Genomic diversity and population structure of three autochthonous Greek sheep breeds assessed with genome-wide DNA arrays

Michailidou, S; Tsangaris, G; Fthenakis, GC; Tzora, A; Skoufos, I; Karkabounas, SC; Banos, G; Argiriou, A; Arsenos, G

Published in:
Molecular Genetics and Genomics

DOI:
10.1007/s00438-018-1421-x

First published: 25/01/2018

Document Version
Peer reviewed version

Genomic diversity and population structure of three autochthonous Greek sheep breeds assessed with genome-wide DNA arrays

S. Michailidou[1,2,*], G. Tsangaris[3], G.C. Fthenakis[4], A. Tzora[5], I. Skoufos[5], S.C. Karkabounas[6], G. Banos[1,7], A. Argiriou[2], G. Arsenos[1]

[1]Laboratory of Animal Husbandry, School of Veterinary Medicine, Department of Animal Production, Ichthyology, Ecology and Environmental Protection, Aristotle University of Thessaloniki, Greece.

[2]Institute of Applied Biosciences, Center for Research and Technology Hellas, 6th Km Charilaou Thermi Road, GR-57001, Thermi, Greece.

[3]Proteomics Research Unit, Center of Basic Research II, Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou St, 11527, Athens, Greece

[4]Veterinary Faculty, University of Thessaly, Karditsa, Greece.

[5]Department of Animal Production, Technological Educational Institute of Epirus, Arta, Greece.

[6]Cell and Molecular Physiology Unit, Laboratory of Experimental Physiology, Medical School, University of Ioannina, Ioannina, Greece.

[7]Scotland's Rural College and The Roslin Institute University of Edinburgh EH25 9RG UK.

*Corresponding author

Phone: +30 2311257521, Fax: +30 2310498270, email: sofia_micha28@certh.gr

Acknowledgements

This work was partially supported by the program 'GOSHOMICS' under the action COOPERATION 2009 (SYNERGASIA 2009), financed by the European Regional Development Fund and national resources, project code: 09SYN-23-990.

Abstract

In the present study, genome-wide genotyping was applied to characterize the genetic diversity and population structure of three autochthonous Greek breeds: Boutsko, Karagouniko and Chios. Dairy sheep are among the most significant livestock species in Greece numbering approximately 9 million animals which are characterized by large phenotypic variation and reared under various farming systems. A total of 96 animals were genotyped with the Illumina's OvineSNP50K microarray beadchip, in order to study the population structure of the breeds and develop a specialized panel of single nucleotide polymorphisms (SNPs), which could distinguish one breed from the others. Quality control on the dataset resulted in 46,125 SNPs, which were used to evaluate the genetic structure of the breeds. Population structure was assessed through principal component analysis (PCA) and admixture analysis, whereas inbreeding was estimated based on runs of homozygosity (ROHs) coefficients, genomic-relationship matrix inbreeding coefficients (F_{GRM}) and patterns of linkage disequilibrium (LD). Associations between SNPs and breeds were analyzed with different inheritance models, in order to identify SNPs that distinguish among the breeds. Results showed high levels of genetic heterogeneity in the three breeds. Genetic distances among breeds were modest, despite their different ancestries. Chios and Karagouniko breeds were more genetically related to each other compared to Boutsko. Analysis revealed 3,802 candidate SNPs that can be used to identify two-breed crosses and purebred animals. The present study provides, for the first time, data on the genetic background of three Greek indigenous dairy sheep breeds as well as a specialized marker panel that can be applied for traceability purposes as well as targeted genetic improvement schemes and conservation programs.

Keywords

Genetic diversity, population structure, OvineSNP50K beadchip, *Ovis aries*, breed identification, conservation, breeding programs

Introduction

Sheep, one of the most economically important livestock species, were brought under domestication in the region that stretches from the Northern Zagros to southeastern Anatolia at around 10500 BP, and then moved to the eastern and western arms of the Fertile Crescent (Zeder 2008). In Europe, sheep were spread throughout the valley of Danube and the Mediterranean Basin. In Greece sheep are reared since 6500 BC (Hatziminaoglou 2006a). Greece has probably one of the broadest and most valuable reservoirs of sheep genetic material in the Mediterranean and para-Mediterranean region (Hatziminaoglu et al. 1990). Breed identification through genomic evaluation is of high priority in order to explore the potential of our national flocks and design breeding schemes, although many scientists support a hypothesis that the concept of breeds and how they are formed is vague (Dodds et al. 2014). The Greek national flock of dairy sheep is among the largest in the E.U., counting approximately 9 million heads. Domestic breeds have been traditionally reared in marginal, mountainous, semi-mountainous and lowland areas and islands, and are well adapted to a wide range of environmental conditions. Due to their high levels of heterogeneity (Table S1), these animals can be reared under various production systems including extensive, semi-intensive or intensive. Sheep production in Greece plays a key role in the annual gross domestic product (GDP); approximately 772,072 tons of milk and 58,416 tons of meat are produced per year (FAOSTAT 2016). A major characteristic of the national flock of sheep is the numerous indigenous breeds which are still reared despite the continuous importation of foreign breeds such as Lacaune and Assaf (Gelasakis et al. 2012). The popularity and aggressive marketing of foreign breeds helped their fast distribution across the country. The latter, combined with the lack of comprehensive national breeding programs for indigenous breeds and the limited application of artificial insemination, has resulted in flocks being dominated by crossbred animals (Hatziminaoglou et al. 2006; Gelasakis et al. 2012). This fact challenges the maintenance of reasonable numbers of purebreds and raises the issue of unique genetic resources being lost.

Based on origin and phenotypic characteristics, there are 20 distinct indigenous Greek breeds of sheep that belong mainly to the Zackel type. However, breeds of the Ruda type are also found. The overall majority of autochthonous breeds in the national flock comprise five breeds (Chios, Karagouniko, Sfakion, Lesvou, Frizarta); the remaining purebreds are reared in small populations in isolated geographical areas or in islands. In the present study we selected three of the most representative breeds of dairy sheep in Greece: Chios, Karagouniko and Boutsko. Chios is the most productive indigenous dairy breed mostly reared in Northern Greece. Over the last three decades, this breed has been monitored though national programs for breeding and performance recording. Karagouniko is a classical lowland dairy sheep breed. Purebred animals and their crosses dominate most flocks in Central Greece. Karagouniko breed is also included in national projects for performance recording. Boutsko, is a representative mountainous breed and is mostly reared in the areas of West Macedonia and Epirus (Northwestern Greece). Boutsko is characterized by its ability to adapt to extreme environmental conditions while effectively exploiting poor mountainous pasture areas with moderate productive and reproductive performance (Hatziminaoglou 2006a).

Although the sheep genome is fully sequenced and published online since 2012 by the ISGC (International Sheep Genomics Consortium), genetic characterization and structure of Greek breeds has so far been limited to single or

few nuclear markers, mostly in form of microsatellite markers (Peter et al. 2007; Pariset et al. 2009; Loukovitis et al. 2016). Genome-wide association studies (GWAS) have only been applied so far in Frizarta breed for body size traits (Kominakis et al. 2017) and in Chios breed to identify genomic regions associated with mastitis (Banos et al. 2017).

