Journal of Biogeography (J. Biogeogr.) (2017) **44**, 259–270

ORIGINAL ARTICLE

Distinguishing migration events of different timing for wild boar in the Balkans

Panoraia Alexandri[1,*], Hendrik-Jan Megens[2], Richard P. M. A. Crooijmans[2], Martien A. M. Groenen[2], Daniel J. Goedbloed[3], Juan M. Herrero-Medrano[2], Lauretta A. Rund[4], Laurence B. Schook[4], Evangelos Chatzinikos[5], Costas Triantaphyllidis[1] and Alexander Triantafyllidis[1]

[1]*Department of Genetics, Development and Molecular Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Macedonia, Greece,* [2]*Animal Breeding and Genomics Centre, 6700AH Wageningen, The Netherlands,* [3]*Braunschwein Zoological Institute, Braunschweig, Germany,* [4]*Laboratory of Comparative Genomics, University of Illinois, Urbana, IL 61801, USA,* [5]*4th Hunting Federation of Sterea Hellas, 10563 Athens, Greece*

*Correspondence: P. Alexandri, Department of Genetics, Development and Molecular Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Macedonia, Greece.
E-mail: palexand@bio.auth.gr

ABSTRACT

Aim We compared the power of different nuclear markers to investigate genetic structure of southern Balkan wild boar. We distinguished between historic events, such as isolation in different refugia during glacial periods, from recent demographic processes, such as naturally occurring expansions.

Location Southern Balkans/Greece.

Methods We sampled 555 wild boars from 20 different locations in southern Balkans/Greece. All individuals were analysed with 10 microsatellites and a subgroup of 91 with 49,508 single nucleotide polymorphisms (SNPs). Patterns of genetic structure and demographic processes were assessed with Bayesian clustering, linkage disequilibrium and past effective population size estimation analysis.

Results Both microsatellite and SNP data analyses detected genetic structure caused by historic events and support the existence of three groups in the studied area. A hybrid zone between two of the groups was also detected. We also showed that genome-wide SNP data analysis can identify recent events in bottlenecked populations.

Main conclusions We inferred the three groups diverged ~50,000–10,000 YR BP when populations contracted to different refugia. Our findings strengthened the evidence that the southern Balkan area was a glacial refugium including further local smaller refugia. Genome-wide genotyping inferred a recent population expansion that can mimic a 'refugium within refugium' scenario. It seems that microsatellite data tend to overestimate genetic structure when genetic drift happens in bottlenecked populations over a short distance. Therefore, genome-wide SNPs are more powerful at inferring phylogeography in natural populations, resolving inconsistencies from mitochondrial and microsatellite data sets.

Keywords
genetic structure, glacial period, Greece, microsatellites, recent migration, single nucleotide polymorphisms, Southern Balkans, *Sus scrofa*, wild boar

INTRODUCTION

Phylogeographic research suggests that genetic variation patterns within and among closely related species carry the signature of the species' demographic past (Knowles, 2009). This is particularly true for European temperate animal species whose distribution was affected by Pleistocene climatic changes. Most of these species survived the Last Glacial Maximum (LGM) in Mediterranean refugia, spreading northwards when climate conditions improved. Glacial refugia have been found in the Iberian Peninsula, Italy, the Balkans and the Caucasus region (Hewitt, 2000). Isolation in refugia resulted in the evolution of unique gene pools that can be detected in phylogeographical patterns and genetic

Journal of Biogeography

differentiation of many species (Avise, 2000). Main post-glacial migration scenarios involve recolonization of central and northern European regions from Mediterranean refugia. This hypothesis is supported by the fossil record (Sommer & Zachos, 2009), although some exceptions have been observed (Stewart & Lister, 2001).

In order to explain the distinct phylogeographic lineages found in some of the Mediterranean areas, the hypothesis of 'refugia within refugia' was also proposed. This hypothesis considers the geographical terrain and the climatic influences during the LGM responsible for the evolution of distinct phylogeographic subgroups (Gomez & Lunt, 2007) and has been observed mostly using mitochondrial DNA.

The Balkan region is physiographically complex, with mountain ranges reaching the alpine zone separated by canyons and valleys. The Pindus mountains, one of the largest and highest mountain ranges in the southern Balkans, was glaciated during Pleistocene cold stages (Hughes *et al.*, 2007) isolating populations, and distinct phylogeographic groups of small mammals, reptiles and amphibians evolved (Sotiropoulos *et al.*, 2007; Ursenbacher *et al.*, 2008). Most of these studies, however, are based on mitochondrial DNA analysis, with rare examples of nuclear marker results for large mammals (e.g. Nikolov *et al.*, 2009). SNPs are the most abundant type of variation in genomes, can easily and cheaply be genotyped in large numbers simultaneously. Assays are readily comparable between laboratories and cover a more representative sample of the entire genome (Morin *et al.*, 2004). SNPs have proven valuable for population studies in human (e.g. Pemberton *et al.*, 2012) and domestic animals (e.g. Kijas *et al.*, 2012) but their use in wild species remains limited (e.g. Gedbloed *et al.*, 2013a,b; Kraus *et al.*, 2013; Iacolina *et al.*, 2016).

Genetic structure of European wild boar populations was also influenced by the last glaciations and a number of refugial populations have been recognized (Scandura *et al.*, 2008). Mitochondrial DNA has shown that the Balkan refugium was important for wild boar expansion to Eastern and Central Europe, and that Southern Balkans still have geographically confined mitochondrial haplotypes (Alexandri *et al.*, 2012). Different wild boar haplotype groups date back to the LGM and a rapid northward expansion from the northernmost edge of the refugium ('leading edge expansion' scenario, Hewitt, 1999), was proposed. However, some aspects of post-glacial expansion, such as the existence of hybrid zones cannot be fully captured by mitochondrial DNA analysis alone. Microsatellites have been used extensively to infer the genetic structure of wild boar populations and to detect individuals with admixed ancestry (Vernesi *et al.*, 2003; Scandura *et al.*, 2008, 2011; Nikolov *et al.*, 2009).

In this study we investigate the genetic structure of European wild boar at one of its southernmost areas of expansion. We used extensive sampling, microsatellite DNA and genome-wide SNP data to estimate the impact of the last glaciation and test the 'refugia within refugium' hypothesis on a large, widespread and iconic mammalian species. Moreover, we show that genome-wide analysis of variation is a better indicator of long-term demography and range expansion than classic microsatellite analysis, providing important details on the biogeography of local populations.

MATERIALS AND METHODS

Population sampling, microsatellite and SNP genotyping

Muscle and liver samples from 555 wild boars were collected from 18 different regions within continental Greece, one in Bulgaria and one in the eastern Aegean island of Samos (Table 1, Fig. 1, see Appendix S1 in supporting information). All samples came from native populations with the exception of Peloponnesus where wild boar was recently reintroduced. DNA was extracted from muscle tissue according to a CTAB (cetyl trimethyl ammonium bromide) protocol (Hillis *et al.*, 1996).

