actually, there is a general lack of data on the biology and ecology of the nowadays Sardinian WB, which strongly limits a full assessment of its conservation value.

Further investigations, applying complete genome sequencing, including ancient Sardinian and Corsican samples, will be useful to address outstanding questions on the origin and evolution of the populations inhabiting these two islands. In addition, further investigations are needed to address the genetic basis and adaptive relevance of their phenotypic distinctiveness, as long with a possible variation due to population substructuring.

### DATA ARCHIVING

The 49 803 autosomal SNP genotypes for 295 WBs and 105 DP (PLINK and STRUCTURE file format) were deposited in the Dryad data repository: doi:10.5061/dryad.8bf48

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

A financial support was provided as research funding to MA (CRP1\_415) and personal grant to LI (CRP2\_384) by the Sardinian Regional Government (LR 7/2007 Promozione della Ricerca Scientifica e dell'Innovazione Tecnologica in Sardegna), the Banco di Sardegna Foundation provided research funding to MS (2013.1315). LI has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 656697. Genotyping was supported by USDA grant 2007-04315 and the EU SABRE project FOOD-CT-2006-01625. RPMAC, H-JM and MAMG have received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) ERC Grant agreement no. 249894. We thank Alexandros Triantafyllidis, Miguel Pérez-Enciso and Mario Barbato for valuable comments. We are also grateful to Pekka Uimari, Peter Dovč, Deodália Dias, Alain Ducos, Alexandros Triantafyllidis and Miguel Pérez-Enciso for providing European wild and domestic samples.

- 
- Abraham G, Inouye M (2014). Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**: e93766.
- Albarella U, Manconi F, Rowley-Conwy P, Vigne J-D (2006). Pigs of Corsica and Sardinia: a biometrical re-evaluation of their status and history. In: Tecchiati U, Sala B (eds). *Archaeozoological studies in honour of Alfredo Riedel*. Ufficio Beni Archeologici: Bologno, Italy, pp 285–302.
- 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, Scandura M, Sprem N (2014). Reintroductions as a management tool for European Ungulates. In: Putman R, Apollonio M (eds). *Behaviour and Management of European Ungulates*. Whittles Publishing: Dunbeath, Scotland, UK.
- Bosse M, Megens HJ, Madsen O, Paudel Y, Frantz LAF, Schook LB *et al.* (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet* **8**: e1003100.
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* **18**: 249–256.
- Canu A, Costa S, Iacolina L, Piatti P, Apollonio M, Scandura S (2014). Are captive wild boar more introgressed than free-ranging wild boar? Two case studies in Italy. *Eur J Wildl Res* **60**: 459–467.
- Crandall KA, Bininda-Emonds ORR, Mace GM, Wayne RK (2000). Considering evolutionary processes in conservation biology. *Trends Ecol Evol* **15**: 290–295.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Ghigi A (1950). La vita degli animali. In: *II Mammiferi ed Uccelli sedentari delle terre continentali*. U.T.E.T.: Milano, Italy. (In Italian).
- Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P *et al.* (2013a). Genome-wide single nucleotide polymorphism analysis reveals recent genetic

- introgression from domestic pigs into northwest European wild boar populations. *Mol Ecol* **22**: 856–866.
- Goedbloed DJ, Van Hooff P, Megens HJ, Langenbeck K, Lutz W, Crooijmans RPMA *et al.* (2013b). Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boars. *BMC Genetics* **14**: 43.
- Gomez-Uchida D, Seeb JE, Smith MJ, Habicht C, Quinn TP, Seeb LW (2011). Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evol Biol* **11**: 48.
- Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogdén R, Limborg MT *et al.* (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Res* **11**: 123–136.
- Kumar S, Skjæveland T, Orr RJS, Enger P, Ruden T, Mevik B *et al.* (2009). Air: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* **10**: 357.
- Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J *et al.* (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**: 1618–1621.
- Major F (1883). Studien zur Geschichte der Wildschweine (Gen. Sus). *Suppl Zool Anz* **140**: 295–300.
- Masseti M, Mazza R (1996). Is there any paleontological "Treatment" for the "Insular syndrome"? *Vie Milieu* **46**: 355–363.
- Megens H-J, Crooijmans RPMA, San Cristobal M, Hui X, Li N, Groenen MAM (2008). Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. *Genet Sel Evol* **40**: 103–128.
- Morin PA, Luikart G, Wayne RK (2004). SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.
- Pemberton TJ, Absher D, Feldman WM, Myers RM, Rosenberg NA, Li JZ (2012). Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* **91**: 275–292.
- Pertoldi C, Wójcik JM, Tokarska M, Kawałko A, Kristensen TN, Loeschcke V *et al.* (2010). Genome variability in European and American bison detected using the BovineSNP50 beadchip. *Conserv Genet* **11**: 627–634.
- Pilot M, Greco C, von Holdt BM, Jedrzejewska B, Randi E, Jedrzejewski W *et al.* (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity* **112**: 428–442.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575.
- R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE *et al.* (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* **4**: e6524.
- Randi E (1995). Conservation genetics of the genus *Sus*. *Ibex* **3**: 6–12.
- Randi E, Apollonio M, Toso S (1989). The systematics of some Italian populations of wild boar (*Sus scrofa* L.): a craniometric and electrophoretic analysis. *Z Säugetierkd* **54**: 40–56.
- Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- Rousset F (2008). Genepop'007: a complete re-implementation of the genepop software for windows and linux. *Mol Ecol Res* **8**: 103–106.
- Scandura M, Iacolina L, Apollonio M (2011a). Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal Rev* **41**: 125–137.
- Scandura M, Iacolina L, Cossu A, Apollonio M (2011b). Effects of human perturbation on the genetic make-up of an island population: the case of the Sardinian wild boar. *Heredity* **106**: 1012–1020.
- Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V *et al.* (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? *Mol Ecol* **17**: 1745–1762.
- Van Zyl T, Jerling JC, Conradie KR, Feskens EJM (2014). Common and rare single nucleotide polymorphisms in the LDLR gene are present in a black south African population and associate with low-density lipoprotein cholesterol levels. *J Hum Genet* **59**: 88–94.
- Vigne J-D, Zazzo A, Saliège J-F, Poplin F, Guilaine J, Simmons A (2009). Pre-Neolithic wild boar management and introduction to Cyprus more than 11,400 years ago. *Proc Natl Acad Sci USA* **106**: 16135–16138.
- Vilaça ST, Bioss D, Zachos F, Iacolina L, Kirschning J, Alves PC *et al.* (2014). Mitochondrial phylogeography of the European wild boar: the effect of climate on genetic diversity and spatial lineage sorting across Europe. *J Biogeogr* **41**: 987–998.
- VonHoldt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H *et al.* (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res* **21**: 1294–1305.
- White S (2011). From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. *Environ Hist* **16**: 94–120.
- Wilkins B (2003). La fauna sarda durante l'Olocene: le conoscenze attuali. *Sardinia, Corsica et Baleares Antiquae* **1**: 181–197. (In Italian).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)