

RESEARCH ARTICLE

Genetic differentiation of mainland-island sheep of Greece: Implications for identifying candidate genes for long-term local adaptation

Antonios Kominakis¹, Eirini Tarsani^{1*}, Ariadne L. Hager-Theodorides¹, Ioannis Mastranestasis², Dimitra Gkelia³, Ioannis Hadjigeorgiou¹

1 Department of Animal Science, Agricultural University of Athens, Athens, Greece, **2** Breeder's Association of the Lesvos sheep, Anaxos Lesvos, Greece, **3** Association of Pastoral Farmers of Epirus, Ioannina, Greece

* etarsani@aua.gr



OPEN ACCESS

Citation: Kominakis A, Tarsani E, Hager-Theodorides AL, Mastranestasis I, Gkelia D, Hadjigeorgiou I (2021) Genetic differentiation of mainland-island sheep of Greece: Implications for identifying candidate genes for long-term local adaptation. PLoS ONE 16(9): e0257461. <https://doi.org/10.1371/journal.pone.0257461>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: April 10, 2021

Accepted: September 1, 2021

Published: September 16, 2021

Copyright: © 2021 Kominakis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are subject to restrictions with regard to their publicly availability imposed by the 'Terra Lemnia' project due to their containing information that could compromise participant privacy and were used under license for the current study. These data can be gained by contacting Mr. George Dimitropoulos (e-mail: egeorged@med-ina.org) upon reasonable request and with permission of the 'Terra Lemnia' project.

Abstract

In Greece, a number of local sheep breeds are raised in a wide range of ecological niches across the country. These breeds can be used for the identification of genetic variants that contribute to local adaptation. To this end, 50k genotypes of 90 local sheep from mainland Greece (Epirus, $n = 35$ and Peloponnesus, $n = 55$) were used, as well as 147 genotypes of sheep from insular Greece (Skyros, $n = 21$), Lemnos, $n = 36$ and Lesvos, $n = 90$). Principal components and phylogenetic analysis along with admixture and spatial point patterns analyses suggested genetic differentiation of 'mainland-island' populations. Genome scans for signatures of selection and genome-wide association analysis (GWAS) pointed to one highly differentiating marker on OAR4 ($F_{ST} = 0.39$, $FLK = 21.93$, FDR p-value = 0.10) that also displayed genome wide significance (FDR p-value = 0.002) during GWAS. A total number of 6 positional candidate genes (*LOC106990429*, *ZNF804B*, *TEX47*, *STEAP4*, *SRI* and *ADAM22*) were identified within 500 kb flanking regions around the significant marker. In addition, two QTLs related to fat tail deposition are reported in genomic regions 800 kb downstream the significant marker. Based on gene ontology analysis and literature evidence, the identified candidate genes possess biological functions relevant to local adaptation that worth further investigation.

Introduction

Post domestication sheep (*Ovis aries*) have spread and distributed over a wider geographic range around the globe and adapted to a variety of diverse, often harsh, environments. Following the global trend, a number of indigenous sheep breeds are found in Greece, in a range of ecological and environmental conditions [1]. Amongst them, two local breeds (Katsika and Oreino) are traditionally kept on the alpine climate zone of the Pindos mountain chain under a vertical transhumance system with livestock displacement from lowlands to highlands in summer and *vice versa*, in wintertime.

Funding: This work was partially supported by the 'Terra Lemnia' project funded by the 'MAVA Fondation pour la Nature'. Ioannis Hadjigeorgiou received this financial support.

Competing interests: The authors have declared that no competing interests exist.

At the south western part of the country, and specifically on natural pastures of Peloponnesus' mountains, local sheep populations prevail, with most distinctive specimens that of the rare breed known as Kokovitiko. Here the pastoral landscape is characterized by a rough terrain covered by diverse and lush vegetation due to the higher amount of rainfall than the eastern part of this peninsula. There is no transhumance practice in this area.

At the other geographical extreme of the country, and specifically on two islands of the North Aegean Sea i.e. Lemnos and Lesvos, two more breeds named after the homonymous islands, are kept in a semi-arid climate zone with low winter temperatures and longer summer droughts ([2,3]). In a similar climatic landscape, in the middle of the Aegean Sea, the sheep of Skyros have been bred in isolation for years on the homonymous island.

The above local sheep present valuable genetic resources and are critical for preserving local genetic diversity and ecosystem integrity, as well as food chain productivity and resilience. With regard to the latter, milk from these ewes is used to produce famous, high quality dairy products, of which some cheeses have been granted protected designations of origin, such as Feta in Epirus and Peloponnesus, Ladotyri in Lesvos (PDO), Kalathaki (PDO), and Melichloro in Lemnos.

Local sheep breeds traditionally raised under extreme environments are an excellent model for studying genetic adaptation [4]. Under these conditions, selective breeding for improved (re)productive performance is limited or non-existent, leaving natural selection as the primary evolutionary acting mechanism. In such cases, natural selection causes distorted patterns of genetic variation at and around selected loci, resulting in local molecular footprints that are manifested as reduced levels of genetic diversity, extended homozygosity tracts and increased differentiation from other populations [5]. Adaptation specific footprints can nowadays be detected using genome-wide data and methods for tracking specific signals in the genome (genome scans). Some scans look for alleles that are correlated with particular features of the environment. Others look for highly differentiated genomic regions between extreme populations that exceed neutral expectation, based on the notion that local adaptation will increase the frequency of locally beneficial alleles and the variation among populations in allele frequencies [6]. Regions in the genome that have been preferentially increased in frequency and fixed in a population because of their functional importance in specific processes are called signatures of selection (SS) [7] and can provide valuable information on the genetic mechanism of local adaptation and speciation [8].

In sheep, SS have been detected in a variety of wild and domestic sheep populations ([9–17]). Putative genomic regions and genes associated with breed formation [18], climate-mediated selection pressure [19] and local environmental adaptation ([10,20]) have also been successfully identified. The above studies highlight the potential of using local sheep breeds raised in different environments as a means to detect adaptive loci contributing to environmental adaptation.

Driven by the diversity of the ecological niches (variable topography, vegetation and climatic conditions) across the country, this study was primarily designed to identify plausible genetic loci associated with local adaptation (LA) in sheep. To this end, mainland sheep populations (Epirus and Peloponnesus) were compared to insular ones (Skyros, Lemnos, and Lesvos), assuming environmentally driven selection throughout: i) has been the primary driving force for LA, and ii) has affected the genomes of sheep raised in similar habitats in the same way.

In the present study, first we examined whether mainland-island present genetically different entities and then conducted genome scans and genome-wide association analysis in attempts to address two questions with regard to local adaptation (LA). Specifically: a) does LA

leave detectable differential footprints in sheep genome? and b) can genetic differentiations lead to detection of functionally relevant genes to LA?

Materials and methods

Ethics statement

All applicable ethical guidelines for the care and use of animals were followed and animal blood samples were collected by trained personnel under strict veterinary rules. All samples and data in our study were collected under the consent of the sheep breeders. The protocol of the present study was approved by the Research Ethics and Ethics Committee of the Agricultural University of Athens (no 54/2020 and no 18/2021) according to the article 23 of law 4521/2018 of the Greek government.

Data and quality control

A total number of $n = 311$ sheep blood samples were collected from 17 herds across 5 regions (3 island and 2 mainland) of Greece (see [S1 Table](#) and [S1 Fig](#)). All animals were genotyped with the Illumina Ovine 50K SNP array by Neogen Europe Ltd (Ayr, UK) and genotypes presented with a GenCall score greater than 0.30 were considered called. Quality control (QC) was performed at a sample and marker level. At the animal level, $n = 12$ samples were excluded due to call rate lower than 0.95. An additional number of 62 related animals were removed based on an identity by descent (IBD) degree of recent shared ancestry ($\hat{\pi}$) threshold of 0.25 (second-degree relatives). A $\hat{\pi}$ threshold higher than the typical one (0.1875 i.e. halfway between third- and second-degree relatives [21]) was used here to avoid potential bias due to cryptic relatedness and the removal of a large number of samples. At the marker level, $n = 14,402$ of the 53,295 original number of SNPs were excluded due to: call rate lower than 0.95, minor allele frequency (MAF) lower than 0.05, deviation from Hardy–Weinberg equilibrium (HWE) using a Fisher exact test $p\text{-value} < 10^{-5}$ and linkage disequilibrium (LD) r^2 levels ($r^2 > 0.50$, window size: 50 SNPs, increment: 5 SNPs) and autosomal mapped SNPs. Application of QC criteria at the marker level resulted in a total number of 38,893 autosomal mapped SNPs retained for further analyses. QC was performed using the SNP & Variation Suite software (version 8.9.0). The distribution of samples across regions is shown in [S1 Table](#) and [S1 Fig](#).

Detection of genetic structure and phylogeny analysis

Detection of genetic structure was assessed using various approaches. First, the genomic relationship matrix (GRM) between all pairs of individuals using genotypes of the 38,893 SNPs was calculated. Next, Principal Component Analyses (PCA) of animals' genetic relationships was conducted and the first two PCs were plotted on the two dimensional space to resolve genetic patterns in the samples. The above analyses were performed using SNP & Variation Suite software (version 8.9.0).