All individuals were genotyped with a panel of ten microsatellite loci: SW055, SW461, SW841, SW1492, SW2021, SW2496, SW2532, SW951, S0225 and S0227. Three microsatellites (SW951, S0225 and S0227) are recommended by the Food and Agriculture Organization (FAO) for pig diversity analysis (Barker *et al.*, 1998). The rest have been successfully used for wild boar population genetics studies (Vernesi *et al.*, 2003). PCR amplification protocols are

Table 1 Sampling sites in southern Balkans-Greece for 555 wild boar individuals, which were used for microsatellite and SNP analysis.

	Region	Number of samples (microsatellites)	Number of samples (SNPs)
1.	Aitoloakarnania (AI)	64	4
2.	Evritania (EV)	40	3
3.	Fokida (FK)	58	4
4.	Fthiotida (FT)	60	6
5.	Voiotia (VI)	47	4
6.	Larisa (LR)	19	4
7.	Magnisia (MG)	6	2
8.	Trikala (TR)	5	3
9.	Ioannina (IO)	26	4
10.	Grevena (GR)	4	-
11.	Pella (PL)	7	3
12.	Thessaloniki (TH)	6	4
13.	Halkidiki (HL)	8	2
14.	Drama (DR)	55	5
15.	Kavala (KB)	25	5
16.	Rodopi (RO)	38	7
17.	Evros (EB)	54	7
18.	Samos (SM)	18	13
19.	Peloponnesus (PN)	10	6
20.	Bulgaria (BLG)	5	5
	Total	555	91

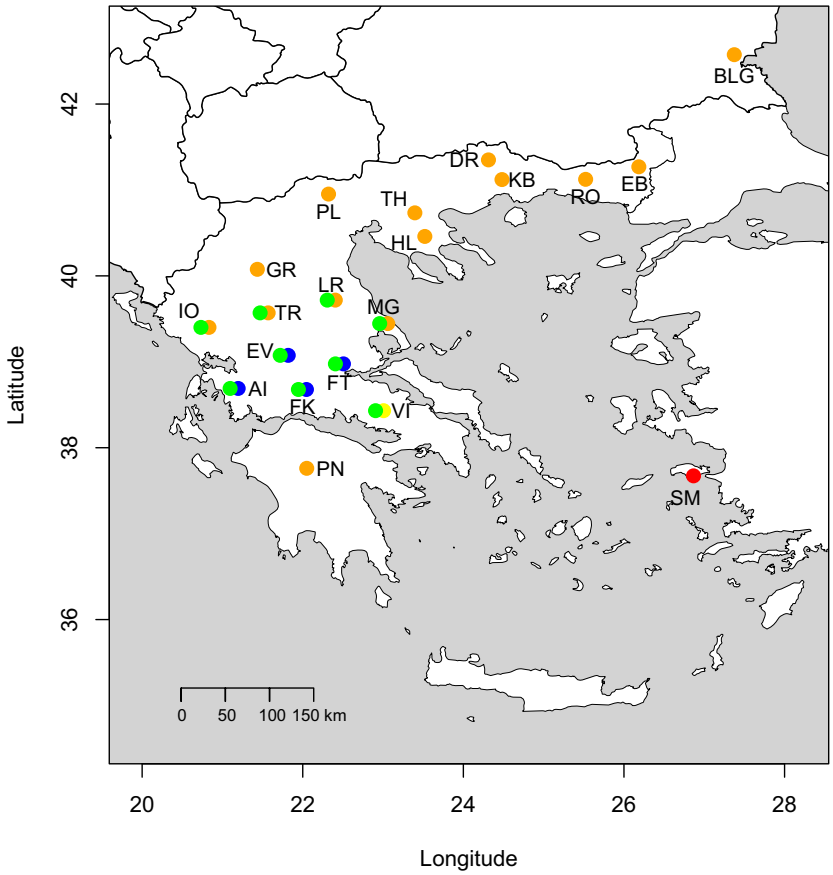

Figure 1 Map of southern Balkans showing sampling sites of 555 wild boars (*Sus scrofa*). Coloured dots correspond to different groups identified with microsatellite and SNP analyses. Orange dots: North group, green: central, red: Samos. Blue, green and yellow dots in areas AI, EV, FK, FT, and VI correspond to subgroups identified in central Greece with microsatellite Structure analysis [Colour figure can be viewed at wileyonlinelibrary.com].

according to Scandura *et al.* (2008). Microsatellite alleles were sized on a 6.5% polyacrylamide gel on a LICOR 4200 automated sequencer (LICOR, Nebraska, USA) using the Saga^{GT} software (LICOR).

Within the Porcine HAPMAP project wild and domestic *Suids* were genotyped with the Porcine SNP60 Beadchip (Amaral *et al.*, 2009; Ramos *et al.*, 2009). For our study, 91 individuals from different areas (Table 1) were analysed for 49,508 autosomal SNPs. Individuals were genotyped according to manufacturer's instructions. All reported samples displayed call rates above 0.9. To test for domestic pig introgression, we used 113 pigs from four commercial breeds occurring in Greece (Duroc $N = 31$, Landrace $N = 27$, Large White $N = 34$ and Pietrain $N = 21$).

Microsatellite data analysis

Observed and expected heterozygosities and allele diversities were computed with Arlequin 3.5 (Excoffier & Lischer, 2010). Allelic richness and private allelic richness were calculated per population using Adze 1.0 (Szpiech *et al.*, 2008). Deviations from Hardy–Weinberg equilibrium (HWE) were tested across each population and locus with Guo & Thompson (1992) test in Genepop 4.2 (Rousset, 2008) using the Markov chain algorithm with 10,000 dememorization steps, 100 batches and 1000 iterations. Mean number of alleles and inbreeding coefficient (F_{IS}) values were estimated with Genetix 4.05 (Belkhir *et al.*, 2001). Genetic differentiation between different populations was estimated by calculating pairwise F_{ST} values in Arlequin.

We estimated existence of possible subpopulations with Structure 2.3.3 (Pritchard *et al.*, 2000; Hubisz *et al.*, 2009) using a number of subpopulations (K) between 1 and 10. We discarded any previous population information and used the admixture model, correlated allele frequencies and ten independent runs of 100,000 iterations after a burn-in of 50,000 iterations for each K. The optimal K value was chosen according to Evanno *et al.* (2005). Isolation by distance for central Greek populations was tested using linear distances for Mantel's test (Mantel, 1967), from the comparison of all pairwise $F_{ST}/(1 - F_{ST})$ values computed in Arlequin.