In the last two decades, autosomal microsatellites have been the most frequently used markers in phylogenetic studies to address genetic diversity, mainly because they are informative, polymorphic and interspersed throughout the entire genome (Guichoux et al. 2011). Yet, their limitation due to inconsistencies in allele size calling and errors in size determination renders these markers unsuitable to apply in animal breeding (Fernandez et al. 2013). The introduction of high-throughput DNA genotyping microarrays established SNPs as more suitable for selective breeding and also more promising in assessing breed differentiation (Garke et al. 2012). Studying and understanding the genetic structure of a breed can help predict future threats caused by adverse environmental changes by accurately assessing each individual's potential and identifying important genes correlated with disease resistance, productivity or survival. Application of these data can improve national breeding schemes and management of genetic resources.

The objective of the present study was to characterize for the first time three autochthonous Greek breeds of dairy sheep using whole-genome genotyping microarray technology for breed identification and assignment, conservation, and breeding.

Materials and methods

Animal sampling

A total of 96 samples of 3 indigenous Greek dairy sheep breeds were randomly collected from 10 farms located in Central and Northern Greece (Fig. 1). All animals were raised under intensive or semi-intensive conditions and fed with a standard diet. Thirty samples belonged to Karagouniko breed, 32 to Boutsko breed and 34 samples Chios breed. Breed discrimination was performed based on each individual's pedigree records in addition to phenotypic characteristics. Blood samples were collected from the jugular vein in tubes containing EDTA and stored in a freezer (-20°C) until further use in the laboratory.

DNA extraction and genotyping

DNA was extracted from whole blood samples using the Nucleospin Blood Quick Pure kit (Macherey-Nagel, Germany) according to the manufacturer's protocol. DNA integrity was assessed by electrophoresis on a 0.8% TAE agarose gel. DNA concentration was quantified with the Qubit® dsDNA BR assay kit (Qubit 2.0 Fluorometer, Invitrogen, Carlsbad, CA, USA) according to Illumina's recommendations for Infinium genotyping assays. Medium-density OvineSNP50K Genotyping Beadchip (Illumina Inc, San Diego, California) containing 54,241 SNPs with an average interval between SNPs of 46kb was used to genotype all sheep samples. Samples were dried down in 1 mMTris-EDTA and 400 ng of each sample was sent to GeneSeek (Neogen Corporation, UK) for genotyping. In brief, DNA samples were amplified through an overnight isothermal reaction at 37°C and then enzymatic fragmentation was applied in order to obtain fragments of optimum length (300-600bp) for hybridization

to the beadchip. Samples were then precipitated and re-suspended in hybridization buffer, and hybridized in the surface of the chip. Imaging and data generation was performed by the Illumina's iScan system.

Data analysis

Genotyping data were automatically pre-processed with Illumina's GenomeStudio software (v.1.9.4) including a preliminary quality control and data analysis step (exclusion or inclusion of samples, SNP and sample call rate generation, allele frequencies). PLINK v1.90 (Purcell et al. 2007) was then used to generate summary statistics and conduct exact tests for deviation from Hardy–Weinberg equilibrium (HWE). Chromosomal coordinates (number and position) for each SNP were assessed through the online platform SNPchiMp v3 based on Oar_v3.1 assembly (Nicolazzi et al. 2015). Further quality control and SNP filtration was performed using the R statistic package (R Development Core Team 2010), by filtering out SNPs with minor allele frequency (MAF) <1%, call rate <0.95, deviation from HWE (P value=0.001) and lacking genomic location. Observed (*Ho*) and expected (*He*) heterozygosity and Wright's inbreeding coefficient (F_{IS}) were calculated for each breed to measure the genetic diversity within breed. Pairwise F_{ST} values were also calculated to test genetic relationships and distance between breeds, using ARLEQUIN 3.5.2.2 software (Excoffier and Lischer 2010).

For the assessment of genetic structure through Bayesian clustering the quality-filtered SNPs were pruned by removing SNPs in linkage disequilibrium (LD) at a threshold of r^2 =0.5. Bayesian clustering was performed in autosomal LD-pruned SNPs using the ADMIXTURE v1.3 software (Alexander and Lange 2011) assuming a number of subpopulations (K) ranging from 1 to 10 and visualized using R plots. Population structure was also examined by principal component analysis (PCA). PLINK was used for the generation of eigenvectors and eigenvalues which were visualized with the GENESIS software demonstrating the relationship between PCA1 and PCA2 (Buchmann and Hazelhurst 2015).

Genomic inbreeding coefficient was also calculated following (VanRaden 2008). Computation of inbreeding values were assessed from the diagonal of the genomic relationship matrix, denoted as F_{GRM}. Estimates of inbreeding coefficients were calculated from autosomes, for each individual using the GCTA program (Yang et al. 2011) starting from the PLINK binary PED files (.bed, .fam, .bim). Runs of homozygosity (ROH) were defined to assess autozygosity (inbreeding) for all individuals in autosomes, in each breed separately, and categorized based on their length and chromosome using PLINK v1.90. ROHs were calculated with the default parameters of the 'Runs of homozygosity' function in the PLINK software except for the total length of ROHs (≥10kb), with 20 variants in the scanning window, allowing one heterozygous SNP and 2 missing calls to estimate homozygosity. The percentage of ROHs per chromosome was calculated as proposed by Al-Mamun et al. (2015), by summing all ROHs (Mbp) on a chromosome multiplied by 100 and then dividing by the number of animals that had ROH on that chromosome multiplied by the respective chromosome length (Mbp). Genomic inbreeding coefficient based on ROH (F_{ROH}) was calculated for each individual from the sum of ROH lengths, divided by the total length of the autosomal genome (kb) covered by SNPs as proposed by McQuillan et al., (2008) with the difference of including the centromeric region. Inbreeding coefficients based on ROH were calculated for four different bins, grouped according to the

length of ROHs (i) all ROHs, ii) <10 Mb, iii) 10-20 Mb, iv) >20 Mb). For all inbreeding estimates, coefficients were calculated separately for each individual and then averaged per breed.

Linkage disequilibrium was also tested for each breed separately to examine recombination of linked SNPs using PLINK v1.90 with default parameters, with SNPs included in this step spanned a distance from 0.001 to 1Mb. The squared correlation coefficient (r^2) curve was estimated by determining the nonlinear least squares fit line using the *nls* function in R. Effective population size (Ne) was estimated separately for each breed, using SNeP version 1.1 (Barbato et al. 2015). Ne estimates at different generations were based on linkage disequilibrium, using the formula suggested by Corbin and co-authors (2012). Estimated effective population sizes against past generations were plotted in R over the last 100 and 3,000 generations ago. For the estimation of ROHs, LD, Ne and population structure, SNPs that were located on chromosome X were excluded, since they express lower mutation rate and a direct comparison of the overall diversity between the X chromosome and autosomes is difficult (Schaffner 2004).