Pairwise Reynold's genetic distances between regions were automatically obtained by the hapFLK software (version 1.4, <https://forge-dga.jouy.inra.fr/projects/hapflk>, [22]) and they were graphically visualized using the NeighborNet graph in SplitsTree5 [23].

Admixture analysis (AA) was also applied to infer assignment probabilities for mainland or island membership for individual samples. To this end, the ADMIXTURE software [24] was employed with K-value equal to 2, assuming two ancestor populations i.e. mainland and island. Admixture results were graphically visualized using Distruct from CLUMPAK (<http://clumpak.tau.ac.il/> [25]).

Spatial point patterns analysis

A spatial point patterns analysis (SPPA) was performed using the elements of the first PC of animals' genetic relationships as the events (observations) to investigate whether genomic relationships are distributed randomly or follow a specific pattern (indicative of structure) across the geographical landscape of sampling sites. A Gaussian kernel (or covariance function) was employed to fit the first-order intensity (number of observations per unit area) of a spatial point pattern using procedure SPP in SAS (ver.9.4). Spatial patterns were then resolved by constructing plots of the nearest neighbor distance and the smoothed kernel density of animals' relationships on the geographical landscape with X and Y representing longitude and altitude of the sampling sites, respectively. The global Moran's index (I) [26] and the Z score were finally estimated via the Variogram procedure in SAS (ver. 9.4) to test the null hypothesis of no spatial, positive or negative spatial auto-correlation for sample values.

Estimation of LD scores and LD decay by distance

First, LD scores for each of the 38,893 markers using two maximum windows sizes of 500 kb and 1000 kb were calculated. Windows were always centered around the marker being scored, unless the window was at the beginning or the end of a region. This estimation was carried out for the island and mainland populations. Finally, the Wilcoxon Rank Sum test as implemented in procedure NPAR1WAY in SAS (ver.9.4) was applied to test the hypothesis that the two LD score distributions differ with respect to their medians.

Second, the decline rate of LD with inter-marker distance (in kb) within genes (using the OAR4.0 annotation) was estimated by fitting the following equation:

$$y_i = \frac{1}{(1 + 4\beta d_i)} + e_i$$

where y_i is the estimated \hat{r}^2 after correcting for sampling error using the formula of Hill and Robertson [27] $\hat{r}^2 = (\hat{r}^2 - 1/N)(1 - 1/N)$ with N twice the number of sampled individuals. Furthermore, d_i is the inter-marker distance (in kb) for marker pair i, β is the coefficient that describes the decline rate of LD with distance (large values indicate low extent of LD) and e_i is a random residual with $e_i \sim (0, s^2)$. A maximum inter-marker distance of 2,500 kb was considered here. This analysis was carried out with procedure NLMIXED in SAS (ver. 9.4) using r^2 values of the markers that had no negative or 0 values (because of correction for sample size). This selection resulted in a total number of 35,911 and 37,276 r^2 values for mainland and island sheep, respectively. Estimation of LD scores and LD r^2 values was performed with SNP & Variation Suite software (version 8.9.0).

Detection of signatures of selection

Signatures of positive selection were detected using various approaches, as described below. First, the F_{ST} ([28,29]) statistic of genetic differentiation based on allele frequency differences between mainland-island populations was calculated for the 38,893 markers, via SNP & Variation Suite software (version 8.9.0). The negative F_{ST} values obtained for 14,214 SNPs were set to 0, since negative values have no biological interpretation. Raw F_{ST} values were then ranked and values exceeding the conservative 99.999th percentile (0.3554) bound of the empirical F_{ST} distribution were considered as outliers. As individual F_{ST} values may be subject to genotyping errors, we also estimated genome-wide smoothed F_{ST} estimates, by averaging F_{ST} values across 500 kb windows. Windows (n = 383 out of 5,243) with less than four SNPs were discarded, which resulted in on an average of 7.82 (SD = 2.14) SNPs per window (minimum = 4,

maximum = 16). Here, smoothed F_{ST} values exceeding the 99.999th percentile (0.1357) bound were considered as outliers pointing to genomic regions under positive selection.

As a complementary approach to F_{ST} , we used the FLK statistic [30] to detect regions of outlying differentiation between mainland and island populations. FLK is an extension of Lewontin and Krakauer (LK) test [31] which accounts for population size heterogeneity and for the hierarchical structure between populations. Specifically, FLK tests the neutrality of polymorphic markers by contrasting their allele frequencies in a set of populations against what would be expected under neutral evolution. Under the null hypothesis of no selection, a neighbor joining tree based on a matrix of Reynolds genetic distances is built and the length of branch is expected to be proportional to the amount of genetic drift in each population.

For each SNP, the FLK estimate and its p-value were obtained using the hapFLK software (version 1.4, <https://forge-dga.jouy.inra.fr/projects/hapflk>, [22]). As input for FLK analysis, we used a ped PLINK file that was constructed with all genotypes and two population identifiers i.e., mainland (n = 90) and island (n = 147) sheep populations. FLK analysis was run for all chromosomes using 20 expectation maximization iterations. This analysis was carried out with hapFLK software (version 1.4, <https://forge-dga.jouy.inra.fr/projects/hapflk>, [22]). To identify SS, FLK p-values were finally corrected for multiple comparisons using the false-discovery rate (FDR) method in R (<http://www.r-project.org/>). Here, a marker with FDR p-value below or equal to 0.1 was considered significant.

A Bayesian framework was additionally applied to evaluate whether each locus is affected by selection using a reversible-jump Markov chain Monte Carlo (MCMC) approach and the BayeScan 2.0 software (<http://www.cmpg.unibe.ch/software/bayescan/>, [32]). During this analysis, the following settings were specified: burn-in of 10,000 iterations followed by 50,000 iterations, a thinning interval of 10 and prior odds equal to 1000 for the neutral model. Following documentation [32], we used a posterior probability (prob) higher than 0.99 to define a marker under selection.

Assessment of selective process

First, we estimated summary statistics of F_{ST} , genetic diversity within (estimated as the average observed Heterozygosity (H_o) of the two populations) and between populations (calculated as difference of H_o between the two populations) in 1 Mb flanking genomic regions surrounding the F_{ST} outlier(s) marker. Then, for the three parameters, non-parametric loess smooth curves obtained via the LOESS procedure in SAS (ver.9.4) were estimated and plotted to detect characteristic patterns after ‘removing’ local noise. Finally, following Booker et al. [6], a visual examination of the characteristic patterns of the loess smooth curves was performed to determine the type of selective process (global adaptation, local adaptation, or genetic drift) that has occurred.

Genome-Wide Association Analysis (GWAS)

A single locus mixed (additive) model GWAS using the EMMAX algorithm [33] was applied to detect significant associations between SNPs (fitted as fixed effect covariates) and the two origins coded as a binary dependent variable (0: island, 1: mainland). No other covariate (e.g., herd) was included in the model as these were confounded with the dependent variable. During this analysis, animals’ relationships were included in the mixed model as random effects via the GRM. Note that inclusion of the GRM has been shown to correct for possible population structure and stratification in the data [34]. Genome-wide significance of SNPs was declared using a FDR p-value lower than 0.05. Q-Q plots were used to analyze the extent to which the observed distribution of the test statistic followed the expected (null) distribution.

This analysis along with the estimation of the genomic inflation factor (λ) was done to assess potential systematic bias due to population structure or to the analytical approach [35]. All above analyses were carried out with the SNP and Variation Suite v8.3.4 (Golden Helix, Inc. 2015).

Detection of positional candidate genes and reported QTLs

First, we searched within 500kb downstream and upstream flanking regions of each significant marker for positional candidate genes using the NCBI database ([36]) and the Oar_v4.0 assembly ([37]). Then, the ids of the significant SNPs were submitted to the Variant Effect Predictor (VEP, <https://www.ensembl.org/Tools/VEP>, [38]) tool to search for published QTLs according to Oar_v3.1 assembly. VEP retrieves QTL information via connections with Animal QTL database (Animal QTLdb) and Online Mendelian Inheritance in Animals (OMIA) database for *Ovis aries*. Positions of QTLs were remapped from Oar_v3.1 to Oar_v4.0 assembly using the Genome Remapping Service from NCBI database [39].

Functional enrichment analysis

Positional candidate genes were subject to Gene Ontology (GO) Biological Process (BP) enrichment analysis using the ClueGO V2.5.7 [40] plug-in in Cytoscape V3.7.2 (<http://cytoscape.org/>, [41]). During this analysis, the latest update (released: 05.03.2021) of GO BP annotations were used for the species of *Ovis aries*. Each time, the input genes were compared with a reference list of ovine genes using the following settings: minimum number of genes per BP = 1, minimum percentage of genes included in GO BP = 3%, minimum GO levels = 3, maximum GO levels = 8 and minimum kappa score for GO BP network connectivity = 0.4. Here, GO BPs with FDR p-values lower than 0.05 for the right-sided hypergeometric test were considered as significantly enriched. The CluePedia V1.5.7 [42] plug-in of Cytoscape(<http://cytoscape.org/>, [41]) was also used to illustrate the significantly enriched GO BPs with their involved genes.