SNP data analyses

To estimate the extent of domestic pig introgression and identify possible hybrids between wild and domestic individuals, we performed an allele spectrum frequency assessment (AFSA, Goedbloed *et al.*, 2013a).

Heterozygosity, allele diversity and F_{ST} values between different populations were calculated using all 49,508 loci with Arlequin. Population structure was assessed using Structure. Ten independent runs of 20,000 iterations following a burn-in of 15,000 were done for each K between 1 and 5. We discarded prior population information and assumed

independent loci and admixture between different clusters. The ΔK method (Evanno *et al.*, 2005) was used to calculate the optimal K. Principal component analysis (PCA) was done with EIGENSTRAT (Price *et al.*, 2006) to examine similarities between wild populations. Isolation-by-distance scenario for the central Greek cluster was tested using the same approach as with microsatellite data.

We used the SNP genotype data to identify consecutive homozygous regions (runs of homozygosity, ROHs), which can be a sign of recent demographic events. We used PLINK 1.07 (Purcell *et al.*, 2007) with adjusted parameters: –homozyg-density 1000, –homozyg-window-het 1, –homozyg-kb 10, –homozyg-window-snp 20. To avoid overestimation of homozygous regions due to rare allele removal, we did not filter data for low allele frequencies. Differences of ROHs between various populations were tested with the χ^2 test of proportions and goodness-of-fit in R 3.2.4 (R Core Team, 2016). To check for recent inbreeding the correlation between ROHs size and longitude was tested with Pearson's product moment correlation in R.

To estimate linkage disequilibrium (LD) patterns between different populations, we excluded SNPs deviating from HWE ($P < 0.001$) and with MAF lower than 0.05. LD (r^2) was estimated for all marker pairs less than 3 Mb apart for each chromosome with HAPLOVIEW 4.2 (Barrett *et al.*, 2005). Effective population sizes for each group were estimated using the equation (McEvoy *et al.*, 2011): $r^2 = {}^1/_{(4N_e c+2)}$ where r^2 is the LD, c is the distance between markers in Morgans and N_e the effective population size. Past *Ne* at generation T was calculated according to Hayes *et al.* (2003) $T = 1/2c$. To account for different recombination rate across porcine chromosomes (Bosse *et al.*, 2012), an average recombination map (Tortereau *et al.*, 2012) was used. *Ne* estimates were obtained by averaging multiple genomic regions (Stumpf & McVean, 2003): chromosomes were divided into 1 Mb bins containing recombination rate information and average r^2 for all SNP pairs.

RESULTS

Domestic pig introgression

Allele Spectrum Frequency Assessment analysis, performed to detect possible domestic pig introgression (based only on SNP genotyping information), identified eight out of 91 analysed individuals as possible hybrids because they displayed raised levels of low MAF SNPs (Appendix S2). These individuals were omitted from all further analyses.

Population structure analysis

Initial STRUCTURE analysis using the entire microsatellite data set ($n = 547$) identified two groups (Fig. 2a, Appendix S3a). The first group included individuals

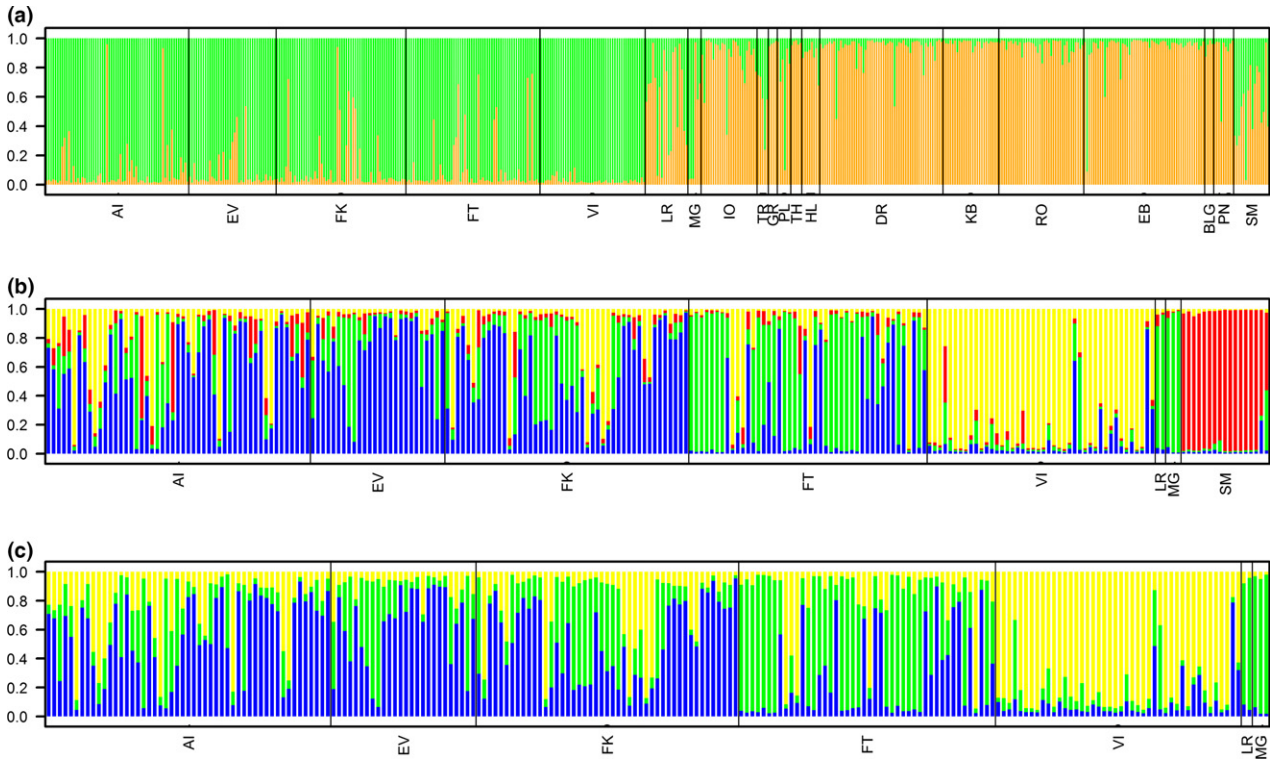

Figure 2 STRUCTURE results showing the most likely number of wild boar populations in southern Balkans-Greece based on microsatellite analysis. Colours correspond to each of the identified groups. (a) for all wild boar samples: orange = northern group, green = rest of samples (b) red = Samos, green – blue – yellow = subgroups identified within central Greece (c) only for central Greek samples [Colour figure can be viewed at wileyonlinelibrary.com].

originating from northern Greece (assigned with probability values $Q > 0.8$). The second group included all other individuals, the majority of which (251 out of 293) were assigned with $Q > 0.7$. Most of the individuals with lower Q values originated from the areas between northern and central Greece (orange and green double dots, Fig. 1). Animals from Peloponnesus clustered with the northern group confirming that this population originated from translocated northern Greek wild boars.

