scientific reports

Check for updates

OPEN

Assessment of breeding nuclei contributions to the genetic diversity and population structure of the Cyprus Chios sheep

Andreas C. Dimitriou[1,2✉], Giannis Maimaris[1] & Georgia Hadjipavlou[1]

Cyprus, facing climate change and desertification, has Europe's second highest population growth. Halloumi cheese's Protected Designation of Origin status boosted goat and sheep milk demand. Optimizing primary sector production is key to balancing food production and environmental sustainability. The present study aims to explore the genetic diversity and population structure of Cyprus Chios sheep, as well as the impact of existing breeding nuclei on the breed's gene pool. Two breeding nuclei aiming to preserve genetic diversity and boost local milk production and four private farms were sampled. Genome-wide data using Illumina 50 K arrays were generated, analyzing over 1000 animals from six farms. At least three distinct genetic groups were identified, two of which are represented by the nuclei's distinct genetic profiles and a third was found in private farms. Calculated metrics indicate a negative correlation of gene flow with geographic distance in case of the Eastern nucleus. Presented results support recent, though limited inbreeding (mean F_{ROH}=0.046), compared to other commercial breeds. Estimated effective population size reveals a stabilizing trend to a relatively high number (Ne ∼ 200) which indicates relatively high diversity that allows further genetic improvement. Revealed genomic diversity highlights the necessity of such exploration before effectively implementing genomic evaluation programs.

Located in the eastern Mediterranean, close to the putative domestication center, the Central/Eastern Anatolia[1,2] Cyprus was one of the first regions where sheep were introduced, as evidenced by zooarcheological records dating the presence of the species on the island back to ∼ 10,000 BP[2–4]. Available genetic and archaeological data suggest that the first domesticated populations originated from adjacent eastern Mediterranean regions (Israel, Lebanon, Syria, and southeast Turkey), while part of this early population was later feralized to the endemic Cypriot mouflon[4–6]. The constant presence of sheep on the island over the last 10 millennia highlights the importance of the species to the local communities for dairy and meat products.

Nowadays, the circum-Mediterranean and Black Sea regions host the largest fraction of the global dairy sheep population, a fact that is also reflected in the typical human diet in these regions[7]. Although dairy small ruminants currently produce only a small fraction of the world's milk (∼ 3,5%), there has been a constant increasing trend that has led to over doubling the total milk production during the last 50 years[8]. This increasing trend is even more steep in the case of Cyprus, where the total sheep milk production has doubled in the last 10 years[8], a fact that is mainly attributed to the registration of the locally produced Halloumi cheese as a Protected Designation of Origin (PDO) product.

The Chios sheep breed has attracted scientific community interest since the 1980s, when the first studies focused on the genetic correlations and heritability of specific dairy traits, such as milk yield and udder characteristics, utilizing pedigree and phenotypic data from the Cypriot breed[9,10] were conducted. Soon after, studies including genetic data of Chios and other sheep breeds suggested that specific important traits could be improved through selection[11–13], indicating that the simultaneous selection of two traits could be challenging if they are genetically linked (i.e. milk yield and composition)[14,15]. Subsequent studies revealed the significance of specific loci and polymorphisms on milk production traits[16], while more recent work suggested candidate gene sets correlated with the same factors[17–19]. It is worth noting that although at the beginning of the millennium it

[1]Agricultural Research Institute, P.O. Box 22016, Lefkosia 1516, Cyprus. [2]Department of Biological Sciences, University of Cyprus, University Campus, 2109 Nicosia, Cyprus. ✉email: andreas.c.dimitriou@gmail.com

was predicted that only 50–100 genes could affect dairy traits[20], this number has increased a hundredfold based on the advancement of data generation methods combined with GWAS analyses[21]. Recently, a study focusing on Greece and France, two of the biggest sheep milk producers in Europe, suggested that the majority of the examined farms in both countries were profitless, with the exception of the intensive Chios farms in Greece[7]. According to recent data, Greece, which is the country of origin of the breed, maintains the majority of Chios sheep individuals in intensive or semi-intensive farming conditions in Northern Greece[22].

Cyprus Chios sheep is the dominant pure breed on the island in terms of the maintained number of individuals and is genetically distinct from the indigenous Cyprus fat tail sheep[23]. Currently, ~20% of the total population belongs to the Cyprus Chios breed, while ~76% are crossbreds. These crossbreds resulted from historical unstructured crossbreeding, between Cyprus Chios and Cyprus fat-tailed, along with Awassi and East Friesian individuals that were in the past imported on the island, but had adaptation issues as purebreds, according to available governmental records[24] or with the more recently introduced exotic breeds. The constant local human population growth along with the increasing reputation of halloumi cheese abroad and subsequent increasing milk demands have led to repetitive imports of a few thousand individuals belonging to exotic commercial genetically improved sheep breeds, such as the Assaf and the Lacaune, aiming to boost the local population's productivity.

The first Chios sheep breed individuals were imported from the Greek island of Chios during the 1950s, while the great majority of the extant population are the ancestors of a second, much more successful introduction of the same origin that occurred in the 1970s. The introduction of the specific breed was favored, among other reasons, due to its great prolificacy, high milk output, and adaptive capacity to Cyprus' harsh environmental conditions. Additional advantages of this breed include early maturity at 8–9 months and extended seasonal breeding, which usually occurs between September and May, while their lactation period lasts six to seven months[25]. Considering all the aforementioned advantages, the Chios Sheep is rendered as the most appropriate breed for semi-intensive sheep production systems in Cyprus.

The Cyprus government maintains two experimental Cyprus Chios breeding nuclei in the central and western parts of the island (Athalassa and Acheleia areas, respectively). These farms were established ~60 years ago and currently enumerate ~400 individuals each. The purpose of these public farms, beyond conserving the genetic diversity of the breed on the island, is also the implementation of a genetic improvement program aiming to increase total milk yield, preserving at the same time the high nutritional value of the produced milk (fat and protein content). These nuclei successively faced challenges in the past, with Scrapie disease prevention being the most critical one, actively selecting resistant individuals and transferring at the same time knowledge and good practices to local farmers. Overall, one of the most important roles of these nuclei is the enrichment of the local population of the Cyprus Chios sheep breed with improved male and female animals carrying desired alleles correlated with milk quality and quantity traits. It is worth noticing that there has been no introduction of new individuals in either nucleus since their foundation, apart from very scarce and limited transportation of individuals from the eastern to the western nucleus during the national scrapie disease crisis. Since the breeding nuclei were established using founder individuals based on the typical phenotypic characteristics of the breed and maintain a small number of breeding animals, we hypothesize that they are not genetically identical to each other and probably do not adequately represent the entire genetic diversity of the local breed. Taking into account the long isolation of both breeding nuclei, the relatively small flock size, and the unidirectional gene flow from nuclei to private farms, we assume that the present patterns of genetic diversity could also be influenced by secondary diversification after their foundation, including random genetic drift and inbreeding.