To trace SNPs that could be used to distinguish animals from different breeds from each other, genome-wide associations between SNPs and breeds were derived using the R package 'SNPassoc' (Gonzalez et al. 2007), assuming four different genetic inheritance models (dominant, recessive, log-additive and co-dominant). Three pairwise breed comparisons were designed to test for significant SNPs after GWAS: Boutsko-Chios (BC), Boutsko-Karagouniko (BK) and Chios-Karagouniko (CK). Statistically significant SNPs for each pair were acquired after Bonferroni correction of p values at level 1e-06. To further test if these pairwise significant SNPs can be used to properly cluster individuals to their origin among the three breeds, de novo hierarchical clustering was performed with the unique statistically significant SNPs derived from each pair of breeds from the log-additive model. Hierarchical clustering of samples was performed through the construction of heat-map dendrogram with 'gplots' R package using the heatmap.2 function (Warnes et al. 2011). Using *hclust* function, each individual is assigned to its own cluster and at each stage the two most similar clusters are joined until there is a unique cluster (Murtagh and Legendre 2014). Hierarchical clustering was applied with five different distance methods (euclidean, manhattan, canberra, binary and minkowski). Statistically significant SNPs were further classified with the Variant Effect Predictor (VEP) online tool in order to determine the association of SNPs on genes, transcripts, protein sequence, as well as regulatory regions (McLaren et al. 2016).

Results

Sample and SNP filtration / descriptive statistics

From the 96 samples only one individual from the Karagouniko breed was excluded from further analysis due to its low call rate (0.8314<0.95). Mean call rate for the remaining 95 samples was 0.9926. Quality control of SNPs was performed by filtering out non-informative SNPs for minor allele frequency (MAF), call rate (call frequency) and Hardy-Weinberg equilibrium (HWE). As such, 1,638 SNPs with call rate<0.95, 1,525 SNPs with MAF<0.01, and 942 SNPs with HWE P≤0.001 were excluded from downstream analysis (Table 1). In addition, 4,011 SNPs that had no assigned location on chromosomes were excluded from the analysis. In total, 46,125 SNPs were kept for analyzing genetic diversity indices (F_{ST}, *Ho*, *He*, F_{IS}), PCA, genome-wide associations between SNPs and breeds

and hierarchical clustering heatmap-dendrograms, whereas for the estimation of ROHs, LD, Ne and admixture analysis, the 44,954 autosomal SNPs were analyzed/used.

Within breed genetic diversity

From the 46,125 filtered SNPs, the percentage of within-breed polymorphic loci was greater than 98% in all cases (Table 2). The highest number of polymorphic loci was found in Chios breed (n=45,568 SNPs), followed by Karagouniko (n=45,445 SNPs) and Boutsko (n=45,225 SNPs) breeds. Although just the Chios breed had been used for the development of OvineSNP50K beadchip within the International Sheep Genomics Consortium (Kijas et al. 2012), the other two breeds presented high levels of polymorphism, too. According to *He* values, Chios breed presented the highest within-breed genetic diversity. Mean *He* for all breeds was 0.411 ranging from 0.407 to 0.413, for Boutsko and Chios breeds, respectively. Nucleotide diversity (π) values were calculated separately for each breed and ranged from 0.399 (Boutsko) to 0.407 (Chios). These results are consistent with *He* for the three breeds.

Different estimates of inbreeding coefficients were calculated by the SNP genotyping data (Table 3). Boutsko and Chios populations demonstrated positive F_{IS} values indicating more homozygotes than expected, whereas for the Karagouniko population only five out of twenty-nine samples had positive F_{IS} values, which is indicative of heterozygote excess in the studied population. In addition, Karagouniko breed presented the lowest values for F_{IS}, F_{GRM} and F_{ROH}. Estimated F_{GRM} values were positive for all breeds. Chios breed expressed the highest F_{IS} and F_{GRM} values, whereas Boutsko breed had the highest F_{ROH} value. The differences on levels of inbreeding reflected by F_{IS}, F_{ROH} and F_{GRM} were modest among breeds. Regarding the inbreeding coefficients calculated for the different bins, $F_{ROH>20Mb}$ expressed the highest values. Moreover, at small lengths, where very short and common ROHs are located due to LD (Mastrangelo et al. 2017), $F_{ROH<10Mb}$ expressed very low values in the three sheep breeds. Across all measures of inbreeding coefficients, Karagouniko seems to be the least inbred breed, whereas Chios and Boutsko breeds are the most inbred.

Between breed genetic diversity

Genetic differentiation of breeds and the level of their relatedness, indicated by pairwise F_{ST} values was relatively low, ranging from 0.02563 to 0.04929 (Table 2). Chios and Boutsko breeds were found to be the most genetically distant from each other with an average pairwise F_{ST} value of 0.04929, whereas Chios and Karagouniko were more closely related to each other. This is not surprising, since Boutsko populations are more isolated from the others due to geographical barriers (Pindus mountain range) and throughout the years, Boutsko purebreds were selected for different breeding purposes.

Runs of homozygosity, linkage disequilibrium patterns and effective population size

Runs of homozygosity were tested for all samples in each breed, to assess whether inbreeding had occurred in the studied populations. According to the parameters used, 865 ROHs were found in total; Chios had the largest number of ROHs (n=404), followed by Boutsko (n=371) and Karagouniko (n=90) with an average of 11.88, 11.59 and 3.10 ROHs per individual, respectively (Fig. 2). The longest ROH was found in an individual of the Chios breed (max. length 56.867Mbp), but on average, the Boutsko population presented the longest ROHs (mean=13.139Mbp).

Three-hundred and forty-four ROHs were longer than 10Mbp (122 for Chios, 190 for Boutsko and 32 for Karagouniko breed). Chromosome 2 (OAR2) had the largest number of ROHs (n=121), followed by OAR1 (n=112). As expected, the total number of ROHs per chromosome decreased with decreasing chromosome length (Fig. 3). The highest percentage was calculated for chromosome 24 (35.3%) and the lowest for chromosome 3 (8.2%).

Extent of linkage disequilibrium patterns were tested with r^2 in the 44,954 autosomal SNPs, separately per breed. As seen in Fig. 4 the LD curve presents a rapid decline in short distances and decreases with increasing intermarker distance. Chios and Karagouniko breeds presented higher rates of LD decay, in comparison to Boutsko breed which displayed slightly increased levels of LD. In general, LD patterns were found to be relatively similar between breeds according to the parameters used; largest mean r^2 was observed in Boutsko (0.121) followed by Chios (0.098) while the smallest r^2 value was observed for Karagouniko breed (0.092). The average intermarker distances were about 270 kb for all breeds. Most of the SNP pairs that were found in complete LD (r^2> 0.9) were spaced at physical distances of 53 kb, 55 kb and 94 kb for Karagouniko, Chios and Boutsko breeds, respectively. Chromosomes 1, 2 and 3 had the largest number of adjacent SNPs in LD, whereas the lowest were observed in chromosomes 21 and 26 (Fig. S1). No SNPs were found in LD for OAR24. Generally, the number of adjacent SNPs in LD tend to decrease with decreasing chromosome length.

Variations of genetic diversity measured by the effective population size showed a decline for all breeds over past generations. According to the genotyped populations in the distant past, for the period of 100 – 3,000 generations ago, Karagouniko breed had the largest Ne across generations, whereas Boutsko had the lowest (Fig. 5a). Approximately 3,000 generations ago Ne was estimated to be 5,463, 4,922 and 4,691 for Karagouniko, Chios and Boutsko breeds, respectively. However, over the last 100 generations a rapid decline is observed for all breeds, with Karagouniko breed suffering the most severe drop in Ne (Fig. 5b). The most recent Ne values, dated 13 generations ago (equals to ~3 years ago, assuming a generation interval of 4 years in sheep) with 73, 97 and 130 population sizes for Boutsko, Karagouniko and Chios breeds, respectively.