Second step of STRUCTURE analysis was done excluding the northern samples. The optimal number of clusters was $K = 4$ (Fig. 2b). The most prominent cluster is located in Samos Island (individuals assigned with Q~1). The other three clusters included individuals from central Greece. When the analysis was repeated with only the central Greece individuals the same three, geographically confined clusters were recognized (Fig. 2c). The largest cluster expanded throughout western parts of central Greece (areas AI, EV and FK, Fig. 1). The second cluster was central expanding eastwards (area FT), while the third was found exclusively at Voiotia (area VI), the easternmost expansion point of the central group.

Initial level of clustering using SNP data recognized two, well defined groups (Fig. 3a, Appendix S3d): Samos and continental Greece. When only the continental individuals were examined, we confirmed separation of the northern and central individuals ($K = 2$, Fig. 3b). All samples that originated from the areas geographically between central and northern Greece were scattered between these two groups (e.g. samples from IO, LR, MG and TR). PCA analysis agreed with these results (Fig. 4). SNP analysis, however, did not confirm the subpopulation structure discovered with microsatellites within central Greece.

When we tested the geographical patterning at central Greece, using F_{ST} estimates and Mantel tests, both microsatellite and SNP data showed a positive correlation between geographical and genetic distance ($r^2 = 0.114$, $P = 0.001$ and $r^2 = 0.314$, $P = 0.001$ respectively).

Genetic variability within groups

Of the two continental groups discovered with STRUCTURE microsatellite analysis, the central had the lowest mean number of alleles and allelic diversity values (Table 2). Moreover, allelic diversity, allelic richness and observed heterozygosity tended to decline towards the eastern boundaries of this group's expansion (Table 3). F_{IS} values were positive for the central group, indicating more homozygotes than expected (Table 2). The northern group had lower F_{IS} and higher heterozygosity values than the central and they did not show any specific pattern among sampling sites.

From the 49,508 SNPs mapped to autosomes in pig genome version 10.2 (http://www.ensembl.org/Sus_scrofa/Info/Index), 35,528 were polymorphic. The northern group had the highest number of polymorphic loci (33,996), followed by the central (27,242) and Samos (19,765 SNPs).

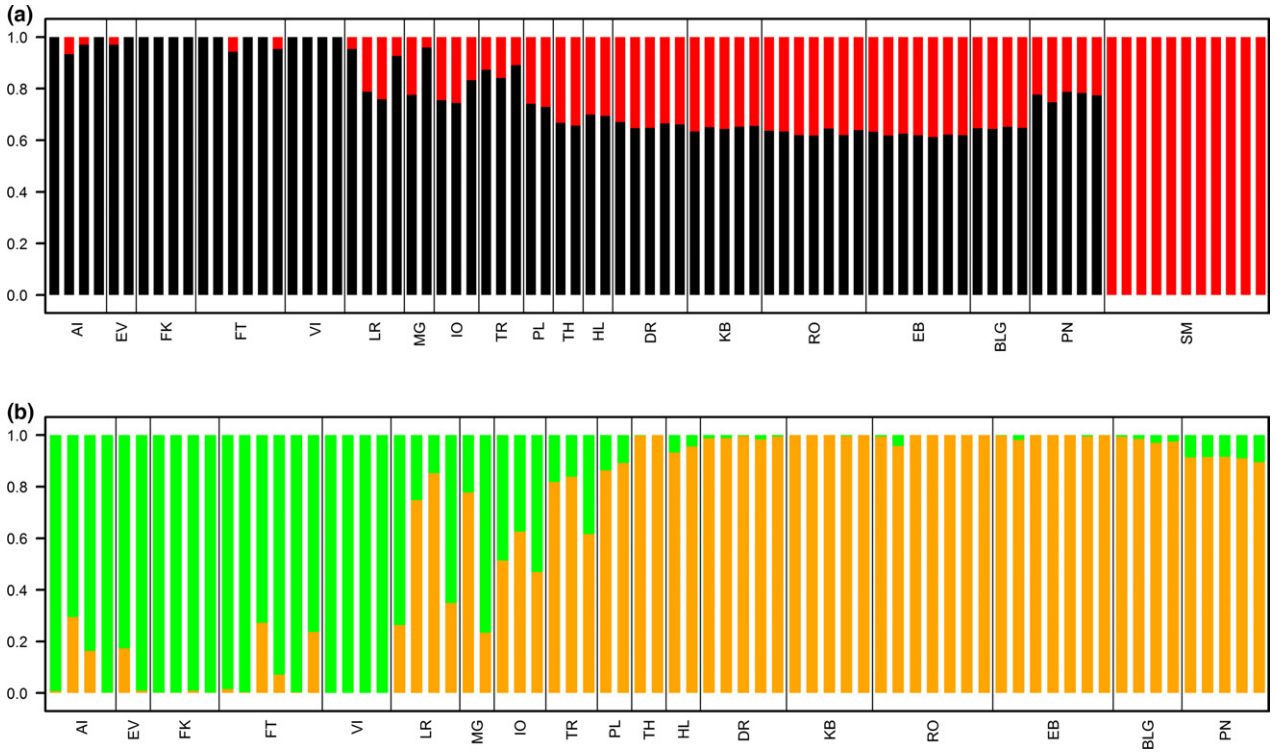

Figure 3 STRUCTURE results showing the most likely number of wild boar populations in southern Balkans-Greece based on SNP analysis. (a) all samples: red = Samos Island, black = continental Greece (b) for continental Greek samples: green = central, orange = northern group [Colour figure can be viewed at wileyonlinelibrary.com].

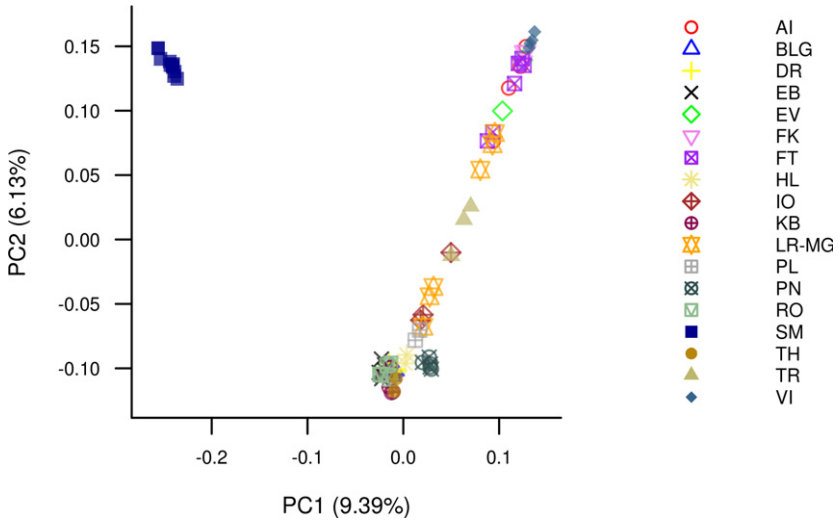

Figure 4 Different wild boar population groups in the Balkans-Greece defined with PCA analysis of genome-wide SNP data. Different individual symbols correspond to different areas of sampling [Colour figure can be viewed at wileyonlinelibrary.com].