Regarding the four private farms included in this study, these are members of the Cyprus Chios sheep breed society and therefore maintain purebred Cyprus Chios sheep. All farms, except OUF, were established over 30 years ago independently from the breeding nuclei. The OUF farm was set up about 15 years ago mainly with male genetic stock from the ENU nucleus and therefore could be characterized as a pseudo-nucleus. All four private farms periodically bought genetically improved Cyprus Chios lambs, mainly males, from the nucleus flocks for breeding. We found no evidence of animals from other farms being introduced, although these farms do exchange animals with breeder society members. All adult breeding males and selected productive, unrelated females, based on farms records, were sampled. Flock sizes varied, with TEP having 1000 sheep, ANT 700, OUM 1500, and OUF 300. Among the private sampled farms, only one farm (TAN) follows a genetic improvement program aiming to increase milk yield while keeping inbreeding to a minimum.

Given the increasing demands for locally produced sheep milk and the lack of knowledge regarding the genetic diversity of the dominant sheep breed on the island, the present study aims to assess the genetic variation and explore the genetic structure of the Chios sheep breed within the island, producing, for the first time, genome-wide data. Additionally, the contribution of breeding nuclei to the genetic pool of private farms as well as the representation of total genetic divergence by the nuclei are investigated.

Generated knowledge is anticipated to inform and advance ongoing genetic improvement programs. Mating strategies for genetic improvement could be redesigned based on the genetic background, homogeneity and inbreeding levels of the local Chios sheep breed, considering at the same time the preservation of genetic diversity.

Results
Genetic diversity – gene flow
In total, 1059 individuals from six farms located throughout the island were genotyped (Fig. 1). After filtering our final dataset included 50,484 SNPs out of 59,727 generated. The average calculated genetic distance (p_distance) within group varied from 9.50 to 13.50% in case of OUF and TAN respectively. Regarding between group divergence, the minimum distance is found between ENU and the OUF pseudonucleus as expected given
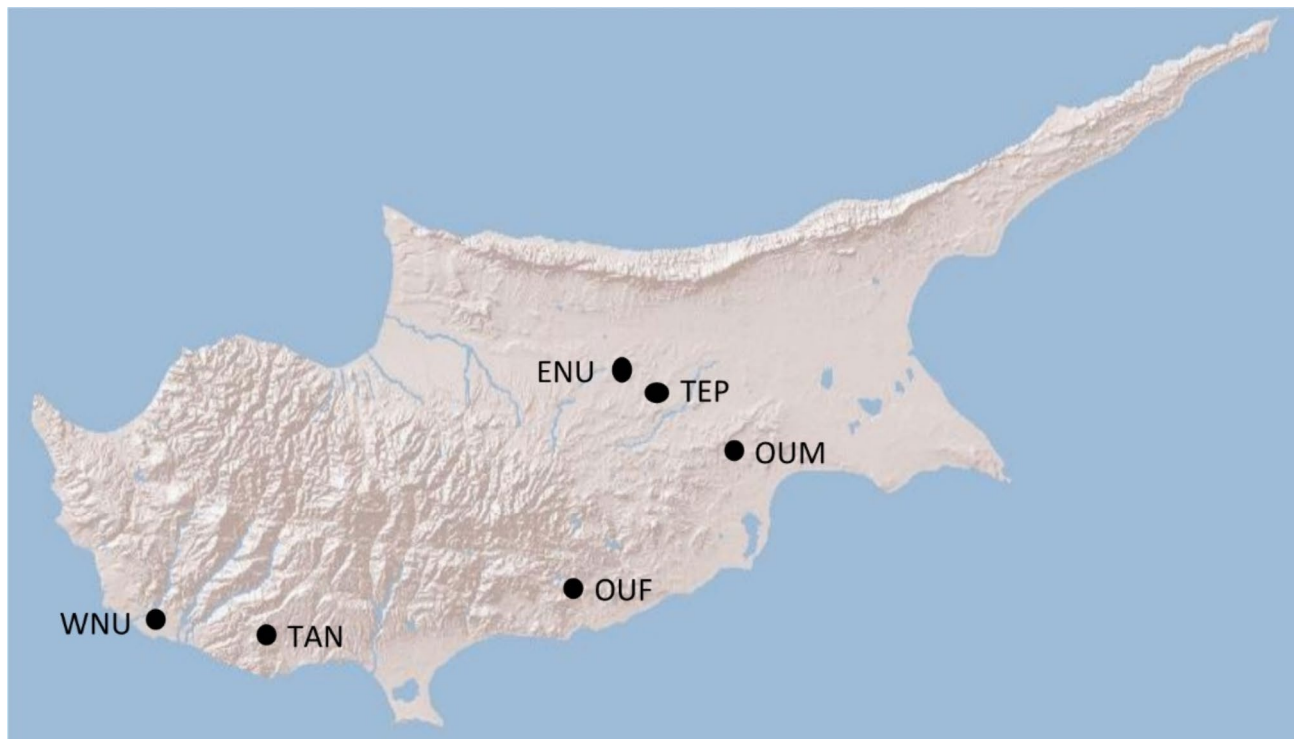
Fig. 1. Cyprus Chios sheep farms from where samples were collected.

Group	ENU	WNU	TEP	TAN	OUM	OUF
ENU	**11.20**					
WNU	13.45	**11.60**				
TEP	12.61	13.99	**11.50**			
TAN	13.74	14.23	14.33	**13.50**		
OUM	12.98	13.58	13.58	14.30	**13.00**	
OUF	11.48	13.40	12.62	13.75	12.85	**9.50**

Table 1. Percentage sequence divergence (p-distance) among farms. Values at the diagonal are within group mean distances.

the origin of individuals at the latter farm. On the other hand the divergence is highest between TAN-TEP groups reaching 14.33% (Table 1).

Genetic diversity explored through Discriminant Analysis of Principal Components (DAPC;[26]) assigned available genotypes to different groups based on the K-means method. According to the calculated Bayesian Information Criterion (BIC;[27]), the most probable solution, corresponding to the lowest value suggests the division of analysed individuals into six groups (K=6). The assignment of individuals to these groups by DAPC is largely in agreement with their farm of origin, with some exceptions. More specifically, TAN and OUM individuals form a single group, while ENU individuals are split into two groups. The OUF pseudo-nucleus appears to be closely related to both ENU groups. A significant number of individuals belonging either to ENU or OUF farms could not be assigned with confidence to one of these two groups based on calculated posterior probabilities indicating the close genetic relationship between each other. Furthermore, according to DAPC assignments, some individuals from OUM and OUF are more closely related to ENU than individuals from their predefined group. The same was also evident in the case of the WNU nucleus, where the assignment of some individuals is more likely to TAN/OUM than their group of origin (Fig. 2). In more detail, individuals from the TEP farm form a single group (group 1). Individuals from ENU and OUF form a closely related cluster of three groups, with ENU split in groups 2 and 6, and OUF in groups 5 and 2 (Fig. 2; yellow:2, black:5, red:6). Finally, TAN and OUM individuals form group 3, while the majority of WNU samples belong to group 4, with some exceptions where posterior probabilities indicate their assignment to group 3.