Results

Genetic structure

The two-dimensional plot of the first two PCs derived from PCA on animals' genomic relationships is shown in Fig 1. PCA revealed the formation of four clusters: one with Lemnos (green color) and Lesvos (brown color) samples, a second with genotypes from Skyros island (red) in the center of the plot, a third with genotypes from Peloponnesus (dark grey color) and the Oreino breed (Epirus, blue color) and finally a fourth cluster with genotypes from Katsika breed (Epirus, blue color, upper right). In the same plot, the first PC that accounted for most of the variation of animals' relationships (7.5%) had the highest discriminatory power and separated island (Lemnos, Lesvos and Skyros) from mainland populations (Peloponnesus and Epirus). The second PC, which accounted for 3.5% of the variation of animals' relationships, classified Epirus genotypes into two groups, each one corresponding to a different breed (Oreino and Katsika). In accordance with PCA results, the unrooted generated phylogenetic tree also suggested genetic discrimination of mainland-island genotypes with one branch containing Lemnos, Lesvos and Skyros samples and other samples from Peloponnesus and Epirus. At the bottom of Fig 1, the results of AA assuming two ancestral populations ($K = 2$) were consistent with the two previous analyses, showing a differential distribution of assignment probabilities of samples across origins with higher assignment probabilities of island membership (denoted as blue color) for samples of Lemnos, Lesvos and Skyros, and diminishing for samples of Peloponnesus and Epirus (Fig 1).

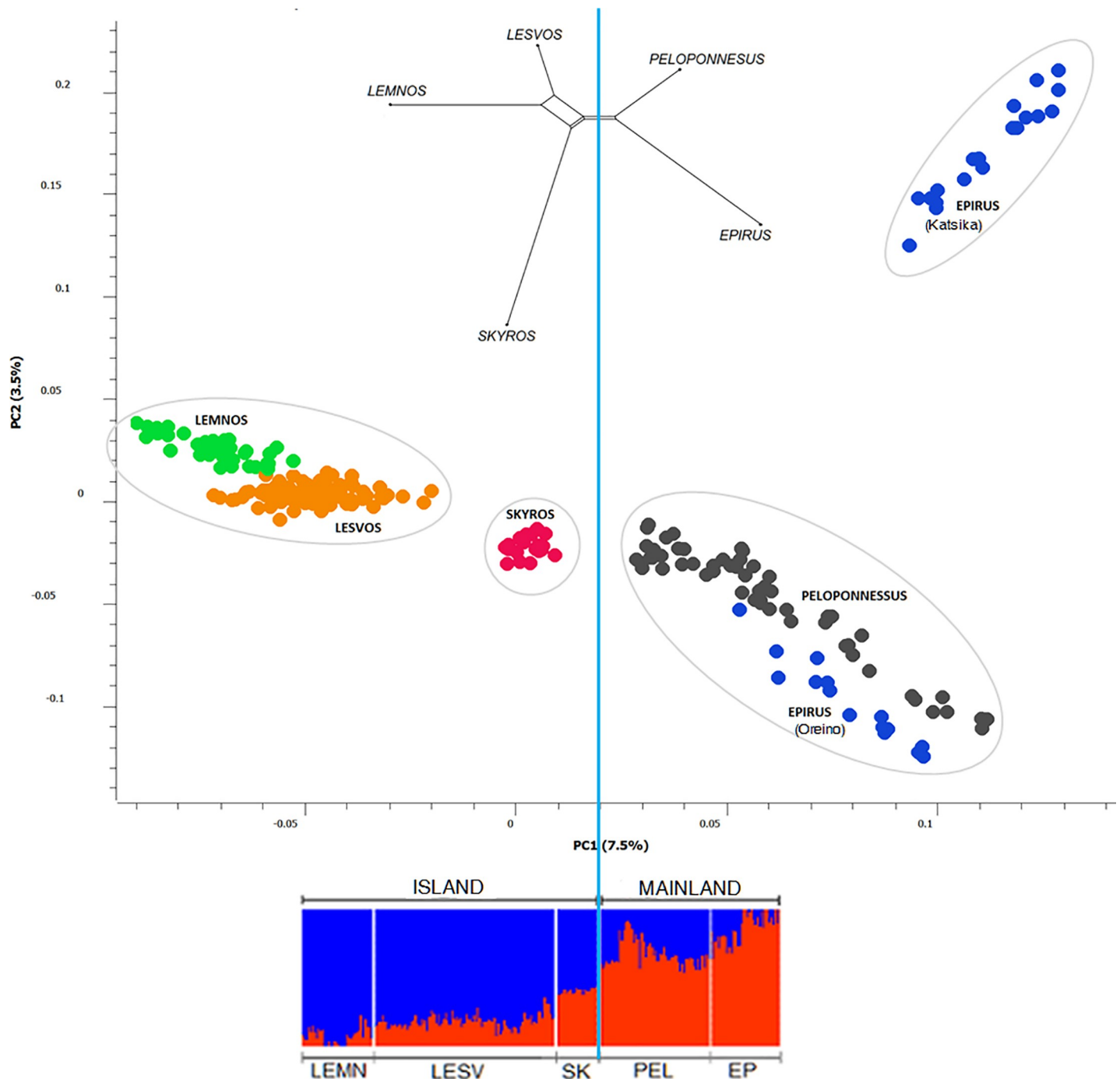


Fig 1. Two-dimensional plot of the first two PCs derived from PCA on animals' genomic relationships. A total number of four clusters were created, one with genotypes from Lemnos (green) and Lesvos (brown) islands, a second with genotypes from Skyros island (red), a third with genotypes from Peloponnesus (dark grey) and Oreino breed of Epirus (blue) and finally a fourth with genotypes of Katsika breed of Epirus (blue). An unrooted phylogenetic tree depicting pairwise Reynolds' genetic distances between regions was also constructed (top). Individual assignment probabilities generated with ADMIXTURE ($K = 2$) are shown at the bottom of the graph. Each color represents a cluster, and the ratio of colored bars is proportional to the assignment probability of an individual to each cluster. Blue and red colors represent island and mainland membership, respectively. A possible partition of populations into mainland-island origin is indicated by the vertical blue colored line in the graph's center.

<https://doi.org/10.1371/journal.pone.0257461.g001>

Spatial point patterns analysis

Results of SPPA of animals' genomic relationships (accounted here by the first PC) are depicted on Fig 2. As the nearest neighbor distance (Fig 2A) and the Gaussian kernel intensity