Table 2 Heterozygosities, mean number of alleles, allelic range and F_{IS} values for identified wild boar populations in southern Balkans-Greece. *N* = sample size, *Ho* = observed heterozygosity, *He* = expected heterozygosity, *M* = mean number of alleles, *R* = allelic range, *D* = allelic diversity.

	Microsatellite analysis							SNP analysis		
Population	*N*	*Ho*	*He*	*M*	*R*	*D*	F_{IS}	*N*	*He*	*Ho*
Central	238	0.531	0.602	6.7	9.8	0.577	0.100	22	0.325	0.267
North	187	0.579	0.631	9.9	12.2	0.550	0.034	27	0.341	0.312
Continental	524*	0.557	0.670	11.5	13.4	0.602		66*	0.325	0.265
Samos	17	0.638	0.618	4.7	7.5	0.639	−0.049	11	0.362	0.366

*Includes only wild boars from natural populations and from the hybrid zone between north and central groups.

Heterozygosity values were again higher for the northern group (Table 2). Within the central Greece, however, there were little differences from west to east (Table 3).

Genetic diversity measured as effective population size (*Ne*) was considerably higher in the northern (*Ne* = 262, SD = 84.8) than the central (*Ne* = 76, SD = 41.27) Greece. When three different possible subgroups, congruent with the microsatellite analysis, were taken into account in the central population, *Ne* values decreased towards the east (AI-EV-FK = 66, FT = 46, VI = 25).

Runs of homozygosity, LD patterns and historical population sizes

To investigate how much recent demographic parameters affected the genomic distribution of ROHs we analysed 77 pure wild boar individuals from different areas (non-native Peloponnesus samples were excluded). Samples were grouped based on geographical origin and STRUCTURE assignment into three groups: Samos, central and northern Greece. The proportion of the genome in ROH of different lengths varied significantly ($P = 0.0005$) across different parts of Greece. The central group had the highest number (47.75) and longest stretches of ROHs (29.22% of ROHs >15 Mb), whereas individuals from northern Greece and Samos had the lowest number and lowest cumulative size of homozygous regions (Fig. 5). Within central Greece there was a west to east increase in number and length of ROHs (Fig. 6). In particular, individuals from the area of Voiotia (VI) had the highest number (50–62) and the longest stretches of homozygosity (795–1224 Mb). However, the correlation between ROHs size and longitude for central group was weak and non-significant ($r^2 = 0.236$, $P = 0.290$).

To estimate the results of past demographic processes and evaluate if the identified groups had different ancestral populations, a LD decay analysis was performed. Individuals were assigned to three groups, northern, central Greece and Samos. LD patterns differed between groups (Fig. 7). Mean r^2 values were largest for Samos (0.20), followed by the central group (0.18), while the northern had the smallest mean r^2 value (0.058). SNP pairs with an average distance around 1 Mb had an average r^2 of 0.198 for Samos, 0.176 for central and 0.05 for the northern group. All three groups experienced the most pronounced population decline around 10,000 generations ago (Fig. 8) based on estimated past effective population sizes.

DISCUSSION

Recent advances in next generation sequencing have enabled genome-wide SNP data for wild populations (Goedbloed *et al.*, 2013a; Kraus *et al.*, 2013), but to date few studies (e.g.

Table 3 Genetic diversity for central Greek wild boar populations measured as observed heterozygosity for microsatellites [*Ho*(*m*)] and SNPs [*Ho*(*SNPs*)], *D* = allelic diversity, *R* = allelic richness and P_R = private allelic richness for microsatellite data (*m*). Sampling areas correspond to Table 1 and Fig. 1.

Sampling area	Longitude	*D* (*m*)	*R* (*m*)	P_R (*m*)	*Ho* (*m*)	*Ho* (*SNPs*)
AI	21.14567	0.582	0.143	3.216	0.542	0.369
EV	21.75568	0.570	0.153	3.110	0.599	0.535
FK	22.17796	0.586	0.149	3.200	0.535	0.411
FT	22.23770	0.581	0.110	3.131	0.518	0.365
LR	22.41074	0.565	0.194	2.211	0.525	0.539
VI	23.01018	0.491	0.124	2.827	0.482	0.431

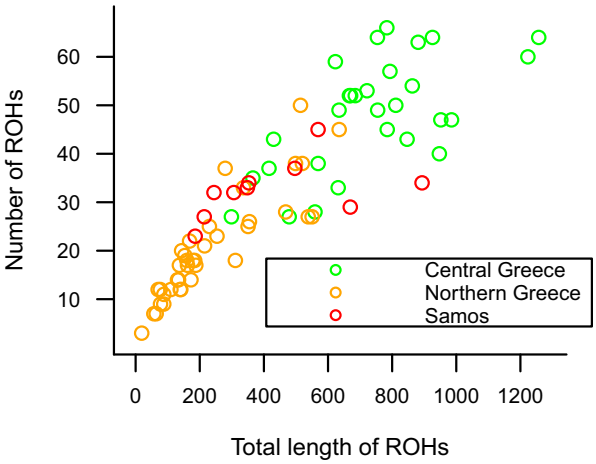

Figure 5 Number and cumulative ROH size for all wild boar individuals based on genome-wide SNP data. Different colours correspond to different groups as defined with STRUCTURE analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

Tokarska *et al.*, 2009; Defaveri *et al.*, 2013; Weinman *et al.*, 2015) have used empirical data to compare SNP with microsatellite results. In this study, we used both markers to assess population structure, past and present expansions of wild boar in southern Balkans and to highlight areas of congruence or disagreement between these markers.

Genetic variability of Greek wild boar: refugial populations and hybrid zones

Both microsatellites and SNPs detected three population groups, each related with a separate geographical area (central, northern Greece and Samos) and reflecting three distinct evolutionary units. These groups were previously recognized with mtDNA (Alexandri *et al.*, 2012), and are now confirmed with multiple markers analysis and more samples.