The generated Maximum likelihood (ML) tree constructed with OrientAgraph[28] in Treemix[29] shows the overall closer genetic relationship between the OUF pseudo-nucleus and the nucleus of origin ENU (Fig. 3). These two farms appear to be closely related with the geographically close TEP farm forming a monophyletic group. The other three farms are grouped together supporting a greater genetic affinity between WNU and TAN
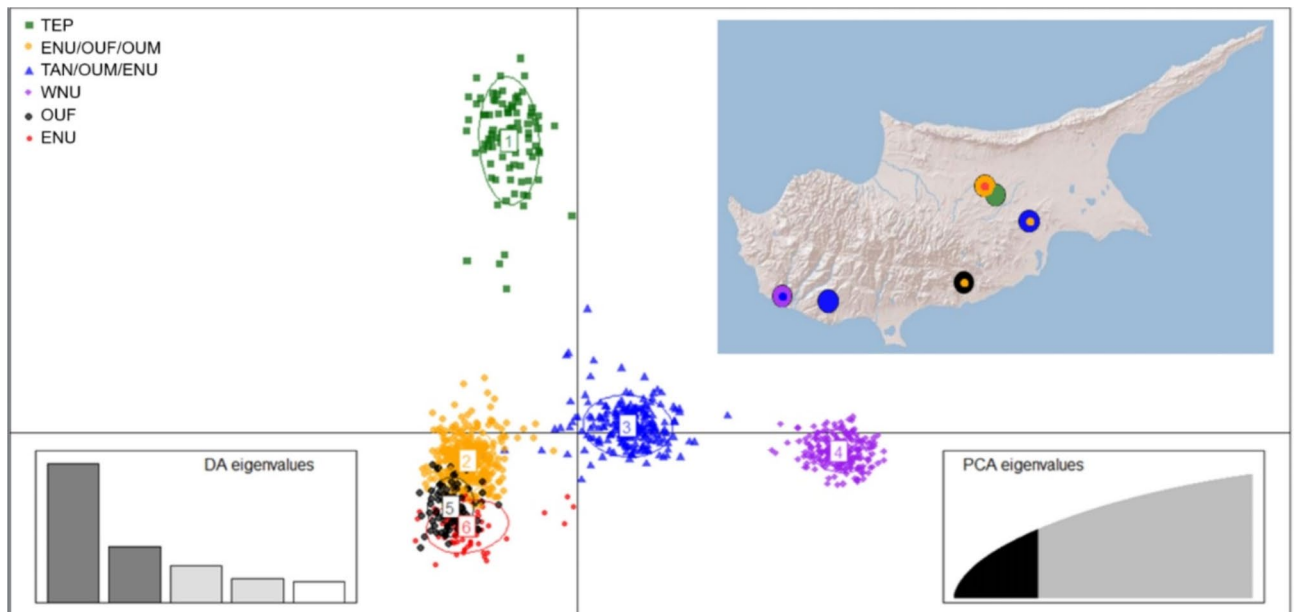
Fig. 2. DAPC scatterplot of all individuals included in this study. Colored circles on the map correspond to DAPC groups. Nested colored circles indicate the assignment of a limited number of individuals to the respective, according to color, alternative group.

farms. Extended gene flow from the ENU-OUF-TEP group towards TAN is highlighted, while migration events from the ENU-OUF group towards the WNU-TAN-OUM or OUM groups seem to have taken place in a lesser degree.

The calculated pairwise F_{ST} values varied from 0.017 to 0.051 between the OUM-ENU / OUM-TAN, and OUF-WNU groups, respectively (Table 2). These findings indicate strong gene flow between ENU-OUM-TAN farms, corroborating the pattern produced by DAPC and Treemix. It is also notable that the following lowest F_{ST} values are exhibited between TEP-ENU-OUM farms, which are geographically closer to each other than with the other farms. Finally, according to the fixation index, gene flow is more limited between the pseudo-nucleus OUF and the western breeding nucleus WNU as compared to the other groups (Table 2).

Population structure

The generated bar plots based on Q values estimated by STRUCTURE[30], highlight the genetic distinctiveness between the two breeding nuclei. The Evanno method[31] was applied in order to determine the number of clusters that best describes our dataset. The ΔK (Delta K) was calculated based on the rate of change in the log probability of data between successive number of clustering solutions. Plotting ΔK against different values of K (clusters) and identifying the peak, which indicates the optimal number of clusters, the best K for our dataset is 3 (Fig. 4). Nevertheless the solution for K=6 is presented for comparative reasons given the initial number of population groups. Based on the solution provided for K= 3 only a few individuals at the western nucleus (WNU) exhibit an admixed genetic profile, possessing nearly equal numbers of polymorphisms found in the two nuclei (Fig. 4). The third population group is represented by individuals from private farms, mainly TAN, exhibiting a unique genetic profile that in not found in either the eastern or the western breeding nucleus. Hence, breeding nuclei do not adequately represent the genetic diversity of the breed since unique genetic variation was found in private farms. However, with the exception of the OUF pseudo-nucleus, the great majority of sheep in private farms exhibit an admixed genetic profile, sharing SNPs from all three identified population groups. As expected, the genetic homogeneity between the "parental" eastern breeding nucleus and the pseudo-nucleus OUF is evident. The significant contribution of breeding nuclei to private farms' gene pool is supported by the considerable number of shared genetic polymorphisms between the individuals included in our analyses. Having a more detailed look at the presented bar plot for K=2 (Fig. 4), the greater genetic influence of the eastern nucleus ENU on the geographically adjacent private farms TEP and OUM rather than on TAN, located in the western part of the island, is underlined. Beyond the identified greater contribution of the ENU breeding nucleus to its geographically closer farms, no further correlation between geographic distance and genetic profile was observed among the farms included in our analyses. Nevertheless, genetic contribution of the eastern nucleus (ENU) is notably more prevalent in private farms compared to the western nucleus, underscoring the significant impact of ENU on the country's gene pool. Focusing on K=6, TEP seems to be more clearly separated from OUF and ENU groups while, in line with DAPC and F_{ST} results, the genetic influence of TAN to TEP and vice versa is becoming more obvious (Fig. 4).
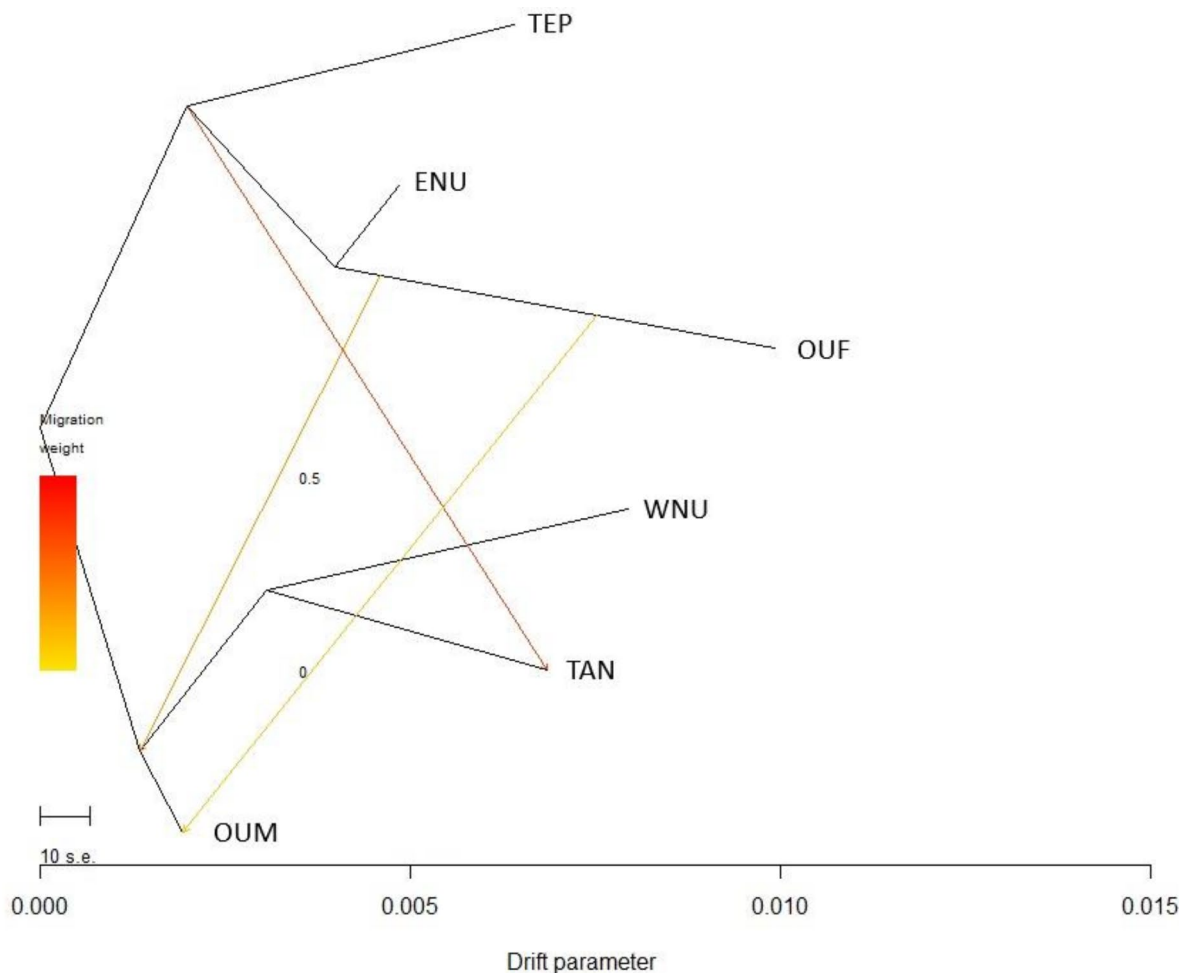
Fig. 3. Phylogenetic tree inferred by implementing Maximum Likelihood Network Orientation (OrientAgraph) within TreeMix software with migration events.