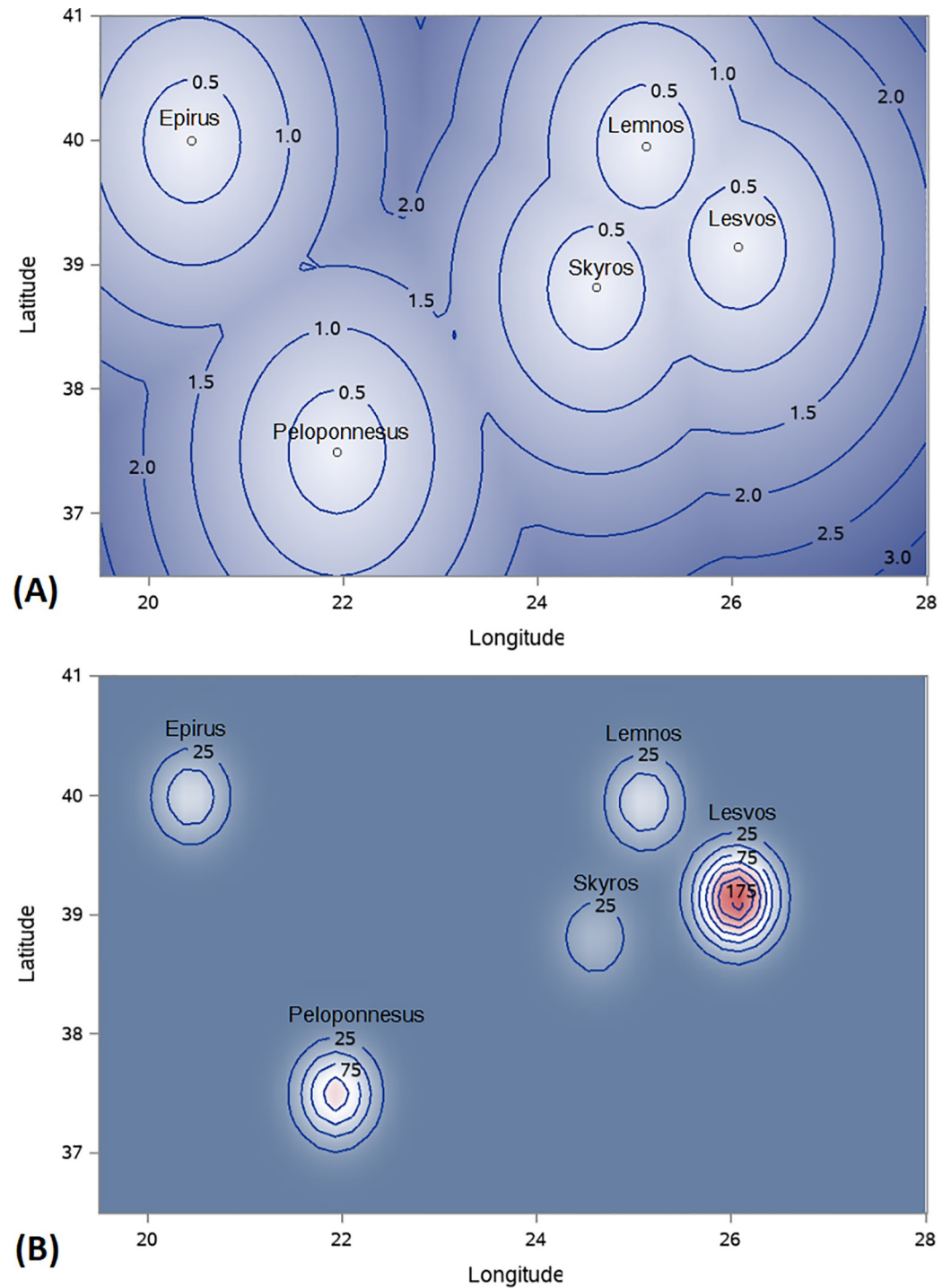


Fig 2. Plots of the mean nearest neighbor distances (A) and Gaussian kernel smoothed intensity (B) of animals' genomic relationships along the geographical landscape. The plots indicate that animals sampled from the same location have higher genetic similarity than distantly placed or randomly selected samples.

<https://doi.org/10.1371/journal.pone.0257461.g002>

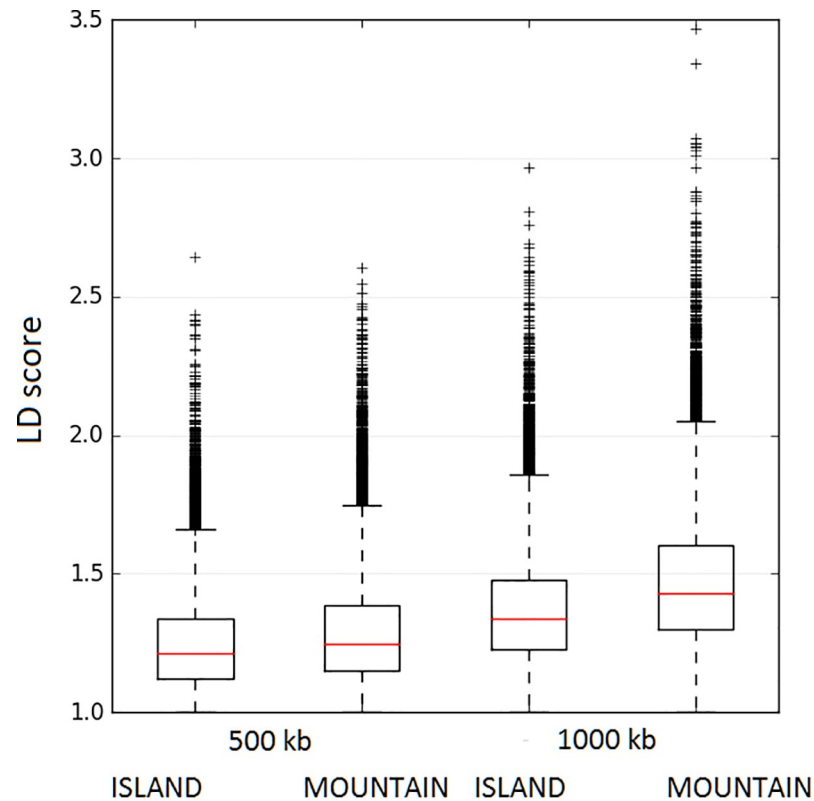


Fig 3. Box plots of LD scores for 500 kb and 1000 kb windows for mainland- island sheep populations.

<https://doi.org/10.1371/journal.pone.0257461.g003>

(Fig 2B) plots portray, animals' genomic relationships are not randomly distributed throughout the study area, and samples with similar values tend to cluster together according to geographical coordinates of sampling locations. In other words, animals sampled from the same location have greater genetic similarity than animals from distant locations or randomly selected. These results were confirmed by the estimated global Moran's I index ($I = 0.8012$, $p < 0.0001$) and the positive Z score ($Z = 83.5$) supporting the hypothesis of non-zero, positive spatial autocorrelation for genomic relationships, respectively (results not shown). Overall, these findings suggest that the five sheep populations are distinct genetic entities that can be further classified based on their origin, mainland vs. island, allowing for further investigation of their genomic characteristics.

Estimation LD score and rate of LD decay by distance

Fig 3 shows Box plots of LD scores across the two origins for two windows sizes: 500 kb and 1000 kb. Median LD score was higher in the mainland vs. island sheep at 500 kb (1.2479 vs. 1.2103, $p < 0.001$, Wilcoxon Two-Sample test) and the difference was more pronounced for the 1000 kb windows (1.4314 vs. 1.3363, $p < 0.001$, Wilcoxon Two-Sample test). Results of LD score estimates were confirmed by β coefficient estimates obtained from nonlinear regression analysis with β estimate higher for island (is) ($\beta_{is} = 0.1173 \pm 0.000953$, $CI(95\%) = 0.1154-0.1191$) when contrasted to its respective counterpart for mainland (m) sheep ($\beta_m = 0.1062 \pm 0.000916$, $CI(95\%) = 0.1044-0.1080$) (results not shown). Note that as the $CI(95\%)$ of the two β estimates do not overlap, the two estimates are statistically significantly different. As β coefficients describe the decline rate of LD by distance, high and low β estimates imply low

Table 1. Significant SNPs obtained by allele frequency-based F_{ST} , FLK analysis and GWAS.

Marker	OAR	Position	F_{ST} (raw)	FLK (p-value, p_{FDR})	GWAS			Major allele (frequency)	
					p-value	$-\log_{10}$ (p-value)	p_{FDR}		
s25289.1	4	72,769,785	0.3909	2.83e-06, 0.11	5.1482e-08	7.2883	0.002	0.1188	B (0.717)

<https://doi.org/10.1371/journal.pone.0257461.t001>

and high LD persistency, respectively. Thus, both LD score and β coefficient estimates indicate lower extent of LD (or lower LD persistency) for the island compared to mainland sheep.

Signatures of selection and GWAS signal(s)

Table 1 summarizes the outlier markers obtained from F_{ST} and FLK along with results of GWAS. Only one marker i.e. s25289.1 (OAR4, position 72,769,785) had a raw F_{ST} value (0.3909), exceeding the 99.999th bound (0.3554) (Fig 4). The same marker displayed a p_{FDR} -value slightly higher than 0.10 for the FLK statistic (FLK p-value = 2.83e-06) and had a p_{FDR} -value (0.002 and $-\log_{10}$ (p-value) = 7.7260) lower than the 0.05 ($-\log_{10}$ (p-value) = 5.890) threshold denoting genome wide significance during GWAS (Fig 4). With regard to the latter, the