Two of the three discovered groups pertain different parts of continental Greece: the 'central' group populates central to north-west Greece and the areas close to the Pindus

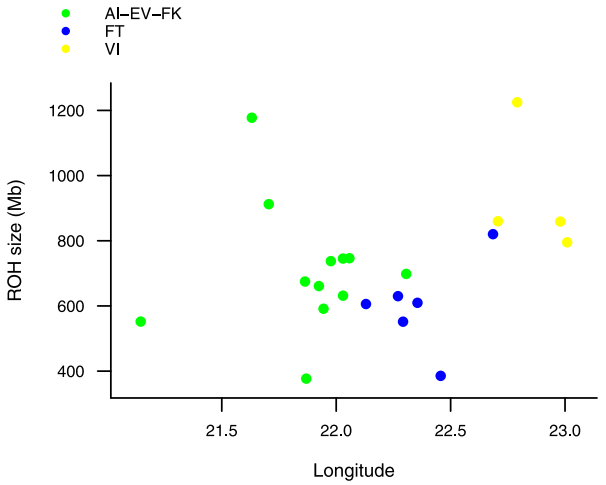

Figure 6 Cumulative size of ROHs per individual plotted against longitude values for wild boars originating from central Greece. Population labels correspond to different sampling sites as well as clusters detected with microsatellite DNA STRUCTURE analysis (Fig. 2c) [Colour figure can be viewed at wileyonlinelibrary.com].

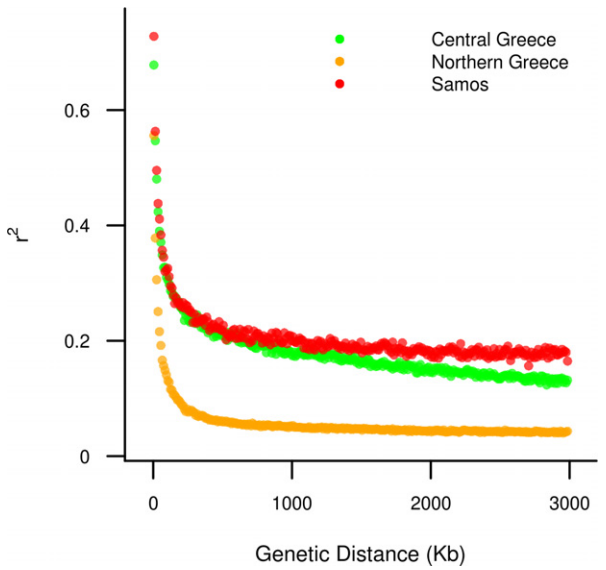

Figure 7 Decay of average r^2 over distance between different wild boar populations in southern Balkans-Greece, for base pairs ranging from 0 Mb to 30 Mb [Colour figure can be viewed at wileyonlinelibrary.com].

mountain range, while the 'northern' expands from the north-eastern part of Pindus to north Greece and Bulgaria. The third group, Samos island, was congruent with the previously described Near Eastern haplogroup (Larson *et al.*, 2005, 2007; Alexandri *et al.*, 2012). The Near Eastern origin of the Samos population was also confirmed by comparison with the Near Eastern group from Anatolia found by Manunza *et al.* (2013) (results not shown).

The implications of our results for the evolutionary history of the wild boar, and for large mammals in general in Europe, are quite profound. The abundance of wild boar fossils

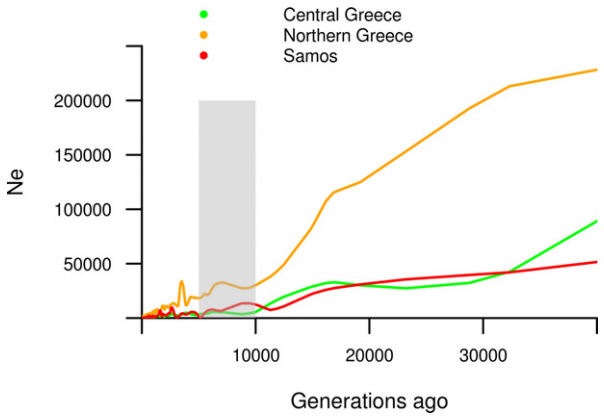

Figure 8 Past Ne estimation for three Greek wild boar groups based on LD analysis. Each chromosome was divided into 1 Mb bins containing recombination rate information and average r^2 for all SNP pairs included in each bin. Grey box indicates Last Glacial Maximum [Colour figure can be viewed at wileyonlinelibrary.com].

in Europe before the Last Glacial Maximum (Vilaça *et al.*, 2014) and the lack of genetic structure during the late Pleistocene demonstrated for other animals (Hofreiter *et al.*, 2004) suggests that a single lineage of wild boar differentiated to three distinct groups in the eastern Mediterranean, within the time span of the LGM. Hence, the separation of the three lineages involves isolation in different refugia. LD analyses support this hypothesis, as the different decay patterns reflect separate demographic histories (Amaral *et al.*, 2008). Demographic analysis shows a decrease of the southern Balkan wild boar populations approximately 10,000 generations ago (Fig. 8). This corresponds to the LGM (~50,000–10,000 years ago, Yokoyama *et al.*, 2000) assuming a generation time between 1 and 5 years (Groenen *et al.*, 2012; Herrero-Medrano *et al.*, 2013). The evolutionary history of wild boar at the southern part of the Balkan Peninsula is therefore consistent with the inferred histories of other temperate European mammals, where similar lineages are a result of population contractions and expansions during the LGM (Hewitt, 2000).

The degree of variation shared between the three inferred geographical and evolutionary entities differs. Our results support the existence of a completely separate group in Samos, with no overlapping zone, despite its geographical proximity to the southern Balkans. Anatolia was an important refugium during the LGM (Rokas *et al.*, 2003) and hybrid zones between European and Near Eastern lineages were found at the area of Thrace for some species (e.g. Michaux *et al.*, 2004). It seems, however, that already established local wild boar populations prevented Near Eastern populations from contributing substantially to the Balkan regions.

On the other hand, our findings show a hybrid zone on both sides of the Pindus range. This suggests allopatric differentiation in each of the refugial populations, followed by a secondary contact. Continental southern Balkans probably only had isolated sparsely occurring wooded areas during the LGM (Tzedakis *et al.*, 2002) which may have acted as local refugia for species such as wild boar that are restricted to tree covered habitats. Populations that survived at these sites during glaciations started expanding when the climate became warmer, and animals originating from different refugia subsequently met, forming hybrid zones.

Detecting subpopulation structure at a local level – discrepancy between microsatellite and SNP results in an inbred population

Microsatellite DNA Structure analysis revealed three local clusters (Fig. 2c) within the central Greek group. Each cluster is well separated and geographically delimited. These results are, however, not supported by SNP analysis, although the same 60K SNP assay was able to distinguish geographically close populations in Western Europe (Goedbloed *et al.*, 2013b).