Population structure analysis

A total of 41,014 SNPs remained after LD-pruning (3,940 SNPs were removed) to be further evaluated with ADMIXTURE. Population genetic structure was investigated through Bayesian clustering, assuming a number of K from 1 to 10. Cross validation (CV) error was the lowest in K=3 (0.54030), indicating the most likely number of different breeds represented in the 95 samples, as proposed by Alexander and Lange (2011). The first group that was differentiated consisted of Boutsko individuals at a low K value of 2, with relatively tight clustering and least admixture (Fig. 6). At K=3 three distinct clusters were formed, one for each breed. Chios and Karagouniko samples formed looser clusters, each one enclosing cores of admixture individuals. Notably, all these individuals were found in the same area in Central Greece. At higher K values (from 4 to 10) CV error reached 0.68190 and classification of breeds started to disappear progressively with minor distinct groups.

Population structure as formed from PCA analysis in 44,954 autosomal SNPs also revealed three distinct clusters, according to their origin (Fig. 7). The first two principal components (PC1 and PC2) explained 13.17% and the first 5 PCs the 20.64% of the total genetic variation. Analysis of principal components showed that Chios breed expresses high levels of genetic variation, which is consistent with its *He* values. Individuals from the Boutsko breed formed a distinct cluster in a greater distance from the other two, with the exception of two samples. These samples originated from the same farm in central Greece and formed a group with other individuals from the Karagouniko and Chios populations, also originated from central Greece. Although all samples were selected based on their phenotypic characteristics and pedigree data, results suggest possible two-breed crosses in some individuals. This is expected, since many stockowners want to improve lactation and meat performance of their ewes and, as such, crosses between different breeds are usual.

Population structure was also assessed by applying more stringent criteria (HWE P value=0.05), resulting in 5,514 autosomal SNPs being kept for the analysis, but the clustering pattern did not change. Both admixture and PCA analysis indicated that the use of OvineSNP50K can be applied in Greek breeds to assess population structure and trace crossbreds in flocks.

Genome-wide association between sheep breeds

In order to search for markers that could be applied in breeding schemes and classify individuals according to breed origin, the three breeds were tested against each other by applying genome-wide association analysis of SNP effects on breed (dependent variable). Three pairwise breed comparisons were designed to test for significant SNPs: BC, BK and CK. In each case, four different models of inheritance were assumed: dominant, co-dominant, log-additive and recessive. The significance level for confidence intervals was set to 1e-06. A preliminary analysis on the 95 samples revealed that five of them, two of the Boutsko and three of the Karagouniko breed clustered with other breeds and, in conjunction with their individual diversity scores and respective admixture and PCA profiles, they were excluded from further analysis. Consequently, 90 samples were included in genomic associations.

Differences in the total number of significant SNPs were observed among the four different models of inheritance (Table 4). Comparison of Boutsko and Chios breeds showed that BC exhibits the highest number of SNPs that can be used to distinguish animals between these two breeds. This was the case independently of the model of inheritance used. This result suggests that these two breeds are the most genetically distant. The latter is consistent with their F_{ST} values and, moreover, corresponds well to the phenotypic and productive characteristics of these two breeds. Regarding the different models of inheritance for BC pair, dominant (n=2,512) and co-dominant (n=2,310) models presented similar number of SNPs that can be further used to analyze genetic distances; log-additive led to the highest number of SNPs (n=2,606), whereas the recessive model was associated with the lowest number of SNPs (n=345). Similar patterns regarding the number of significant SNPs were observed for BK and CK pairs. The pair BK, whose breeds are highly differentiated as well, presented almost the half number of the BC's pair statistically significant SNPs, indicating that Boutsko breed is more related to Karagouniko than to Chios breed. Specifically, 1,227, 1,209, 1,361 and 96 statistically significant SNPs can be used to discriminate Boutsko and Karagouniko breeds, when applying dominant, co-dominant, log-additive and recessive inheritance models, respectively.

Evaluation of the CK pair revealed that significantly fewer SNPs compared to the other two pairs can be used to differentiate between the Chios and Karagouniko breeds. Fewer than 500 SNPs can be analyzed when dominant, co-dominant and log-additive models were applied. In the case of the recessive model, a minute percentage (>0.1%) of total SNPs can be used to further associate genetic distances of these breeds. The latter is an indicator that Chios and Karagouniko are two breeds that are closely related to each other.

To test if these groups of significant SNPs can be applied across all three breeds simultaneously, a hierarchical dendrogram was constructed to screen purebreds, without breed assignment. In total, by applying all inheritance models, 4,581 unique SNPs were identified for further evaluation. Since it is not possible to know the manner of inheritance for each SNP, as most animal traits are polygenic, the log-additive model was selected for further analysis because it is assumed to reflect closer trait inheritance in real life. All possible pairwise breed comparisons for the log-additive model were calculated (Fig. S2) and SNPs were acquired after Bonferroni correction for further evaluation. Six hundred and forty-three SNPs were common between the three sheep pairs and were removed, resulting in 3,802 unique SNPs derived from the log-additive model. Significant SNPs were distributed in all chromosomes, with the highest number of SNPs observed on chromosomes 1 (409 SNPs), 2 (461 SNPs) and 3 (368 SNPs) (Fig. 8). The majority of them (48.72%) were located in intergenic regions and a lower percentage (42.09%) was located within introns (Table 5). Concerning their putative impact on protein function, most SNPs were found to be mostly harmless or unlikely to change protein behavior. Only SNP rs421765058 (CGA/TGA) in aquaporin 10 (AQP10) was found to introduce a premature stop codon, leading to a shortened polypeptide. For 348 SNPs there was no available information on impact on gene regulation and protein primary structure, and they are not included in Table 5.

Construction of a hierarchical heat map dendrogram based on the 3,802 significant SNPs revealed 3 distinct clusters, one for each breed, independently of the distance method used. Since results were consistent among the different distance methods applied, euclidean distances were used to visualize clustering results, as they reflect in the most representative way absolute distances between samples. Cluster ‘A’ comprised exclusively all the samples of Boutsko breed, and appeared to be more distant from the other clusters (Fig. 9). The other two clusters ‘B’ and ‘C’ consisted of samples of the Karagouniko and Chios breeds, respectively. Moreover, when PCA was performed with the 3,802 SNPs, the first two principal components accounted for 40.27% of the total variation, indicating that this specialized marker panel is more informative than the 44,954 SNPs which accounted for the 13.17% (PC1 + PC2) of the total variation of the studied breeds (Fig. S3). These results provide strong evidence that the 3,802 SNPs can be applied for identification in purebreds and crossbreds of these breeds.