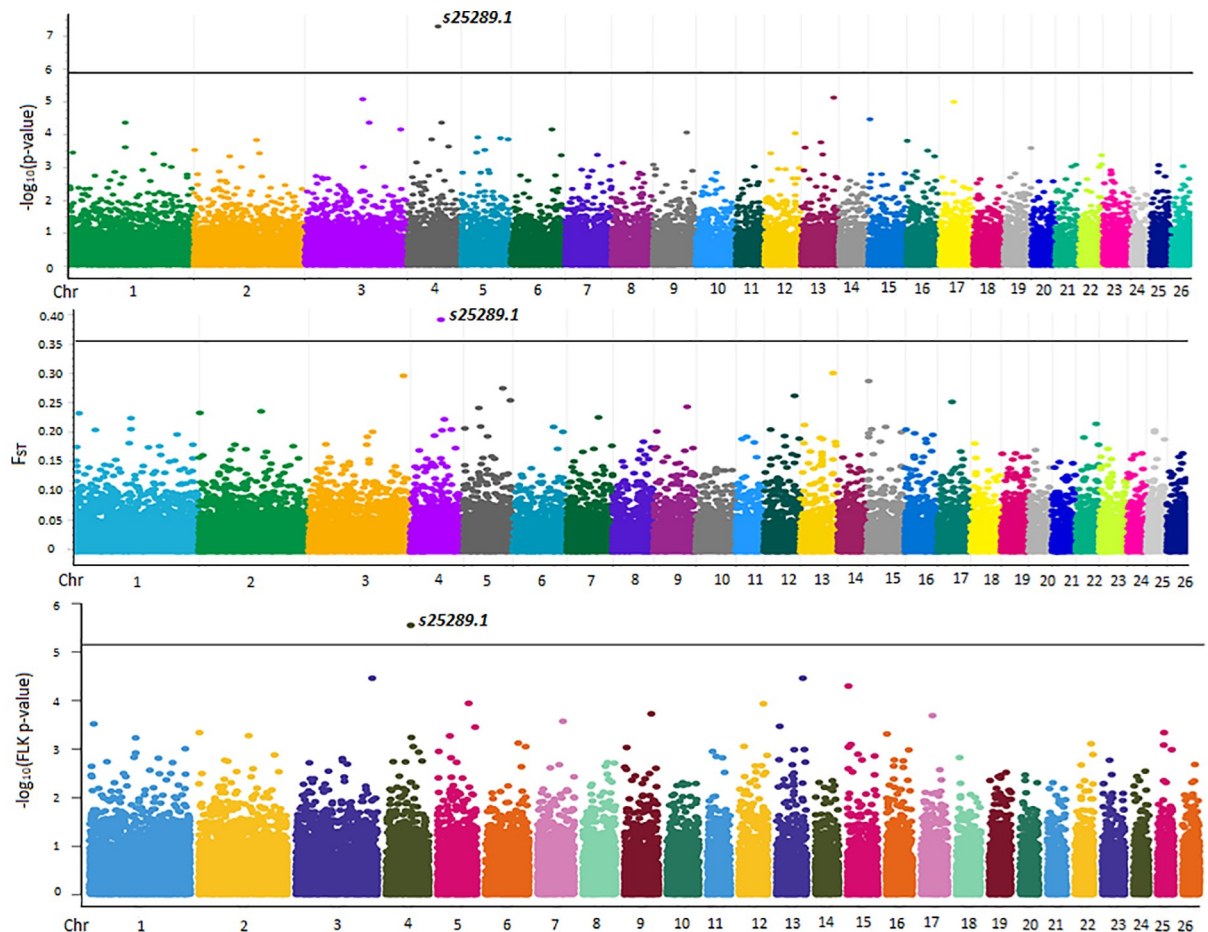


Fig 4. Manhattan plots of $-\log_{10}$ (p-values) of SNPs (top), raw F_{ST} estimates (middle) and FLK statistic (bottom) across autosomes. Grey line on the upper Manhattan plot denote threshold ($-\log_{10}$ (p-value) = 5.893) for genome wide significance, in the middle the 99.999th percentile bound (0.3554) and the threshold for the FLK statistic (bottom). The id of significant SNP is also presented. FLK plot was constructed with the CMplot package (<https://github.com/YinLiLin/R-CMplot>) in R (<http://www.r-project.org/>).

<https://doi.org/10.1371/journal.pone.0257461.g004>

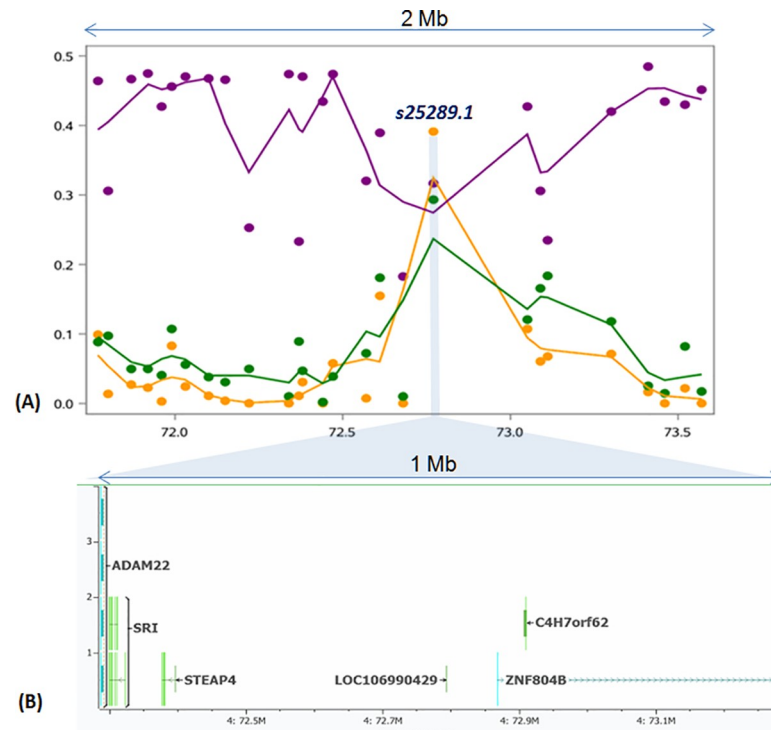


Fig 5. (A) Loess curves of F_{ST} (orange line), within diversity (purple line) and between diversity (green line) for 1 Mb genomic distances around the F_{ST} outlier marker (*s25289.1*). Smoothed lines were obtained via LOESS procedure in SAS (ver.9.4) using a smoothing parameter equal to 0.21 for the three variables and 27 points (markers) (B) Positional candidate genes within 500 kb down- and up-stream the significant marker (*s25289.1*) are also shown here.

<https://doi.org/10.1371/journal.pone.0257461.g005>

estimated λ value of the observed p-values ($\lambda = 0.9870$) along with the Q-Q plot of expected vs. observed SNP p-values denoted no population structure or other artifact(s) in the data (S2 Fig). With regard to smoothed F_{ST} estimates, only one window-based estimate ($F_{ST} = 0.1382$) exceeded the 99.999th bound (0.1357) of the empirical window-based F_{ST} distribution and this window included marker *s25289.1*. Results of F_{ST} outlier(s) detection were also confirmed with Bayescan with a posterior probability (prob) equal to 0.9946 (p_{FDR} -value = 0.005, $F_{ST} = 0.18836$, results not shown).

Evidence for local adaptation

Patterns of the three summary statistics (F_{ST} , genetic diversity within (avHo) and between populations (dHo)) in 1 Mb genomic regions surrounding the F_{ST} outlier marker are shown in Fig 5A. As shown in Figure, the peak of F_{ST} values (orange line) at marker position is accompanied by a decrease in avHo (purple line) and a notable peak of dHo (green line). The latter, which is typical of long-term local adaptation, excludes the scenario of global adaptation and genetic drift ([6,51]). All of the above analyses indicate that marker *s25289.1* is in a genomic region containing genetic loci that are most likely involved in the long-term LA process.

Positional candidate genes and published QTLs

A total number of 6 positional candidate genes i.e., 5 protein coding genes (*ZNF804B*, *C4H7orf62* or *TEX47*, *STEAP4*, *SRI* and *ADAM22*) and one pseudogene (*LOC106990429*) were detected within the searched genomic areas around the significant marker (*s25289.1*) on OAR4 (Table 2 and Fig 5B). The nearest gene to the significant SNP was *LOC106990429* while

Table 2. Positional candidate genes around the *s25289.1* significant marker on chromosome 4.

Gene ID	Gene description	Extent of gene (bp) ^a	Distance from marker (bp)
LOC106990429	histone-lysine N-methyltransferase SETMAR-like	72777738–72778750	8965
ZNF804B	zinc finger protein 804B	72852055–73433337	82270
CAH7orf62 or TEX47	testis expressed 47	72880869–72895751	125966
STEAP4	STEAP4 metalloproteinase	72360288–72381700	388085
SRI	sorcini	72283566–72307498	462287
ADAM22	ADAM metalloproteinase domain 22	72158944–72275420	494365

^a Positions were based on Oar_v4.0 assembly according to NCBI database.

<https://doi.org/10.1371/journal.pone.0257461.t002>

ADAM22 gene was located furthest away from the marker (494 kb). In addition, two published QTLs (ID126988, 4:66,776,665–70,273,454 and ID127002, 4:68,271,877–69,172,400) related to tail fat deposition are reported, distanced ~800 kb downstream the marker. Note that at a more distant location (~1.62 Mb) upstream the marker resides another member of the *STEAP* family genes, namely *STEAP2* (OAR4: 74,395,870–74,420,792).

Functional enrichment analysis

All genes were recognized by ClueGO. As seen in Fig 6, functional enrichment analysis resulted in 41 significantly enriched GO BPs with three member genes (*TEX47*, *STEAP4* and *SRI*). The maximum number of enriched GO BPs ($n = 32$) were detected for *SRI* gene while *TEX47* gene presented the minimum number of GO BPs ($n = 4$). With regard to the number of most significantly enriched GO BP terms, 3 GO BPs were revealed for *TEX47* gene while one GO BP was found for *STEAP4* and *SRI* genes (S2 Table). Specifically, 3 enriched GO BPs i.e., blue light signaling pathway (GO:0009785), cellular response to blue light (GO:0071483) and blue light photoreceptor activity (GO:0009882) presented the same low FDR p-value (0.006) for *TEX47* gene (S2 Table). For *STEAP4* gene, iron import into cell biological process (GO:0033212) showed the lowest FDR p-value (0.007) while the most significantly enriched GO BP for *SRI* gene was the negative regulation of cardiac muscle contraction (GO:0055118) with FDR p-value as high as 0.006.