For small-scale (micro-geographical) population structure, microsatellite-based Structure analysis can produce more clusters, especially when the populations examined are characterized by an isolation-by-distance pattern (supported both by microsatellite and SNP analyses). Isolation by distance causes deviations from random mating which can overestimate the number of clusters detected with microsatellite DNA (Frantz *et al.*, 2009).

Closer examination of the central group shows that it experienced a more pronounced bottleneck during the LGM (Fig. 8) which is reflected in the present day overall LD patterns (Fig. 7). Moreover, central Greek populations have the lowest values of heterozygosity, low *Ne* and the highest number of long homozygous regions. ROH's geographical pattern within central Greece is even more informative showing a west to east increase in cumulative ROHs size (Fig. 6). However, the correlation between ROHs size and longitude locally is weak due to the east–west increase of homozygous regions in a population where long ROHs were already present.

Based on LD and ROH analysis it is likely that a recent wild boar expansion, from the Pindus mountains eastwards created smaller inbred populations in the eastern part of central Greece. In human populations, long ROHs are related to inbreeding and migration events, which decrease population size and increase the probability of Identity By Descent (IBD) (Pemberton *et al.*, 2012). IBD distribution among unrelated individuals can infer demographic events that occurred in the last tens of generations (Palamara *et al.*, 2012). The existence of large numbers of long ROHs is consistent with recent inbreeding (Bosse *et al.*, 2012).

Inbreeding is usually related to translocations of few individuals (Frankham *et al.*, 2002), often used to re-stock declining or extinct populations which was a common practice for European wild boar (Vernesi *et al.*, 2003). The eastwards expansion of the central Greek group, however, appears to be a result of natural migration. Until 1980s, the

southernmost expansion area for wild boar in mainland Greece did not exceed 39th parallel north around the south-east part of Pindus range. In the following years traditional agriculture practices changed, fragmented wild boar habitats unified and local wild boar numbers increased. This population expansion resulted in migration towards south-eastern habitats reaching the 38th parallel (E. Chatzinikos personal observation). This is confirmed by our results: isolation-by-distance patterns, lower microsatellite DNA variation and no new alleles detected in the eastern clusters, show that the migration event involved solely the central Greek population.

Finding empirical evidence of inbreeding events, bottlenecks and subsequent expansions in the wild is often hard, as natural dispersal minimizes mating between relatives (Pusey & Wolf, 1996). Our study shows that these migration events have more chances to be detected with genome-wide markers, which increase the power of statistical analyses.

Domestic pig introgression

The suspected genetic admixture with domestic pigs in the sampled wild boars was similar with other European wild boar populations (5–10%, Scandura *et al.*, 2008; Goedbloed *et al.*, 2013a); 8.79% of the studied individuals were found to be possible hybrids. AFSA can inform about the extent and the source of domestic pig introgression by using SNP data. This technique essentially leverages the ascertainment bias introduced in the SNP chip design that resulted in part of the SNPs to be specific to pigs. As there are thousands of such SNPs on the assay, the analysis allows detection even of 5th or 6th generation backcrosses (Goedbloed *et al.*, 2013a). Microsatellites are not as effective at detecting hybrids: microsatellite alleles are usually shared between wild and domestic individuals and the F_{ST} values observed between them were too low (Scandura *et al.*, 2008, 2011). In addition, the chance of finding pig-specific alleles, even if they could be identified, in 3rd or 4th generation backcrosses would be small when using a handful of markers.

ACKNOWLEDGEMENTS

The authors thank the hunters and hunting associations for their help in sample collection. We also thank Dr Gus Rose for editing, graphs and discussion. Microsatellite analysis was financed by the 4th hunting Federation of Sterea Hellas through a grant awarded to C.T. SNP genotyping was supported by USDA grant 2007-04315 and the EU SABRE project FOOD –CT-2006-01625. R.P.M.A.C., H-J.M. and M.A.M.G. received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013) ERC grant agreement no 249894. SNP data analysis was performed in Animal Breeding and Genomics Centre, Wageningen University through an ESF Research Networking Programme Conservation Genomics Grant awarded to P.A.

REFERENCES

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

Amaral, A.J., Megens, H.J., Crooijmans, R.P.M.A., Heuven, H.C.M. & Groenen, M.A.M. (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics*, **179**, 569–579.

Amaral, A.J., Megens, H.J., Kerstens, H.H.D., Heuven, H.C.M., Dibbits, B., Crooijmans, R.P.M.A., Den Dunnen, J.T. & Groenen, M.A.M. (2009) Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC Genomics*, **10**, 374.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Barker, J.S.F., Hill, W.G., Bradley, D. *et al.* (1998) *Measurement of Domestic Animal Diversity (MoDAD): Original Working Group Report*. FAO, Rome.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. (2001) *GENETIX 4.05, logiciel sous Windows pour a genetique des populations*. Laboratoire Genome, Populations, Interactions, CNRS UMR 1571, Universite de Montpellier II, Montpellier, France.

Bosse, M., Megens, H.J., Madsen, O., Paudel, Y., Frantz, L.A., Schook, L.B., Crooijmans, R.P. & Groenen, M.A.M. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *Plos Genetics*, **8**, e1003100.

Defaveri, J., Vitaniemi, H., Leder, E. & Merila, J. (2013) Characterizing genic and nongenic molecular markers: comparison of SNPs and microsatellites. *Molecular Ecology Resources*, **13**, 377–392.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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

Frankham, R., Briscoe, D.A. & Ballou, J.D. (2002) *Introduction to conservation genetics*. Cambridge University Press, New York.

Frantz, A.C., Cellina, S., Krier, A., Schley, L. & Burke, T. (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.

Goedbloed, D.J., Megens, H.J., Van Hooft, P., Herrero-Medrano, J.M., Lutz, W., Alexandri, P., Crooijmans, R.P.M.A., Groenen, M., van Wieren, S.E., Ydenberg, R.C. & Prins, H.H.T. (2013a) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology*, **22**, 856–66.

Goedbloed, D.J., van Hooft, P., Megens, H.J., Langenbeck, K., Lutz, W., Crooijmans, R.P.M.A., van Wieren, S.E., Ydenberg, R.C. & Prins, H.H.T. (2013b) Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genetics*, **14**, 43. doi:10.1186/1471-2156-14-43.

Gomez, A. & Lunt, D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* (ed. by N. Ferrand and S. Weiss), pp. 155–188. Springer, Dordrecht.

Groenen, M.A.M., Archibald, A.L., Uenishi, H. *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, **491**, 393–398.

Guo, S.W. & Thompson, E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–72.

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E. (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, **13**, 635–43.

Herrero-Medrano, J.M., Megens, H.J., Groenen, M.A.M., Ramis, G., Bosse, M., Perez-Ensisco, M. & Crooijmans, R.P.M.A. (2013) Conservation genomics of domestic and wild pig populations from the Iberian Peninsula. *BMC Genetics*, **14**, 106.