Discussion

We used the medium-density OvineSNP50K beadchip to estimate genetic diversity and identify candidate SNPs for a panel that could be applied to discriminate among animals from three common Greek sheep breeds (Chios, Karagouniko and Boutsko) and their crosses. The results of the present study suggest that these sheep populations are highly heterogeneous compared to other *O. aries* breeds (Kijas et al. 2012) as well as their wild counterparts, the

bighorn and the thinhorn sheep (Miller et al. 2011). On average, Greek breeds possessed greater genetic diversity based on *He* values, estimated from genotyping microarrays (mean *He*=0.41) to other breeds reared in the Mediterranean basin such as three Sicilian (mean *He*=0.38) (Mastrangelo et al. 2014), nineteen Italian (mean *He*=0.35) (Ciani et al. 2014) or eight Algerian (mean *He*=0.37) (Gaouar et al. 2017) breeds. Higher genetic heterogeneity was also observed compared to Australian sheep breeds like Merinos (mean *He*=0.38), Poll Dorset (mean *He*=0.30) and Border Leicester (mean *He*=0.34) (Al-Mamun et al. 2015).

These results are in agreement with previous studies based on the mtDNA of Mediterranean sheep (Pariset et al. 2011). In contrast, our results are not consistent with those produced by microsatellite markers, where mean expected heterozygosity in Greek sheep breeds was 0.74 (Ligda et al. 2009). Other studies in several European sheep breeds show the same discrepancy between using Illumina's Ovine50K (mean *He*=0.33) (Kijas et al. 2012) versus microsatellite markers (mean *He*=0.72) (Peter et al. 2007). This can be explained by the nature of SNPs (bi-allelic) in comparison to microsatellite markers (many different alleles), which render results derived from these two methods incomparable. The high levels of genetic heterogeneity in Greek populations are mainly due to the predominantly random mating within breeds. Artificial insemination, despite its advantages, has never been frequently used in sheep breeding schemes in Greece so far, and selection for specific traits and genotypes has not been applied systematically, thereby further contributing to high levels of heterogeneity within breeds (Hatziminaoglou et al. 2006; Valergakis et al. 2010; Gelasakis et al. 2012). From a national perspective, it is important to define purebreds and preserve diversity within breed, as it can be exploited for the genetic upgrading of flocks by selecting from a rich pool for special traits affecting animal performance.

Our results showed that Chios individuals exhibit greater genetic diversity compared to Boutsko and Karagouniko, as well as other foreign breeds; Chios breed was shown to be more diverse and differentiated compared to three Sicilian sheep breeds and six of other origins, including Lacaune and Merinos (Mastrangelo et al. 2014). The results also showed that Boutsko breed is genetically more distant from Chios and Karagouniko. This is most likely related to geographic and socioeconomic conditions, since Boutsko breed is traditionally reared in the Pindus mountain range by nomads whereas Chios and Karagouniko are more intensively reared in the lowlands. Moreover, Boutsko breed has not been subjected to any breeding program, whereas some selection for improved milk production has been taking place on and off for the past few years in Chios and Karagouniko. Genetic distances between breeds did not change even when more stringent criteria (HWE>0.05) and different inheritance models were applied, which reduced the number of SNPs from 46,125 to 5,514. Differences were only observed when the recessive inheritance model was applied, but this model cannot be used for the parallel evaluation of thousands SNPs, since the high number of SNPs just not justify a recessive model of inheritance. Moreover, the dominant notion is that when there is uncertainty about the genetic model behind the genotype-phenotype association, then the additive genotypic test is a good compromise (Purcell et al. 2007). Genetic homogeneity of Boutsko breed was observed despite the multiple sampling locations in different regions, in contrast to Karagouniko, where higher heterogeneity levels were observed although all individuals originated from the same region. Despite differing in visible phenotypic characteristics, Chios and Karagouniko breeds appeared to share a similar genetic background. Karagouniko was found to be genetically closer to Chios in the present study, despite having a common ancestor with Boutsko (Zackel sheep)

(Hatziminaoglou 2006b). Regardless of the common ancestor or the genetic distances, our results indicate that these three breeds can easily and accurately be discriminated using a subset of SNP from the OvineSNP50K beadchip.

Genetic diversity in livestock is mostly based on calculated heterozygosity and Wright's *F* statistics (Lenstra et al. 2012). Furthermore, the length of ROHs and the extent of LD between two adjacent markers reflect levels of inbreeding and may be used to estimate genetic variation within a population or a breed. In our study, analysis of inbreeding through the long ROHs indicated higher inbreed levels in Boutsko compared to Chios and Karagouniko. This may be reflective of more random mating patterns in these latter two breeds than in Boutsko. Additionally, the smallest effective population size in Boutsko breed may suggest a small founding population during breed formation. The natural breeding areas of the Boutsko breed (isolated regions of the Pindus mountain range) might have contributed to the creation of a highly homogeneous population. Overall, trends of Ne in Greek breeds, were in accordance with sheep breeds worldwide, presenting a sharp decline over time, mainly due to domestication, intensive selection and the limited use of AI (Kijas et al. 2012; Barbato et al. 2017; Chitneedi et al. 2017; Purfield et al. 2017). Inbreeding estimates from pedigree data were not calculated in this study, since genealogy of the majority of animals used was unavailable. Hence, comparison of inbreeding among genomic and pedigree data was not possible. However, many studies have shown that F_{ROH} seems to be a more powerful approach for detecting inbreeding depression than any other alternative, due to the advantage of not being affected by estimates of allele frequency or incompleteness of the pedigree data (Keller et al. 2011; Zhang et al. 2015).

From the aspect of farm locations, no particular geographical pattern was observed regarding the number or length of ROHs. In general, the total number of ROHs for each breed was relatively low and similar to results from studies on Australian sheep crosses (Al-Mamun et al. 2015).

Throughout the years inbreeding has been implicated for decreasing biodiversity, inviting a higher prevalence of genetic diseases, lower fertility and a decrease in disease resistance (Felius et al. 2015). Hence, although the number and length of ROHs is indicative of purebreds, particular attention should be given when including/selecting these animals in breeding schemes, since high ROHs run the risk of introducing susceptibility traits. Moreover, results can vary since there is no consensus or standardized criteria used to define a ROH. True extent of homozygosity can be underestimated through not clearly defined SNPs due to e.g. hemizygous deletions, or due to the pruning, or not, of SNPs that are in strong LD before the data can be used for ROHs (Howrigan et al. 2011; Ku et al. 2011). OAR2 had the largest number of ROHs, and as a general trend, the number of ROHs decreased with decreasing chromosome length, which was consistent with Al-Mamun et al. (2015) and Chitneedi et al. (2017) in Australian and Spanish sheep breeds, respectively. However, OAR24 was the chromosome with the highest average percentage in ROHs, followed by OAR20 and OAR21. OAR24 was the chromosome with the highest coverage of ROHs in the Churra breed, too (Chitneedi et al. 2017), but interestingly, no ROHs were found in OAR24 of the Australian breeds (Al-Mamun et al. 2015). Our results showed that the region between 32Mbp to 34Mbp of OAR24 displayed the largest number of occurrences in ROHs. This region is related to quantitative trait loci (QTL) for body weight traits, average daily gain and milk yield in sheep (Raadsma et al. 2009). Overall, SNPs that were the most frequently found in ROHs were located in genes such as *TRPM6* (transient receptor potential cation channel, subfamily M, involved

in cellular protein modification processes), *CNOT2* (CCR4-NOT transcription complex, subunit 2, protein tyrosine phosphatase, receptor type, D, involved in cell differentiation and stem cell maintenance), *NRXN1* (neurexin 1, involved in neurophysiological regulations), *PSIP1* (PC4 and SFRS1 interacting protein 1, involved in chromatin structure and gene regulation), *BNC2* (basonuclin 2, involved in coat pigmentation in sheep) and *EPHA2* (EPH receptor A2, involved in protein signaling pathways) (Zdobnov et al. 2017).