Group	ENU	WNU	TEP	TAN	OUM
WNU	0.035				
TEP	0.022	0.040			
TAN	0.026	0.029	0.029		
OUM	0.017	0.022	0.022	0.017	
OUF	0.022	0.051	0.039	0.041	0.030

Table 2. Estimated F_{ST} values among farms.

Runs of homozygosity

In total, 19,576 Runs of Homozygosity (ROH) segments were identified in our dataset, spanning from 137 to 2699 on chromosomes 24 and 2, respectively. Long fragments exceeding 5 Mb are the most abundant in all farms, supporting recent inbreeding (Fig. 5A, B). Average inbreeding based on the mean F_{ROH} for all examined Chios sheep was 0.046 (Fig. 5C). Focusing on each group, the estimated F_{ROH} was highest in the ENU breeding nucleus (0.053) and lowest for the TAN group (0.023). The same inbreeding coefficient is relatively high in TEP (0.051), WNU breeding nucleus (0.046), and OUF (0.045), indicating limited introduction of genetically distinct individuals into these groups, whereas F_{ROH} is rather low in the OUM (0.033) group. Additional data regarding the F_{ROH} size distribution across groups and individuals are given in Fig. 5C. With respect to the distribution of these runs across the genome, it seems that the lowest F_{ROH} mean per chromosome was detected in chromosome 1, while chromosomes 22 and 25 possess the highest frequencies of ROH (Fig. 6). The exploration of historical demographics using SNeP indicates that the effective population size of the Cyprus Chios sheep experienced
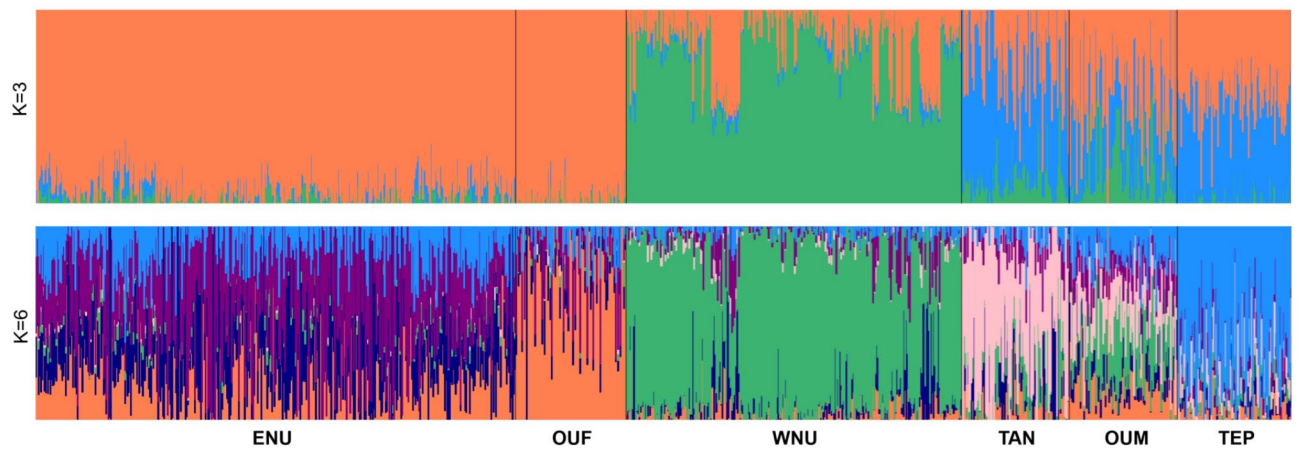
Fig. 4. Population structure inferred using STRUCTURE software based on genome-wide SNPs for K = 3 and K = 6. Vertical lines represent different individuals, while different colors correspond to the assignment probabilities of individuals to putative population clusters according to estimated Q values.
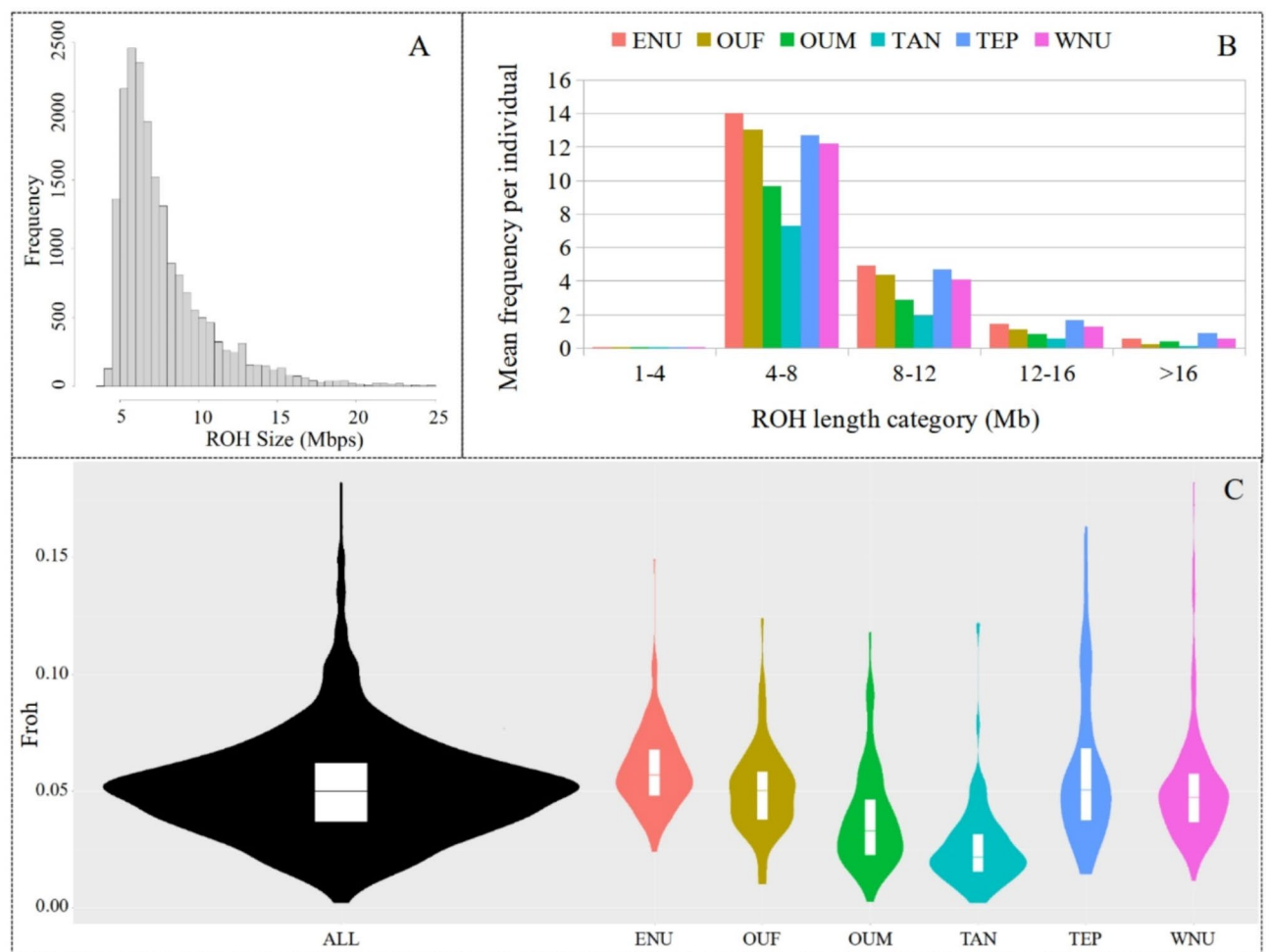