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt, G.M. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hillis, D., Mable, B.K., Larson, A., Davis, S.K. & Zimmer, E.A. (1996) *Nucleic acids IV: sequencing and cloning*. Sinauer & Associates, Sunderland, MA.

Hofreiter, M., Serre, D., Rohland, N., Rabeder, G., Nagel, D., Conard, N., Münzel, S. & Pääbo, S. (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences USA*, **101**, 12963–12968.

Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hughes, P.D., Gibbard, P.L. & Woodward, J.C. (2007) Geological controls on Pleistocene glaciation and cirque form in Greece. *Geomorphology*, **88**, 242–253.

Iacolina, L., Scandura, M., Goedbloed, D.J., Alexandri, P., Crooijmans, R.P.M.A., Larson, G., Archibald, A., Apollonio, M., Schook, L.B., Groenen, M.A.M. & Megens, H.J. (2016) Genomic diversity and differentiation of a managed island wild boar population. *Heredity*, **116**, 60–67.

Kijas, J.W., Lenstra, J.A., Hayes, B. *et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *Plos Biology*, **10**, e1001258.

Knowles, L.L. (2009) Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.

Kraus, R.H.S., Van Hooft, P., Megens, H.J., Tsvey, A., Fokin, S.Y., Ydenberg, R. & Prins, H.H.T. (2013) Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Molecular Ecology*, **22**, 41–55.

Larson, G., Dobney, K., Albarella, U. *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Larson, G., Albarella, U., Dobney, K. *et al.* (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences USA*, **104**, 15276–15281.

Mantel, N. (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–000.

Manunza, A., Zidi, A., Yeghoyan, S., Balteanu, V.A., Carsai, T.C., Scherbakov, O., Ramirez, O., Eghbalsaied, S., Castelló, A., Marcadé, A. & Amills, M. (2013) A high throughput genotyping approach reveals distinctive autosomal genetic signatures for European and Near Eastern wild boar. *PLoS ONE*, **8**, e55891.

McEvoy, B.P., Powell, J.E., Goddard, M.E. & Visscher, P. (2011) Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Research*, **21**, 821–829.

Michaux, J.R., Libois, R., Paradis, E. & Filippucci, M.G. (2004) Phylogeographic history of the yellow-necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and Middle East. *Molecular Phylogenetics and Evolution*, **32**, 788–798.

Morin, P.A., Luikart, G., Wayne, R.K. & SNP Working Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Nikolov, I.S., Berbhard, G., Markov, G. & Kuehn, R. (2009) Population genetic structure of wild boar *Sus scrofa* in Bulgaria as revealed by microsatellite analysis. *Acta Theriologica*, **54**, 193–205.

Palamara, P.F., Lencz, T., Darvasi, A. & Pe'er, I. (2012) Length distributions of identity by descent reveal fine-scale demographic history. *American Journal of Human Genetics*, **91**, 809–822.

Pemberton, T.J., Absher, D., Feldman, M.W. *et al.* (2012) Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics*, **91**, 275–292.

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.

Pritchard, J.K., 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, M.A.R., Bender, D., Maller, J., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.

Pusey, A.E. & Wolf, M. (1996) Inbreeding avoidance in animals. *Trends in Ecology and Evolution*, **11**, 201–106.

R Core Team (2016) *R: a language and environment for statistical computing*. R Foundation for statistical computing, Vienna, Austria.

Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A. *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.

Rokas, A., Atkinson, R.J., Webster, L., Csoka, G. & Stone, G.N. (2003) Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*. *Molecular Ecology*, **12**, 2153–2174.

Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–6.

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

Scandura, M., Iacolina, L., Cossu, A. & Apollonio, M. (2011) Effects of human perturbation on the genetic make-up of an island population: the case of the Sardinian wild boar. *Heredity*, **106**, 1012–1020.

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

Sotiropoulos, K., Eleftherakos, K., Dzukic, G., Kaiezic, M.L., Legakis, A. & Polymeni, R.M. (2007) Phylogeny and biogeography of the alpine newt *Mesotriton alpestris* (Salamandridae, Caudata), inferred from mtDNA sequences. *Molecular Phylogenetics and Evolution*, **45**, 211–226.

Stewart, J.R. & Lister, A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**, 608–613.

Stumpf, M.P. & McVean, G.A. (2003) Estimating recombination rates from population-genetic data. *Nature Reviews Genetics*, **4**, 959–68.

Szpiech, Z.A., Jakobsson, M. & Rosenberg, N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.

Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J.M., Kristensen, T.N., Loeschcke, V., Gregersen, V.R. & Bendixen, C. (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, **103**, 326–322.

Tortereau, F., Servin, B., Frantz, L., Megens, H.J., Milan, D., Rohrer, G., Wiedman, R., Beever, J., Archibald, A., Schook, L.B. & Groenen, M.A.M. (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*, **13**, 586.

Tzedakis, P.C., Lawson, I.T., Frogley, M.R., Hewitt, G.M. & Preece, R.C. (2002) Buffered tree population changes in a quaternary refugium: evolutionary implications. *Science*, **297**, 2044–2047.

Ursenbacher, S., Schweiger, S., Tomovic, L. *et al.* (2008) Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): evidence for high genetic diversity and multiple refugia in the Balkan peninsula. *Molecular Phylogenetics and Evolution*, **46**, 1116–1128.

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

Vilaça, S.T., Biosa, D., Zachos, F. *et al.* (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.

Weinman, L.R., Solomon, J.W. & Rubenstein, D.R. (2015) Comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, **3**, 502–511.

Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L.K. (2000) Timing of the Last Glacial Maximum from observed sea-level minima. *Nature*, **406**, 713–716.

SUPPORTING INFORMATION

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

Appendix S1 Wild boar samples used for the microsatellite and SNP analysis.

Appendix S2 Detected wild boar-domestic pig hybrids with introgressed SNPs.

Appendix S3 L(K) and ΔK per K for each STRUCTURE run.

DATA ACCESSIBILITY

Microsatellite genotypes and SNP data in PLINK format used in this study are available from Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.t722h.

BIOSKETCH

Panoraia Alexandri is interested in the effect of demography and selection on the genetic structure of wild animals. This is well integrated in the focus of the research group from the Aristotle University (Greece), that is, the genetic analysis of Greek animal species.

Author contributions: CT, AT, PA and EC designed and planned the study. PA, EC and CT collected the specimens. PA performed the laboratory analysis. PA, AT, HJM, DJD, JMHM carried out the statistical analyses. RPMAC, MAMG, LAR and LBS supervised the SNP analysis and contributed to the manuscript. PA, AT, HJM and CT wrote the manuscript. All authors read, commented on and approved the final manuscript.

Editor: Malte Ebach