In the present study, LD was studied through r^2, mostly because in comparison to D', r^2 is a more reliable factor in studies with small sample sizes and more useful in predicting the power of association mapping (Khatkar et al. 2008). The fact that LD decayed rapidly with increasing intermarker distance points to the need for denser SNP panels to undertake successful genome-wide association studies in future (Meadows et al. 2008; Mastrangelo et al. 2014; Al-Mamun et al. 2015). Linkage disequilibrium data are critical for GWAS since they play a fundamental role in gene mapping and could serve as a tool for fine mapping of complex disease (Pritchard and Przeworski 2001). Sharp decline in LD has been observed in other sheep breeds as well, like the Chinese Sunite breed (Zhao et al. 2014), many Sicilian breeds (Mastrangelo et al. 2014), some Australian breeds (Border Leicester, Merino, Poll Dorset) (Al-Mamun et al. 2015) and breeds originating from New Zealand (Texel, Primera, Lamb Supreme) (Brito et al. 2017). However, compared to other livestock species such as cattle (Lu et al. 2012) or pigs (Zanella et al. 2016; Grossi et al. 2017), the extend of LD is much lower in sheep. The less intensive selection in sheep than in other domestic species (Kijas et al. 2012) characterized by different recombination events, population dynamics, as well as breed selection bottlenecks (Meadows et al. 2008) could have influence the differences in the extent of LD. In agreement with the results obtained from ROHs, the lower LD decay in Boutsko breed, confirmed the increased levels of homozygosity compared to the other two breeds.

Trends of Ne in Greek breeds, were in accordance with sheep breeds worldwide, presenting a sharp decline over time, probably due to the intense selection pressure and inbreeding (Kijas et al. 2012; Barbato et al. 2017; Chitneedi et al. 2017; Purfield et al. 2017). About 50 generations ago, Ne in Boutsko, Chios and Karagouniko breed was 229, 358 and 342, respectively. Among the three breeds, Chios has only been studied in the past by the SheepHapMap project (Kijas et al. 2012) and similar Ne size was obtained (Ne=334). However, the lowest estimate of Ne in our study (13 generations ago) for Chios breed dropped at 130, indicating the intense selection processes on this breed. It is important to note that discrepancies in Ne estimates between studies may be observed, because this index might be strongly biased when the sample size is small, probably as a result of the LD generated by the sampling process (Mastrangelo et al. 2017).

Having investigated population structure and diversity levels within and between breeds, we tried to identify candidate SNPs that could be used to identify the breed origin of animals. Such an application would be of particular interest to traceability issues and breeding programs. In the past, similar approaches have been followed with microsatellite markers (Dalvit et al. 2008; Niu et al. 2012; Sardina et al. 2015) or specific genes (Fontanesi et al. 2011). With the progress in genotyping technologies, genome-wide SNP microarrays in livestock have replaced simple sequence repeats (SSRs) for breed traceability purposes (Ramos et al. 2011; Dimauro et al. 2013) and for parentage testing (Clarke et al. 2014; Heaton et al. 2014). Since the sheep industry of autochthonous breeds in

Greece relies almost exclusively on pedigree data with very limited records on past genealogy, we searched for an alternative method to better assign individuals to their origin. Using the OvineSNP50K beadchip, Heaton et al. (2014) found 163 SNPs that can be used as parentage SNPs, which are also ideal molecular identifiers for tracing sheep products to their source. Fifteen of these SNPs were common with our dataset, some of which were located in *TBC1D4* (TBC1 domain family, member 4), *FCHSD2* (FCH and double SH3 domains 2), *PPP1R13B* (protein phosphatase 1, regulatory subunit 13B), *LOC101104482* (alpha-1-macroglobulin-like), *TDRD7* (tudor domain containing 7), *NAALADL2* (N-acetylated alpha-linked acidic dipeptidase-like 2), *LIMA1* (LIM domain and actin binding 1), *DPH6* (ATP binding domain 4) and *SH2D4A* (SH2 domain containing 4A) genes. The majority of significant SNPs were located in introns and intergenic regions, which is expected and in agreement with other studies, since it is known that most of the trait-associated SNPs in GWAS studies are located in genes without obvious biological significance on the analyzed phenotypes, or they are located in the intergenic regions or introns of certain genes (Fan et al. 2010; Hayes and Goddard 2010). In our study, only SNP rs421765058, which is located in *AQP10* on chromosome 1, was among the significant ones that impacts a gene coding region. In sheep, rs421765058 is a nonsense mutation, introducing a premature stop codon that leads to a shortened polypeptide. Many studies in mammals suggest that several aquaporin proteins collaborate in the lactosynthesis in the mammary gland (Mobasheri and Barrett-Jolley 2014). Aquaporins are important since they primarily transport water across the plasma membrane driven by osmotic gradients, and in particular, *AQP10* acts mainly in the epithelial cells of the gastrointestinal tract (Ishibashi et al. 2009). It was recently discovered that *AQP10* in sheep (and other ruminants) may be a pseudogene, thus it does not produce a functional protein even if translated (Tanaka et al. 2015). Taken together, environmental and genetic mechanisms influencing their expression levels should be further analyzed to assess if a particular expression pattern exists between breeds with different milk yields, such as the ones included in the present study.

The results of the present study suggest that Greek sheep breeds maintain high levels of genetic resources based on diversity indices. Presence of genetic variation can be exploited in the design and implementation of appropriate breeding schemes. Even in Chios breed, which has been subjected to some selection pressure over the years, the genetic diversity still remains high. FAO recommends an effective population size of 50 or more per generation to maintain the fitness in a breed (FAO 1998). From this aspect, these three Greek breeds do not seem to suffer from inbreeding depression that could lead to reduced biological fitness (Leroy 2014). The fact that some animals were revealed to be crossbreds despite their recorded pedigree points to the need for a more effective way to record and identify individuals. For many years we had limited knowledge on the evolution and domestication of Greek breeds through the transit of sheep from the southeastern Anatolia to the Mediterranean basin. This study sets the first step towards the genomic mapping of the Greek national flock. These results, alongside with the proteomic characterization of animal milk profiles (Anagnostopoulos et al. 2016) will be used for breed identification and traceability purposes. Further analyses involving larger population sizes and other breeds should be carried out to further validate and extend our results. The importance of conserving purebreds is recognized worldwide since irreversible loss of breeds could lead to decreased adaptability and survival rates to upcoming changing environmental conditions. Unfortunately, local breeds are threatened more and more since stockowners prefer

intensive sheep farming with foreign breeds that produce more than the local ones. European Union policies, report the importance of conservation and sustainable use of these unique genetic resources, so as to be used for future generations in threats like disease outbreaks and epidemics. Since E.U. is committed to halting this loss of biodiversity, breed identification can be used to characterize rare breeds or breeds that are in danger of extinction, since apart from preserving their rare alleles, E.U. spends large amounts of funds/subsidies each year for breeding purebreds and not their crosses.