Fig. 5. (**A**) Distribution of ROH fragment sizes frequency for all individuals included in our analysis, (**B**) Mean sum of ROH per animal within each ROH length category per farm, (**C**) Kernel density (violin) plots showing the distribution of ROH lengths in total and each group. From left to right, with black color the F_{ROH} for all specimens included in our analysis followed by each farm's results.
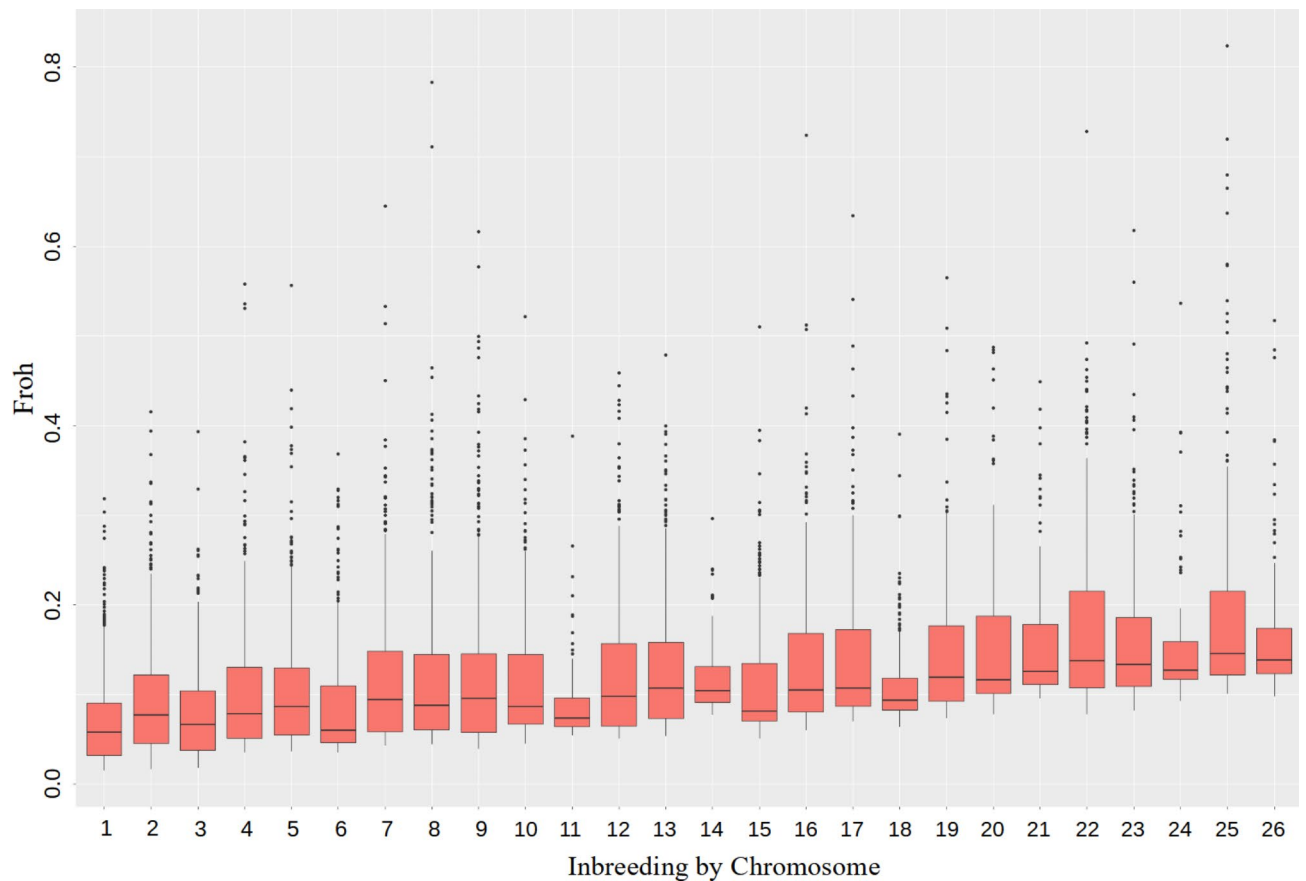
Fig. 6. Boxplot of the inbreeding coefficient (F_{ROH}) estimated from runs of homozygosity for each chromosome across all sampled individuals.
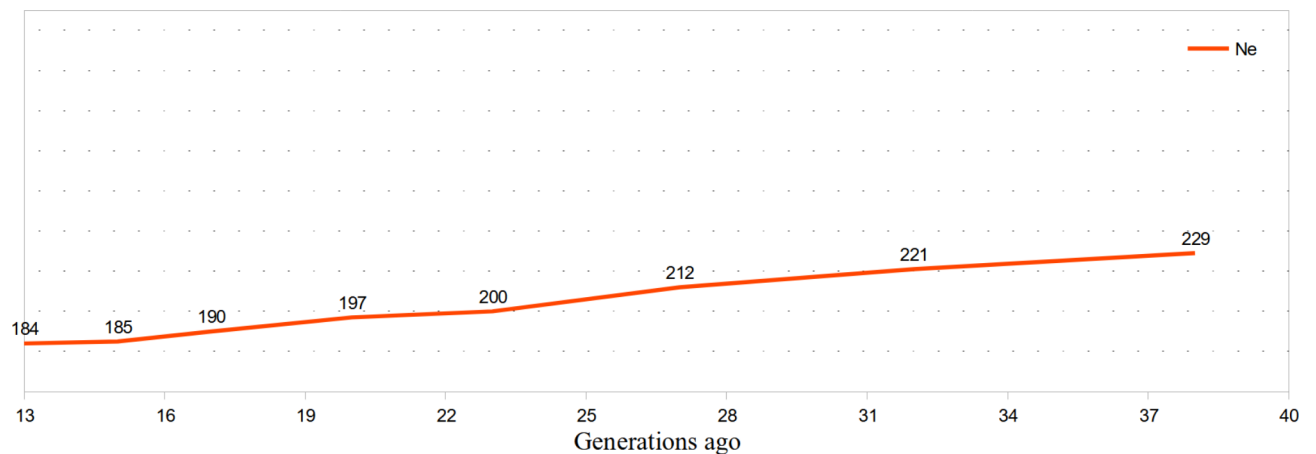