Discussion

Different LD patterns in mainland— island sheep

The current study sought to answer the intriguing question of whether LA leaves distinct genome footprints manifested in terms of LD structure. The current findings provided an answer to this question, indicating that mainland sheep have higher LD persistency than island sheep. In both populations, LD decays fast that is typical for local breeds experiencing weak artificial selection in contrast to ‘commercial’ breeds where LD extends for larger genomic distance [43]. Given that the mating system for such breeds is relatively comparable across the country, the slight differences in LD decay between the two populations are most likely due to differences in effective population sizes. Yang et al. [44] have also observed overall slow decay rate and a high level of LD for the plateau (Tibetan and Yunnan–Kweichow) sheep breeds when compared to breeds from Northern and Eastern China the latter exhibiting a rapid decay rate and low LD levels. In buffaloes, different patterns for LD decay for swamp and their river counterparts are reported, with higher rates for LD decay in the swamp when compared to the river group [45]. In farmed Atlantic salmon, different LD patterns between populations adapted to different environments across hemispheres have been observed [46] with authors

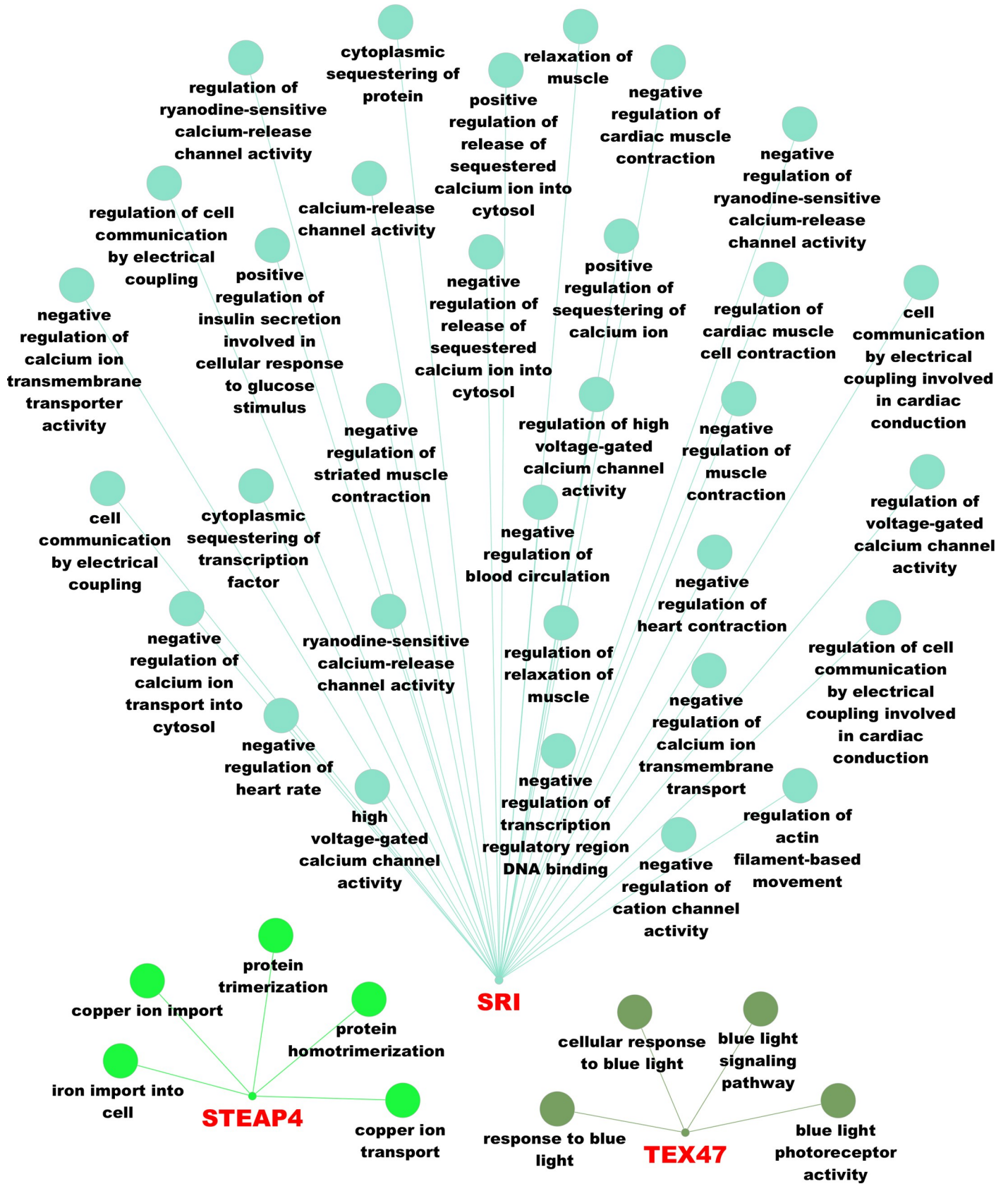


Fig 6. Three network representations of the total number of significantly enriched GO BP terms (nodes with khaki, light green and light blue color) and their respective member genes (red labels). Figure was constructed via ClueGO V2.5.7 and CluePedia V1.5.7 plugins in Cytoscape.

<https://doi.org/10.1371/journal.pone.0257461.g006>

emphasizing that exposure to different management and environmental conditions over just a few generations be sufficient to generate large changes in the genetic structure of farmed salmon populations with the same genetic origin. In line with results for salmon, populations of common carp originating from different geographic regions from China showed distinct LD decay patterns with distance [47]. Since different LD patterns can emerge for a variety of reasons, including mating system, sample size, population background, and demography [48], it's impossible to pinpoint and quantify each factor's exact contribution.

Geography or environmental related adaptation?

The genetic patterns discovered here corroborate the genetic similarity of Lemnos and Lesvos sheep, probably due to past gene flow [3]. Furthermore, they uncover the genetic uniqueness of the Skyros and the Katsika sheep from Epirus. Finally, they suggest a genetic relatedness between the Oreino (in Greek meaning mountainous) Epirus sheep with the mountainous Peloponnesus sheep. The latter suggests either a common ancestor for the two populations or past gene flow between Epirus and Peloponnesus mountainous sheep following possibly human population migration routes.

Intuitively, the genetic variation of the studied populations herein is interpreted in terms of environmental adaptation. Yet, as the studied populations are placed on an east-west geographic axis, genetic variation may equally fit a geographical axis instead of an environmental axis. In line with such a scenario, results of population structure and SPPA are supportive of an east-west geographical gradient with Skyros sheep positioned in the middle of the two geographical extremes (Lemnos—Lesvos and Epirus). As geography overlaps with climatic zones in the way that (east) insular sheep are kept on semi arid areas and (western) mainland sheep in rainy, humid and high altitude areas, it is impossible to identify the exact source of genetic differentiation. In general, regional origin has been identified as the main determinant of genetic differentiation of numerous European sheep breeds with three Greek breeds (Chios, Kymi and Lesvos) placed intermediate on an east–west cline between the fat-tailed Asian and European sheep [49]. Since the animals of the present study have been kept at their sampling sites for decades, we assess that genetic differentiation is more likely reflecting adaptation processes to local environmental conditions than possible regional differences.

Evidence for long term local adaptation

Current results suggested that mainland sheep differed genetically from island sheep, with genetic differentiation reaching its peak at marker *s25289.1* on OAR4. Genomic regions with such outlier markers are intuitively interpreted as evidence of LA. Nonetheless, apart from LA, extreme F_{ST} outliers can also be generated by global adaptation and genetic drift [6]. Since different processes leave different footprints on the genomic landscape, examining the pattern of summary statistics, such as F_{ST} , genetic variation within and between populations in genomic regions surrounding F_{ST} outliers, may aid in determining which type of selective processes has occurred ([6,50]). Following this rationale, the peak of dH_o observed in Fig 5 conforms to long-term LA arising from spatially antagonistic pleiotropy ([6,51], Fig 3, case C). The importance of *s25289.1* was also confirmed by association analysis, raising the possibility that the surrounding genomic region contains one or more genes involved in long-term LA in sheep. We will discuss this issue in more detail in the next section.

Functional candidate genes

Three of the candidate genes were associated with significantly enriched GO terms. Of these, *TEX47* is associated with responses to blue light, including photoreceptor activity. In animals, light in the range from 400 to 500 nm is sensed via blue and ultraviolet photoreceptors that are called cryptochromes. Cryptochromes act as components of the circadian clock that control daily physiological and behavioral rhythms and as photoreceptors that mediate entrainment of the circadian clock to light [52].

The next candidate gene i.e. *SRI* was found to be involved in multiple BPs including positive regulation of insulin secretion involved in cellular response to glucose stimulus, excitation-contraction coupling in the heart, calcium homeostasis in the heart sarcoplasmic reticulum and modulation of the activity of RYR2 calcium channels [53].

With regard to *STEAP4*, associated GO BP terms revealed its involvement in iron and copper import/transport that is critical to the maintenance of cellular homeostasis [54]. In humans, the gene has been linked to obesity [55–57], while in sheep, *STEAP4* [58] and other members of the same gene family (*STEAP3*) have been associated with excess fat accretion in tails [59]. The latter is a valuable energy reserve, playing a particular role in adaptation to harsh conditions (e.g. drought seasons, extreme cold winters and food shortages) allowing fuel reserves to be mobilized when underfed and restored during the more favorable season [60]. Apart from the regulatory circuits of fat storage and mobilization in white adipocytes, the adipose tissue provides a unique thermogenic mechanism (via the brown and beige adipocytes) to maintain euthermy in response to cold exposure ([61,62]) that is critical for the survival of the newborns [63]. Finally, the adipose tissue acts as an endocrine organ and produces numerous bioactive factors such as adipokines and lipokines that communicate with other organs and modulate a range of metabolic pathways [64]. All of the above points to the adipose tissue's role in the adaptive regulation of systemic metabolic homeostasis in response to nutrient excess or fasting, cold exposure [62], or physiological status.