Ethical approval

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Conflict of Interest

The authors declare that they have no conflict of interest.

Figure Captions

Fig. 1 Geographical map indicating the location of sheep farms

Fig. 2 Total number and length (Mbp) of runs of homozygosity (ROHs) per animal. Different colors indicate different breeds

Fig. 3 Distribution of runs of homozygosity (ROHs) on chromosomes and percentage of coverage per chromosome. Each bar represents the total number of ROHs spotted per chromosome. The red line shows the average percentage of ROHs for each chromosome

Fig. 4 Linkage disequilibrium decay measured by r^2 of SNP pairs over all autosomal chromosomes against physical inter-marker distances (kb) of three sheep breeds

Fig. 5 Effective population size (Ne) estimation for the three sheep breeds over a) the last 100 generations b) a period of 100 to 3,000 generations ago

Fig. 6 Admixture analysis at K=2 and K=3 for 41,014 autosomal LD-pruned SNPs. At K=3 CV error was the smallest, indicating the most probable number of populations in 95 samples. Each individual is presented by a vertical bar. Different colors indicate different clustering groups

Fig. 7 Principal component analysis of the first two axes in 95 sheep samples and 44,954 autosomal SNPs. PC1 explains 8.20% and PC2 4.97% of the total genetic variance

Fig. 8 Distribution of the 3,802 significant SNPs across all three breeds, derived from the log-additive model of inheritance

Fig. 9 Heatmap – dendrogram of the three sheep breeds studied. Samples are presented on the horizontal axis, whereas 3,802 significant SNPs are illustrated on the vertical axis and are grouped with dendrograms based on Euclidean distances

References

Al-Mamun HA, Clark SA, Kwan P, Gondro C (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genetics Selection Evolution 47:90

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246

Anagnostopoulos AK, Katsafadou AI, Pierros V, Kontopodis E, Fthenakis GC, Arsenos G, Karkabounas SC, Tzora A, Skoufos I, Tsangaris GT (2016) Milk of Greek sheep and goat breeds; characterization by means of proteomics. Journal of Proteomics 147:76-84

Banos G, Bramis G, Bush SJ, Clark EL, McCulloch MEB, Smith J, Schulze G, Arsenos G, Hume DA, Psifidi A (2017) The genomic architecture of mastitis resistance in dairy sheep. Bmc Genomics 18:624

Barbato M, Hailer F, Orozco-terWengel P, Kijas J, Mereu P, Cabras P, Mazza R, Pirastru M, Bruford MW (2017) Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Scientific Reports 7:7623

Barbato M, Orozco-TerWengel P, Tapio M, Bruford MW (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics 6

Brito LF, McEwan JC, Miller SP, Pickering NK, Bain WE, Dodds KG, Schenkel FS, Clarke SM (2017) Genetic diversity of a New Zealand multi-breed sheep population and composite breeds' history revealed by a high-density SNP chip. Bmc Genetics 18:25

Buchmann R, Hazelhurst S (2015) The 'Genesis' Manual. University of the Witwatersrand, Johannesburg

Chitneedi PK, Arranz JJ, Suarez-Vega A, Garcia-Gamez E, Gutierrez-Gil B (2017) Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Anim Genet 48:436-446

Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, Napolitano F, Carta A, Usai G, D'Andrea M, Marletta D, Ciampolini R, Riggio V, Occidente M, Matassino D, Kompan D, Modesto P, Macciotta N, Ajmone-Marsan P, Pilla F (2014) Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Animal Genetics 45:256-266

Clarke SM, Henry HM, Dodds KG, Jowett TWD, Manley TR, Anderson RM, McEwan JC (2014) A High Throughput Single Nucleotide Polymorphism Multiplex Assay for Parentage Assignment in New Zealand Sheep. Plos One 9:e93392

Corbin LJ, Liu AYH, Bishop SC, Woolliams JA (2012) Estimation of historical effective population size using linkage disequilibria with marker data. Journal of Animal Breeding and Genetics 129:257-270

Dalvit C, De Marchi M, Dal Zotto R, Zanetti E, Meuwissen T, Cassandro M (2008) Genetic characterization of the Burlina cattle breed using microsatellites markers. Journal of Animal Breeding and Genetics 125:137-144

Dimauro C, Cellesi M, Steri R, Gaspa G, Sorbolini S, Stella A, Macciotta NPP (2013) Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. Animal Genetics 44:377-382

Dodds KG, Auvray B, Newman SAN, McEwan JC (2014) Genomic breed prediction in New Zealand sheep. Bmc Genetics 15:92

Excoffier L, Lischer HEL (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

Fan B, Du ZQ, Gorbach DM, Rothschild MF (2010) Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. Asian-Australasian Journal of Animal Sciences 23:833-847

FAO (1998) Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Management of Small Populations at Risk, Rome.

FAOSTAT (2016) Food and Agriculture Organization of the United Nations. FAOSTAT Database. Rome, Italy: FAO. Retrieved November 18, 2016 from http://www.fao.org/faostat/en/#data/QL.

Felius M, Theunissen B, Lenstra JA (2015) Conservation of cattle genetic resources: the role of breeds. Journal of Agricultural Science 153:152-162

Fernandez ME, Goszczynski DE, Liron JP, Villegas-Castagnasso EE, Carino MH, Ripoli MV, Rogberg-Munoz A, Posik DM, Peral-Garcia P, Giovambattista G (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genet Mol Biol 36:185-191

Fontanesi L, Beretti F, Dall'Olio S, Portolano B, Matassino D, Russo V (2011) A melanocortin 1 receptor (MC1R) gene polymorphism is useful for authentication of Massese sheep dairy products. Journal of Dairy Research 78:122-128

Gaouar SBS, Lafri M, Djaout A, El-Bouyahiaoui R, Bouri A, Bouchatal A, Maftah A, Ciani E, Da Silva AB (2017) Genome-wide analysis highlights genetic dilution in Algerian sheep. Heredity 118:293-301

Garke C, Ytournel F, Bed'hom B, Gut I, Lathrop M, Weigend S, Simianer H (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. Anim Genet 43:419-428

Gelasakis AI, Valergakis GE, Arsenos G, Banos G (2012) Description and typology of intensive Chios dairy sheep farms in Greece. Journal of Dairy Science 95:3070-3079

Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, Estivill X, Moreno V (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644-645

Grossi DA, Jafarikia M, Brito LF, Buzanskas ME, Sargolzaei M, Schenkel FS (2017) Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs. Bmc Genetics 18:6

Guichoux E, Lagache L, Wagner S, Chaumeil P, Leger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591-611

Hatziminaoglou I (2006a) Sheep and goats in Greece and in the world. Ed. Jahoudi, Thessaloniki Greece. 2nd edition.