Fig. 7. Estimated effective population size (Ne) trajectory for Cyprus Chios Sheep inferred using SNeP.

a slow gradual decline after the introduction of the breed on the island in the 1950s (Fig. 7). However, this trajectory seems to be smoothened towards stabilization in recent years for which estimations are available.

Discussion

This is the first study to produce and analyze genome-scale genetic data for the Cyprus Chios sheep breed. Despite the narrow geographic limits of Cyprus, notable genetic divergence was revealed, identifying at least three distinct, genetically divergent entities (Figs. 2 and 4). Two of the three identified genetic groups correspond to the breeding nuclei, which appear to have diversified independently after their genetic isolation, as represented

by their unique SNP-based genetic profiles. Nuclei's genetic distinctiveness is also corroborated by the presented ML tree topology where the two farms belong to separate monophyletic groups more closely related with private farms than to each other (Fig. 3). Some admixed individuals sharing polymorphisms with both nuclei (Fig. 4) found at WNU could be explained by transfer of scrapie resistant males from the ENU to the WNU nucleus a decade ago, in the framework of national scrapie prevention efforts. Existing nuclei have a relatively recent common origin and have been following an ongoing directional selection towards increasing milk yield. They have also implemented the same strategy to purify the nuclei's gene pools from homozygous and heterozygous animals susceptible to scrapie disease. Thus, the genetic divergence between the two nuclei could be attributed to additional factors beyond selection or genetic improvement. Both groups are totally isolated since they have never had introductions of genetically unrelated individuals in the last 70 years, apart from limited scrapie-resistant male transfer from ENU to WNU explained above. The long isolation, the small population size of both farms (~ 400 individuals), and ongoing recent inbreeding within farms, as indicated by the relatively high F_{ROH} and distribution of ROH sizes (Fig. 5A), strengthen the hypothesis that genetic drift played an important role on differentiation. A third genetic group is represented by individuals from OUM, TEP and TAN farms exhibiting a unique, yet admixed, genetic profile with polymorphisms that are not present in breeding nuclei.

The strong genetic influence of ENU on OUF is confirmed by all presented analyses and metrics. In agreement with available ENU animal transfer records the gene flow towards OUM and TAN private farms and the WNU nucleus, as migration route, is confirmed by Treemix analysis (Fig. 3). The latter is corroborated by the low F_{ST} values between ENU - OUM and TAN – OUM. These findings highlight the consistent effort of OUM to increase the farm's livestock genetic value by enriching the gene pool mainly with males from the three groups, actively selecting individuals favoring a specific trait (milk yield). Data from the TEP farm, which is geographically located close to the ENU, suggests an admixed genetic profile derived from a unidirectional transfer of animals from the eastern nucleus. This is also supported both by the presented tree topology, relatively low genetic distances and estimated F_{ST}. The most genetically isolated farm, in terms of fixation index estimations, is the OUF pseudo-nucleus, which represents a subset of the ENU genetic diversity that probably further diverged after the separation from the original founding population fifteen years ago. Among the farms included in the present study, the genetic contribution of WNU nucleus appears to be more important for the TAN and OUM farms.

The calculated mean F_{ROH} for all Cyprus Chios sheep is analogous or relatively low compared to other European[32,33] or neighboring Turkish, Greek, and Egyptian sheep breeds[34–36], indicating limited inbreeding. The latter is also corroborated by the calculation of the same coefficient in the case of the Cyprus Damascus goat where F_{ROH} is significantly higher (~ 0.075; unpublished data). Nevertheless, the reproductive management strategy seems to vary within each studied breeding group, as F_{ROH} differs significantly across groups. The eastern breeding nucleus ENU seems to exhibit the highest inbreeding coefficients, which are a result of the small flock size and the lack of introduction of outbred individuals. However, the western nucleus (WNU) seems to maintain somewhat lower F_{ROH}, close to the mean calculated value of the breed. This fact could be attributed to (i) the enrichment of the western nucleus/population with males from the eastern population ~ 10 and ~ 5 years ago, (ii) the larger flock size of the western nucleus in the past, and (iii) the delayed start of a comprehensive evaluation and selection strategy of animals towards increasing milk yield at the western nucleus. Remarkably, the TAN group maintains low F_{ROH} (0.023), hence low inbreeding, which is a result of its relatively large flock size (~ 700 individuals) and the comprehensive record of maintained data according to which, concurrently, the avoidance of inbreeding and genetic improvement can be secured. A good practice followed by the specific farm to limit inbreeding is the preservation of a diverse gene pool as suggested by the mean genetic distance among individuals as compared to the others (Table 1). The high frequency of relatively long ROH fragments in all sampled farms (Fig. 5B), compared to other studies on sheep breeds[37–39], indicates recent inbreeding, probably in the last few generations. On the other hand, the low frequency of small fragments (< 5 Mb) shows that the contribution of ancient inbreeding to the estimated metrics is limited. Furthermore, the detection of higher ROH frequencies specifically in chromosomes 22 and 25, sets these two chromosomes as candidate carriers of genes favored by current genetic improvement schemes for dairy and/or adaptation traits in sheep (Fig. 6). Therefore, future research should give special attention to these genomic regions. Regarding the estimated effective population size, the initial decreasing trend is expected, considering the isolation of a small group of individuals from a bigger source population in Greece (Fig. 7). However, the relatively constant Ne values over the last generations (Ne ~ 200) indicate diversity preservation and implies that there is room for genetic improvement.