With regard to rest genes with no detected enriched GO terms, *ADAM22* is a member of the ADAM (A Disintegrin And Metalloprotease) gene family that are involved in a broad range of biological processes such as cell fusion, spermatogenesis and development [65] and in the regulation of cell adhesion and inhibition of cell proliferation [66]. The gene has been associated with rumen microbial composition in sheep [67] and with reproductive traits in Nelore cattle [68].

Finally, *ZNF804B* is found to be located in genomic regions with overlapping SS and Copy Number Variations that were associated with selective adaptation in cattle [69].

Conclusions

In conclusion, here we report a clear genetic differentiation of mainland vs. island Greek indigenous breeds that can be explained either by longitudinal or environmental factors or both. Furthermore, a genomic region (OAR4: 72.4–73.3Mb) associated with LA was identified. Consistently with previous findings [10], this region is likely to be important for local adaptation in sheep and harbors five functionally relevant genes, *TEX47*, *STEAP4*, *SRI*, *ADAM22* and *ZNF804B*, that warrant further investigation.

Supporting information

S1 Fig. Map of sampling locations of sheep herds in Greece.
(TIF)

S2 Fig. Quantile-quantile (Q-Q) plot of GWAS results for the two origins (island vs. mainland). Blue dots denote the $-\log_{10}(p\text{-value})$ obtained from the single locus mixed model and the red line represents the expected values for the null hypothesis under no association. (TIF)

S1 Table. Distribution of samples per origin and region within origin (n: Number of samples per region). (DOCX)

S2 Table. Results by Gene Ontology biological process (GO BP) enrichment analysis. (DOCX)

Acknowledgments

We would like to express our heartfelt gratitude to all contributing farmers for granting us access to their genetic material.

Author Contributions

Conceptualization: Antonios Kominakis, Ioannis Hadjigeorgiou.

Formal analysis: Antonios Kominakis, Eirini Tarsani.

Funding acquisition: Ioannis Hadjigeorgiou.

Investigation: Antonios Kominakis, Ariadne L. Hager-Theodorides, Ioannis Mastranestasis, Dimitra Gkelia, Ioannis Hadjigeorgiou.

Methodology: Antonios Kominakis, Eirini Tarsani.

Project administration: Antonios Kominakis, Ioannis Hadjigeorgiou.

Software: Eirini Tarsani.

Supervision: Antonios Kominakis, Ioannis Hadjigeorgiou.

Visualization: Antonios Kominakis, Eirini Tarsani.

Writing – original draft: Antonios Kominakis, Eirini Tarsani, Ariadne L. Hager-Theodorides, Ioannis Hadjigeorgiou.

References

1. Kominakis A, Rogdaki E. Rare indigenous sheep breeds of Greece. Workshop of Save Foundation: Rare livestock breeds in Greece, Athens. 2005. Available: http://www.agrobiodiversity.net/greece/pdf/Rare_indigenous_sheepbreeds.pdf.
2. Mastranestasis I, Ekateriniadou L V., Ligda C, Theodorou K. Genetic diversity and structure of the Lesvos sheep breed. *Small Rumin Res.* 2015; 130: 54–59. <https://doi.org/10.1016/j.smallrumres.2015.07.015>
3. Kominakis A, Tarsani E, Hager-Theodorides AL, Mastranestasis I, Hadjigeorgiou I. Clustering patterns mirror the geographical distribution and genetic history of Lemnos and Lesvos sheep populations. Chiang T-Y, editor. *PLoS One.* 2021; 16: e0247787. <https://doi.org/10.1371/journal.pone.0247787> PMID: 33657161
4. Abied A, Bagadi A, Bordbar F, Pu Y, Augustino SMA, Xue X, et al. Genomic Diversity, Population Structure, and Signature of Selection in Five Chinese Native Sheep Breeds Adapted to Extreme Environments. *Genes (Basel).* 2020; 11: 494. <https://doi.org/10.3390/genes11050494> PMID: 32365888
5. Librado P, Orlando L. Detecting Signatures of Positive Selection along Defined Branches of a Population Tree Using LSD. Hernandez R, editor. *Mol Biol Evol.* 2018; 35: 1520–1535. <https://doi.org/10.1093/molbev/msy053> PMID: 29617830

6. Booker TR, Yeaman S, Whitlock MC. Global adaptation complicates the interpretation of genome scans for local adaptation. *Evol Lett*. 2021; 5: 4–15. <https://doi.org/10.1002/evl3.208> PMID: 33552532
7. Pérez O'Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassell CP, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol*. 2014; 46: 19. <https://doi.org/10.1186/1297-9686-46-19> PMID: 24592996
8. JOOST S, BONIN A, BRUFORD MW, DESPRÉS L, CONORD C, ERHARDT G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol Ecol*. 2007; 16: 3955–3969. <https://doi.org/10.1111/j.1365-294X.2007.03442.x> PMID: 17850556
9. Fariello M-I, Servin B, Tosser-Klopp G, Rupp R, Moreno C, Cristobal MS, et al. Selection Signatures in Worldwide Sheep Populations. Caramelli D, editor. *PLoS One*. 2014; 9: e103813. <https://doi.org/10.1371/journal.pone.0103813> PMID: 25126940
10. Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity (Edinb)*. 2016; 116: 255–264. <https://doi.org/10.1038/hdy.2015.94> PMID: 26555032
11. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. Barendse W, editor. *PLoS One*. 2017; 12: e0176780. <https://doi.org/10.1371/journal.pone.0176780> PMID: 28463982
12. de Simoni Gouveia JJ, Paiva SR, McManus CM, Caetano AR, Kijas JW, Facó O, et al. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livest Sci*. 2017; 197: 36–45. <https://doi.org/10.1016/j.livsci.2017.01.006>
13. Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P. Application of Selection Mapping to Identify Genomic Regions Associated with Dairy Production in Sheep. Kaltenboeck B, editor. *PLoS One*. 2014; 9: e94623. <https://doi.org/10.1371/journal.pone.0094623> PMID: 24788864
14. Megdiche S, Mastrangelo S, Hamouda M Ben, Lenstra JA, Ciani E. A combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in merino and merino-derived sheep breeds. *Front Genet*. 2019; 10: 1025. <https://doi.org/10.3389/fgene.2019.01025> PMID: 31708969
15. Signer-Hasler H, Burren A, Ammann P, Drögemüller C, Flury C. Runs of homozygosity and signatures of selection: a comparison among eight local Swiss sheep breeds. *Anim Genet*. 2019; 50: 512–525. <https://doi.org/10.1111/age.12828> PMID: 31365135
16. Hu X-J, Yang J, Xie X-L, Lv F-H, Cao Y-H, Li W-R, et al. The Genome Landscape of Tibetan Sheep Reveals Adaptive Introgression from Argali and the History of Early Human Settlements on the Qinghai-Tibetan Plateau. Nielsen R, editor. *Mol Biol Evol*. 2019; 36: 283–303. <https://doi.org/10.1093/molbev/msy208> PMID: 30445533
17. Mastrangelo S, Bahbahani H, Moioli B, Ahbara A, Al Abri M, Almathen F, et al. Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. Mateescu R, editor. *PLoS One*. 2019; 14: e0209632. <https://doi.org/10.1371/journal.pone.0209632> PMID: 31199810
18. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. Tyler-Smith C, editor. *PLoS Biol*. 2012; 10: e1001258. <https://doi.org/10.1371/journal.pbio.1001258> PMID: 22346734
19. Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, et al. Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol*. 2014; 31: 3324–3343. <https://doi.org/10.1093/molbev/msu264> PMID: 25249477
20. Liu Z, Ji Z, Wang G, Chao T, Hou L, Wang J. Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics*. 2016; 17. <https://doi.org/10.1186/s12864-016-3212-2> PMID: 27809776
21. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010; 5: 1564–1573. <https://doi.org/10.1038/nprot.2010.116> PMID: 21085122
22. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*. 2013; 193: 929–941. <https://doi.org/10.1534/genetics.112.147231> PMID: 23307896
23. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. Oxford Academic; 2006. pp. 254–267. <https://doi.org/10.1093/molbev/msj030> PMID: 16221896
24. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; 19: 1655–1664. <https://doi.org/10.1101/gr.094052.109> PMID: 19648217

25. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour.* 2015; 15: 1179–1191. <https://doi.org/10.1111/1755-0998.12387> PMID: 25684545
26. Moran PAP. Notes on Continuous Stochastic Phenomena. *Biometrika.* 1950; 37: 17. <https://doi.org/10.2307/2332142> PMID: 15420245
27. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet.* 1968; 38: 226–231. <https://doi.org/10.1007/BF01245622> PMID: 24442307
28. Wright S. Coefficients of Inbreeding and Relationship. *Am Nat.* 1922; 56: 330–338. <https://doi.org/10.1086/279872>
29. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N Y).* 1984; 38: 1358. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791
30. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, et al. Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics.* 2010; 186: 241–262. <https://doi.org/10.1534/genetics.104.117275> PMID: 20855576
31. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics.* 1973; 74: 175–195. <https://doi.org/10.1093/genetics/74.1.175> PMID: 4711903
32. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics.* 2008; 180: 977–993. <https://doi.org/10.1534/genetics.108.092221> PMID: 18780740
33. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet.* 2010; 42: 348–354. <https://doi.org/10.1038/ng.548> PMID: 20208533
34. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006; 38: 203–208. <https://doi.org/10.1038/ng1702> PMID: 16380716
35. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet.* 2011; 19: 807–12. <https://doi.org/10.1038/ejhg.2011.39> PMID: 21407268
36. NCBI. NCBI Ovis aries Annotation Release 102. [cited 15 Mar 2021]. Available: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/102/.
37. NCBI. Oar_v4.0—oviAri4—Genome—Assembly—NCBI. [cited 15 Mar 2021]. Available: https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2/.
38. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016; 17: 122. <https://doi.org/10.1186/s13059-016-0974-4> PMID: 27268795
39. NCBI. Coordinate remapping service: NCBI. [cited 15 Mar 2021]. Available: <https://www.ncbi.nlm.nih.gov/genome/tools/remap>.
40. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009; 25: 1091–3. <https://doi.org/10.1093/bioinformatics/btp101> PMID: 19237447
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003; 13: 2498–2504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658
42. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics.* 2013; 29: 661–663. <https://doi.org/10.1093/bioinformatics/btt019> PMID: 23325622
43. Qanbari S. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. *Frontiers in Genetics.* Frontiers Media S.A.; 2020. p. 1304. <https://doi.org/10.3389/fgene.2019.01304> PMID: 32010183
44. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol Biol Evol.* 2016; 33: 2576–2592. <https://doi.org/10.1093/molbev/msw129> PMID: 27401233
45. Lu XR, Duan AQ, Liang SS, Ma XY, Liang XW, Deng TX. Genome-wide analysis reveals genetic diversity, linkage disequilibrium, and selection for milk traits in Chinese buffalo breeds. *bioRxiv.* bioRxiv; 2019. p. 701045. <https://doi.org/10.1101/701045>
46. López ME, Linderoth T, Norris A, Lhorente JP, Neira R, Yáñez JM. Multiple Selection Signatures in Farmed Atlantic Salmon Adapted to Different Environments Across Hemispheres. *Front Genet.* 2019; 10: 901. <https://doi.org/10.3389/fgene.2019.00901> PMID: 31632437

47. Xu J, Jiang Y, Zhao Z, Zhang H, Peng W, Feng J, et al. Patterns of Geographical and Potential Adaptive Divergence in the Genome of the Common Carp (*Cyprinus carpio*). *Front Genet.* 2019; 10: 660. <https://doi.org/10.3389/fgene.2019.00660> PMID: 31354795
48. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet.* 2010; 41: 346–356. <https://doi.org/10.1111/j.1365-2052.2009.02011.x> PMID: 20055813
49. Ciani E, Mastrangelo S, Da Silva A, Marroni F, Ferenčaković M, Ajmone-Marsan P, et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genet Sel Evol.* 2020; 52: 25. <https://doi.org/10.1186/s12711-020-00545-7> PMID: 32408891
50. Bierne N. THE DISTINCTIVE FOOTPRINTS OF LOCAL HITCHHIKING IN A VARIED ENVIRONMENT AND GLOBAL HITCHHIKING IN A SUBDIVIDED POPULATION. *Evolution (N Y).* 2010; 64: 3254–3272. <https://doi.org/10.1111/j.1558-5646.2010.01050.x>
51. Sakamoto T, Innan H. The evolutionary dynamics of a genetic barrier to gene flow: From the establishment to the emergence of a peak of divergence. *Genetics.* 2019; 212: 1383–1398. <https://doi.org/10.1534/genetics.119.302311> PMID: 31171654
52. Cashmore AR. Cryptochromes: Enabling plants and animals to determine circadian time. *Cell.* Cell Press; 2003. pp. 537–543. <https://doi.org/10.1016/j.cell.2003.08.004>
53. Franceschini S, Ilari A, Verzili D, Zamparelli C, Antaramian A, Rueda A, et al. Molecular basis for the impaired function of the natural F112L sorcin mutant: X-ray crystal structure, calcium affinity, and interaction with annexin VII and the ryanodine receptor. *FASEB J.* 2008; 22: 295–306. <https://doi.org/10.1096/fj.07-8988com> PMID: 17699613
54. Scarl RT, Lawrence CM, Gordon HM, Nunemaker CS. STEAP4: Its emerging role in metabolism and homeostasis of cellular iron and copper. *Journal of Endocrinology.* BioScientifica Ltd.; 2017. pp. R123–R134. <https://doi.org/10.1530/JOE-16-0594>
55. Zhang CM, Chi X, Wang B, Zhang M, Ni YH, Chen RH, et al. Downregulation of STEAP4, a highly-expressed TNF- α -inducible gene in adipose tissue, is associated with obesity in humans. *Acta Pharmacol Sin.* 2008; 29: 587–592. <https://doi.org/10.1111/j.1745-7254.2008.00793.x> PMID: 18430367
56. Elbein SC, Kern PA, Rasouli N, Yao-Borengasser A, Sharma NK, Das SK. Global gene expression profiles of subcutaneous adipose and muscle from glucose-tolerant, insulin-sensitive, and insulin-resistant individuals matched for BMI. *Diabetes.* 2011; 60: 1019–1029. <https://doi.org/10.2337/db10-1270> PMID: 21266331
57. Xu HM, Cui YZ, Wang WG, Cheng HX, Sun YJ, Zhao HY, et al. Expression and clinical significance of obesity-associated gene STEAP4 in obese children. *Genet Mol Res.* 2016; 15. <https://doi.org/10.4238/gmr.15048705> PMID: 27808366
58. Xu SS, Ren X, Yang GL, Xie XL, Zhao YX, Zhang M, et al. Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (*Ovis aries*). *Anim Genet.* 2017; 48: 560–569. <https://doi.org/10.1111/age.12572> PMID: 28677334
59. Bakhtiarzadeh MR, Alamouti AA. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. *Sci Rep.* 2020; 10: 13525. <https://doi.org/10.1038/s41598-020-70527-8> PMID: 32782325
60. Chilliard Y, Ferlay A, Faulconnier Y, Bonnet M, Rouel J, Bocquier F. Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *Proceedings of the Nutrition Society.* CAB International; 2000. pp. 127–134. <https://doi.org/10.1017/S002966510000015X>
61. Harms M, Seale P. Brown and beige fat: Development, function and therapeutic potential. *Nature Medicine.* Nat Med; 2013. pp. 1252–1263. <https://doi.org/10.1038/nm.3361> PMID: 24100998
62. Chouchani ET, Kajimura S. Metabolic adaptation and maladaptation in adipose tissue. *Nature Metabolism.* Nature Research; 2019. pp. 189–200. <https://doi.org/10.1038/s42255-018-0021-8> PMID: 31903450
63. Ojha S, Robinson L, Yazdani M, Symonds ME, Budge H. Brown adipose tissue genes in pericardial adipose tissue of newborn sheep are downregulated by maternal nutrient restriction in late gestation. *Pediatr Res.* 2013; 74: 246–251. <https://doi.org/10.1038/pr.2013.107> PMID: 23788058
64. Luo L, Liu M. Adipose tissue in control of metabolism. *Journal of Endocrinology.* BioScientifica Ltd.; 2016. pp. R77–R99. <https://doi.org/10.1530/JOE-16-0211>
65. Souza JSM, Lisboa ABP, Santos TM, Andrade MVS, Neves VBS, Teles-Souza J, et al. The evolution of ADAM gene family in eukaryotes. *Genomics.* 2020; 112: 3108–3116. <https://doi.org/10.1016/j.ygeno.2020.05.010> PMID: 32437852

66. Liu H, Shim AHR, He X. Structural characterization of the ectodomain of a disintegrin and metalloproteinase-22 (ADAM22), a neural adhesion receptor instead of metalloproteinase. Insights on adam function. *J Biol Chem.* 2009; 284: 29077–29085. <https://doi.org/10.1074/jbc.M109.014258> PMID: 19692335
67. S M, OA A, MA A. Association between Host Genetics of Sheep and the Rumen Microbial Composition. 2020 [cited 15 Mar 2021]. <https://doi.org/10.21203/RS.3.RS-111081/V1>
68. Costa RB, Camargo GMF, Diaz IDPS, Irano N, Dias MM, Carvalheiro R, et al. Genome-wide association study of reproductive traits in Nellore heifers using Bayesian inference. *Genet Sel Evol.* 2015; 47: 67. <https://doi.org/10.1186/s12711-015-0146-0> PMID: 26286463
69. Ana M. Perez O'Brien, Yuri T. Utsunomiya, Gabor Meszaros, Derek M. Bickhart, George E. Liu, Jose F. Garcia, et al. ISAFG 2013 5th International Symposium on Animal Functional Genomics PROGRAMME AND ABSTRACT BOOK p1 | 5 th Symposium on Animal Functional Genomics-ISAFG 2013. Are Copy Number Variants causal changes for adaptive selection? On the overlap of CNV and VarLD Selection Signatures in cattle. 2013. p. 63.