Hatziminaoglou I (2006b) Sheep and goats in Greece and worldwide. Yahoudi-Yapouli Press, Thessaloniki

Hatziminaoglou I, Liamadis D, Avdi M (2006) Introduction to animal production. Yahoudi-Yapouli Press, Thessaloniki

Hatziminaoglou J, Zervas NP, Boyazoglu J (1990) Prolific dairy sheep breeds in Greece. In: Bougler J. (ed.), Tisserand J.-L. (ed.). Les petits ruminants et leurs productions laitières dans la region méditerranéenne. . Montpellier: CIHEAM. p. 25-30 (Options Méditerranéennes: Série A.Séminaires Méditerranéens; n. 12)

Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. Genome 53:876-883

Heaton MP, Leymaster KA, Kalbfleisch TS, Kijas JW, Clarke SM, McEwan J, Maddox JF, Basnayake V, Petrik DT, Simpson B, Smith TPL, Chitko-McKown CG, Consortium ISG (2014) SNPs for Parentage Testing and Traceability in Globally Diverse Breeds of Sheep. Plos One 9

Howrigan DP, Simonson MA, Keller MC (2011) Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. Bmc Genomics 12:460

Ishibashi K, Hara S, Kondo S (2009) Aquaporin water channels in mammals. Clinical and Experimental Nephrology 13:107-117

Keller MC, Visscher PM, Goddard ME (2011) Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. Genetics 189:237-249

Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Al Cavanagh J, Barris W, Schnabel RD, Taylor JF, Raadsma HW (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. Bmc Genomics 9:187

Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, Consortium ISG (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. Plos Biology 10:e1001258

Kominakis A, Hager-Theodorides AL, Zoidis E, Saridaki A, Antonakos G, Tsiamis G (2017) Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. Genetics Selection Evolution 49:41

Ku CS, Naidoo N, Teo SM, Pawitan Y (2011) Regions of homozygosity and their impact on complex diseases and traits. Human Genetics 129:1-15

Lenstra JA, Groeneveld LF, Eding H, Kantanen J, Williams JL, Taberlet P, Nicolazzi EL, Solkner J, Simianer H, Ciani E, Garcia JF, Bruford MW, Ajmone-Marsan P, Weigend S (2012) Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Anim Genet 43:483-502

Leroy G (2014) Inbreeding depression in livestock species: review and meta-analysis. Animal Genetics 45:618-628

Ligda C, Altarayrah J, Georgoudis A, Consortium E (2009) Genetic analysis of Greek sheep breeds using microsatellite markers for setting conservation priorities. Small Ruminant Research 83:42-48

Loukovitis D, Siasiou A, Mitsopoulos I, Lymberopoulos AG, Laga V, Chatziplis D (2016) Genetic diversity of Greek sheep breeds and transhumant populations utilizing microsatellite markers. Small Ruminant Research 136:238-242

Lu D, Sargolzaei M, Kelly M, Li C, Vander Voort G, Wang Z, Plastow G, Moore S, Miller SP (2012) Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. Front Genet 3:152

Mastrangelo S, Di Gerlando R, Tolone M, Tortorici L, Sardina MT, Portolano B, International Sheep Genomics C (2014) Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC Genet 15:108

Mastrangelo S, Portolano B, Di Gerlando R, Ciampolini R, Tolone M, Sardina MT, Con ISG (2017) Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. Animal 11:1107-1116

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. Genome Biol 17:122

McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, MacLeod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF (2008) Runs of homozygosity in European populations. American Journal of Human Genetics 83:359-372

Meadows JRS, Chan EKF, Kijas JW (2008) Linkage disequilibrium compared between five populations of domestic sheep. Bmc Genetics 9:61

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Molecular Ecology Resources 11:314-322

Mobasheri A, Barrett-Jolley R (2014) Aquaporin Water Channels in the Mammary Gland: From Physiology to Pathophysiology and Neoplasia. Journal of Mammary Gland Biology and Neoplasia 19:91-102

Murtagh F, Legendre P (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification 31:274-295

Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C, Pirani A, Soans C, Brew F, Jorjani H, Evans G, Simpson B, Tosser-Klopp G, Brauning R, Williams JL, Stella A (2015) SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. BMC Genomics 16:283

Niu LL, Li HB, Ma YH, Du LX (2012) Genetic variability and individual assignment of Chinese indigenous sheep populations (Ovis aries) using microsatellites. Animal Genetics 43:108-111

Pariset L, Cuteri A, Ligda C, Ajmone-Marsan P, Valentini A, Consortium E (2009) Geographical patterning of sixteen goat breeds from Italy, Albania and Greece assessed by Single Nucleotide Polymorphisms. BMC Ecol 9:20

Pariset L, Mariotti M, Gargani M, Joost S, Negrini R, Perez T, Bruford M, Marsan PA, Valentini A (2011) Genetic Diversity of Sheep Breeds from Albania, Greece, and Italy Assessed by Mitochondrial DNA and Nuclear Polymorphisms (SNPs). The scientific world journal 11:1641-1659

Peter C, Bruford M, Perez T, Dalamitra S, Hewitt G, Erhardt G (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. Animal Genetics 38:37-44

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: Models and data. American Journal of Human Genetics 69:1-14

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81:559-575

Purfield DC, McParland S, Wall E, Berry DP (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One 12:e0176780

R Development Core Team (2010) R: A language and environment for statistical computing. In. R Foundation for Statistical Computing, Vienna, Austria

Raadsma HW, Thomson PC, Zenger KR, Cavanagh C, Lam MK, Jonas E, Jones M, Attard G, Palmer D, Nicholas FW (2009) Mapping quantitative trait loci (QTL) in sheep. I. A new male framework linkage map and QTL for growth rate and body weight. Genet Sel Evol 41:34

Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM (2011) Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. Animal Genetics 42:613-620

Sardina MT, Tortorici L, Mastrangelo S, Di Gerlando R, Tolone M, Portolano B (2015) Application of microsatellite markers as potential tools for traceability of Girgentana goat breed dairy products. Food Research International 74:115-122

Schaffner SF (2004) The X chromosome in population genetics. Nature Reviews Genetics 5:43-51

Tanaka Y, Morishita Y, Ishibashi K (2015) Aquaporin10 is a pseudogene in cattle and their relatives. Biochemistry and Biophysics Reports 1:16-21

Valergakis GE, Gelasakis AI, Oikonomou G, Arsenos G, Fortomaris P, Banos G (2010) Profitability of a dairy sheep genetic improvement program using artificial insemination. Animal 4:1628-1633

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91:4414-4423

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2011) gplots: Various R programming tools for plotting data. R Foundation for Statistical Computing, Vienna, Austria. http://CRAN.R-project.org/package=gplot

Yang JA, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. American Journal of Human Genetics 88:76-82

Zanella R, Peixoto JO, Cardoso FF, Cardoso LL, Biegelmeyer P, Cantao ME, Otaviano A, Freitas MS, Caetano AR, Ledur MC (2016) Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. Genetics Selection Evolution 48:24

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV (2017) OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Research 45:D744-D749

Zeder MA (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proceedings of the National Academy of Sciences of the United States of America 105:11597-11604

Zhang QQ, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G (2015) Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. Bmc Genetics 16:88

Zhao FP, Wang GK, Zeng T, Wei CH, Zhang L, Wang HH, Zhang SZ, Liu RZ, Liu Z, Du LX (2014) Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations. Livestock Science 170:22-29