Findings on estimated F_{ROHs} and F_{STs} could be used to minimize future inbreeding by promoting an advanced mating strategy through unrelated male exchange between farms able to improve the desired phenotypic traits, conserving the genetic diversity of the Cyprus Chios sheep breed to the maximum feasible degree.

Although the genetic preservation of local commercial breeds is not as widely acknowledged/studied as for indigenous breeds, due to cultural and socioeconomic reasons[40], their role in food security remains crucial, considering their adaptations to the local environment and high productivity compared to indigenous ones. Commercial genetically improved breeds are still superior to local breeds in terms of productivity, a fact that is directly correlated to the genetic potential of local breeds that allowed them to adapt to local environmental conditions[41]. In the face of current climate change events, the preservation of the genetic diversity of already fairly productive and environmentally adapted local breeds is vital, aiming to ensure sustained food security. In comparison with exotic breeds, local commercial breeds also maintain an advantage, having attained through decades of adaptation, a balance between productivity and resilience to environmental perturbations such as (but not limited to) heat stress and disease resistance in climate change hotspots such as the eastern Mediterranean[42]. Although exotic commercial genetically improved breeds may produce more in highly intensive systems in the environmental conditions of continental Europe, maintaining productivity in Cyprus would require extended

periods of acclimatization and adaptation as well as extremely intense breeding conditions and thus, high input production systems. Taking into account these numbers, regarding the genetic diversity of Cyprus' livestock as well as the proven productivity and resilience of Chios sheep under intensive farming conditions and adaptability to changing environments[7,43], the genetic investment in the focal breed seems to be the most parsimonious way to achieve higher local milk productivity.

Nevertheless, the genetic divergence pattern at the population level deems necessary a more reliable downstream analysis related to the identification of important genetic regions correlated with specific traits. Unlike humans, where population stratification was determined by geographic isolation followed by genetic drift, in dairy livestock the same result may arise due to various reasons, including artificial selection, gene hitchhiking, and genetic drift[44,45]. The importance of sub-population diversity knowledge is highlighted by numerous examples in scientific literature supporting the idea that an unknown or misleading population structure could lead to high false positive or negative GWAS rates[46,47]. Hence, generated genetic data regarding the population structure of the breed could enhance efforts in this direction, minimizing the noise attributed to genetic background and focusing on groups with a common profile. Provided data can also inform valuable guidance for the design of an accurate genetic reference population to pursue reliable genomic evaluations including the private sector[48]. In addition, apart from allowing for the investigation of the breed's genetic structure, genomic data can enhance scientific research in various related fields. The evaluation of genetic resources is fundamental, among others, for the promotion of conservation[49], genetic improvement, and overall management of available livestock[50,51].

Overall, the revealed Cyprus Chios sheep diversification pattern and genetic structure calls for (i) redesign of both breeding nuclei to better preserve and represent the existing genetic diversity on the island; (ii) introduction of non-genetically related highly productive individuals in the government flocks to reduce inbreeding; (iii) designing a comprehensive reference population adequately representing the breed's genetic diversity to pursue genomic evaluations of Cyprus Chios sheep and (iv) reexamination of the role of flock size, genetic drift, and selection effects on commercial breeding nuclei genetic diversity for all small ruminant breeds following a nucleus breeding structure in Mediterranean countries and elsewhere.

Further studies are needed to explore possible correlations between population structure and milk quantity and quality traits (e.g., fat content, protein content, somatic cell count) for the focal breed. In the case of a positive correlation between specific traits of interest and genetic background, these should be taken into account in the national genetic improvement strategy. Accumulated data and knowledge are anticipated to allow researchers and farmers to focus on specific goals and traits, following a more efficient and direct route leading to the genetic improvement of the country's small ruminant population. Finally, data generated within the framework of this study could be utilized for future sheep crossbred population studies.

Methods
Sampling - data generation
Individuals from the two governmental breeding nuclei and four private farms that are geographically dispersed across the island and members of the Breeder Society of Cyprus Chios sheep Breeders were sampled. Blood samples were taken from 1064 purebred Cyprus Chios sheep individuals across six farms distributed all over Cyprus. More specifically, 409 individuals were sampled from ENU, 286 from WNU, 96 from TEP, 92 from OUF, 90 from TAN and 91 from OUM. A map presenting all sampling locations was constructed with the R package Mapview v.2.11.2[52].

Genomic data generation
Total genomic DNA was extracted from available specimens using the MagNA Pure 96 DNA and Viral NA Large Volume Kit on the MagNA Pure 96 System. DNA quantity and purity were assessed with NanoDrop 2000/200c (Thermo Fisher Scientific Inc., USA). DNA samples were forwarded to Weatherbys facilities in Ireland for genotyping. Genome-wide SNPs were retrieved using the illumina OvineSNP50v3_XT bead chip, while quality control and raw data files were generated with Illumina GenomeStudio Software v2.0.5.

Sex chromosomes as well as individuals and SNPs with more than 10% missing data were removed from the final dataset using GenomeStudio v2.0.5. Data were also filtered, removing SNPs with low minor allele frequencies, setting the threshold to 0.05 using PLINK v1.9 software[53].

Genetic diversification - population structure - inbreeding
The absolute genetic divergence (p-distance) was calculated using MEGA v.11[54] between and within predefined groups based on individual's farm of origin. The diversification was further explored with the non-model-based, independent from Hardy-Weinberg equilibrium, DAPC[26] to evaluate the degree of genetic affinity of sampled populations, considering predefined groups (farms). The calculated membership probability represents the overall genetic background of an individual, indicating the genetic divergence between examined populations. The analysis was implemented using the R package Adegenet v2.0.0[55]. The optimal number of clusters was determined by applying the Bayesian Information Criterion (BIC) to K-means clustering solutions.

A Maximum Likelihood tree was constructed with OrientAGraph (v1.0;[28]) within Treemix v1.13[29] employing the -allmigs and -mlno aiming to perform a network orientation and evaluate all possible ways of adding a migration edge to the base tree. Data files were converted from plink format to .vcf and then to treemix format using vcf2treemix.py python script[56]. Results were visualized using plotting functions on R (v4.2.2;[57]).

Genome-wide data were also evaluated under a model-based Bayesian clustering approach, aiming to explore the existence of population structure patterns under this alternative. For this purpose, available data were properly formatted and fed to Structure v.2.3.4. Ten replicates for each K varying from 1 to 10 were run

in parallel with STRUCTURE THREADER[58] for 50,000 generations, setting burn-in to 10,000. The best K for our data was estimated by monitoring the calculated mean Ln likelihood and with the ΔK Evanno method conducted in STRUCTURE HARVESTER[31]. Unimodality or alternative solutions among runs was assessed with CLUMPAK[59]. Structure results were visualized using the python script distruct2.3.py[60,61].

Inbreeding – effective population size

The gene flow between predefined groups was further assessed using the R package Hierfstat[62] for the calculation of pairwise population fixation index (F_{ST}) under the WC model[63]. Lower F_{ST} values indicate the vivid exchange of animals and hence gene flow between examined farms while high values correspond to genetic isolation. To verify the statistical validity of our findings, a total of 100 bootstrap replicates were carried out. The extent of inbreeding within farms and for all Cyprus Chios sheep breed individuals included in our analyses was estimated through Runs of Homozygosity (ROH) detection across all SNPs. The analysis was conducted under the following settings: (i) a maximum gap between adjacent SNPs of 250 Kb, (ii) a minimum SNP density of 100 Kb per ROH, (iii) allowance of up to one missing SNP and one heterozygous genotype per ROH, and (iv) a missing genotype threshold based on ROH length: 0 missing for > 1 Mb, 1 for > 4 Mb, 2 for > 8 Mb, 4 for > 12 Mb, and 5 for > 16 Mb, using PLINK v1.9 software[53]. Generated results were plotted in violin format using the R package detectRUNS v. 0.9.6[64]. Finally aiming to evaluate how recent the observe inbreeding is the distribution on ROH sizes against frequency were plotted using the same package.

The effective population size (Ne) in the studied sheep breed was estimated using SNeP v1.1[65], which employs the linkage disequilibrium (LD) method. The average r^2 for markers separated by various genomic distances was calculated for this purpose.

Data availability

The datasets generated and analysed during the current study are not publicly available due to restrictions of the funding scheme before the official end of the project in 2026 but are available from the corresponding author on reasonable request.

References

1. Abell, J. T. et al. Urine salts elucidate early neolithic animal management at Asikli Hoyuk, Turkey. *Sci. Adv.* **5**, eaaw0038. https://doi.org/10.1126/sciadv.aaw0038 (2019).
2. Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci.* **105**, 11597–11604. https://doi.org/10.1073/pnas.0801317105 (2008).
3. Atag, G. et al. Population genomic history of the endangered Anatolian and Cyprian Mouflons in Relation to Worldwide Wild, Feral, and domestic Sheep lineages. *Genome Biol. Evol.* **16** https://doi.org/10.1093/gbe/evae090 (2024).
4. Vigne, J. D. et al. The transportation of mammals to Cyprus sheds light on early voyaging and boats in the Mediterranean Sea. *Eurasian Prehistory*. **10** (1–2), 157–176 (2013).
5. Guerrini, M. et al. Molecular DNA identity of the mouflon of Cyprus (Ovis orientalis ophion, Bovidae): Near Eastern origin and divergence from western Mediterranean conspecific populations. *Syst. Biodivers.* **13**, 472–483. https://doi.org/10.1080/14772000.2015.1046409 (2015).
6. Sanna, D. et al. The first mitogenome of the cyprus mouflon (Ovis Gmelini ophion): New insights into the phylogeny of the Genus Ovis. *PLoS ONE*. **10**, e0144257. https://doi.org/10.1371/journal.pone.0144257 (2015).
7. Pulina, G. et al. Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. *J. Dairy. Sci.* **101**, 6715–6729. https://doi.org/10.3168/jds.2017-14015 (2018).
8. FAOSTAT. Food and Agriculture Organization of the United Nations, (2024).
9. Mavrogenis, A. P. Environmental and genetic factors influencing milk production and lamb output of chios sheep. *Livest. Prod. Sci.* **8**, 519–527. https://doi.org/10.1016/0301-6226(82)90029-X (1982).
10. Mavrogenis, A. P., Papachristoforou, C., Lysandrides, P. & Roushias, A. Environmental and genetic factors affecting udder characters and milk production in Chios sheep. *Génétique sélection évolution*. **20** (477). https://doi.org/10.1186/1297-9686-20-4-477 (1988).
11. Ligda, C., Papadopoulos, T., Mavrogenis, A. & Georgoudis, A. in *Breeding programmes for improving the quality and safety of products. New traits, tools, rules and organization?* (eds Gabiña D. & Sanna S.) 55–59 (2003).
12. Mavrogenis, A. P. & Papachristoforou, C. Genetic and phenotypic relationships between milk production and body weight in Chios sheep and Damascus goats. *Livest. Prod. Sci.* **67**, 81–87. https://doi.org/10.1016/S0301-6226(00)00187-1 (2000).
13. Nikolaou, M., Kominakis, A. P., Rogdakis, E. & Zampitis, S. Effect of mean and variance heterogeneity on genetic evaluations of Lesbos dairy sheep. *Livest. Prod. Sci.* **88**, 107–115. https://doi.org/10.1016/j.livprodsci.2003.09.025 (2004).
14. Kominakis, A., Rogdakis, E., Vasiloudis, C. & Liaskos, O. Genetic and environmental sources of variation of milk yield of Skopelos dairy goats. *Small Ruminant Res.* **36**, 1–5. https://doi.org/10.1016/S0921-4488(99)00105-4 (2000).
15. Volanis, M., Kominakis, A. & Rogdakis, E. Genetic analysis of udder score and milk traits in test day records of Sfakia dairy ewes. *Arch. Anim. Breed.* **45**, 69–77. https://doi.org/10.5194/aab-45-69-2002 (2002).
16. Orford, M. et al. A single nucleotide polymorphism in the acetyl-coenzyme A acyltransferase 2 (ACAA2) gene is associated with milk yield in Chios sheep. *J. Dairy. Sci.* **95**, 3419–3427. https://doi.org/10.3168/jds.2011-4732 (2012).
17. Banos, G. et al. Genetic and genomic analyses underpin the feasibility of concomitant genetic improvement of milk yield and mastitis resistance in dairy sheep. *PLoS ONE*. **14**, e0214346. https://doi.org/10.1371/journal.pone.0214346 (2019).
18. Li, R., Ma, Y. & Jiang, L. Review Research progress of dairy Sheep milk genes. *Agriculture* **12**, 169. https://doi.org/10.3390/agriculture12020169 (2022).
19. Rezvannejad, E., Asadollahpour Nanaei, H. & Esmailizadeh, A. Detection of candidate genes affecting milk production traits in sheep using whole-genome sequencing analysis. *Veterinary Med. Sci.* **8**, 1197–1204. https://doi.org/10.1002/vms3.731 (2022).
20. Hayes, B. & Goddard, M. E. The distribution of the effects of genes affecting quantitative traits in livestock. (2001).
21. Meuwissen, T., Hayes, B. & Goddard, M. Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* **6**, 6–14. https://doi.org/10.2527/af.2016-0002 (2016).
22. Gelasakis, A. I., Valergakis, G. E., Fortomaris, P. & Arsenos, G. Farm conditions and production methods in Chios sheep flocks. *J. Hellenic Veterinary Med. Soc.* **61**, 111–119. https://doi.org/10.12681/jhvms.14880 (2018).

23. Lawson Handley, L. J. et al. Genetic structure of European sheep breeds. *Heredity (Edinb)*. **99**, 620–631. https://doi.org/10.1038/sj.hdy.6801039 (2007).
24. DAG. (ed. *Department of Agriculture),* Ministry of Agriculture, Rural Development and Environment (2023).
25. Basdagianni, Z., Sinapis, E. & Banos, G. Evaluation of reference lactation length in Chios dairy sheep. *Animal* **13**, 1–7. https://doi.org/10.1017/S1751731118000769 (2019).
26. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. https://doi.org/10.1186/1471-2156-11-94 (2010).
27. Schwarz, G. Estimating the dimension of a model. *Annals Stat.* **6**, 461–464 (1978).
28. Molloy, E. K., Durvasula, A. & Sankararaman, S. Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics* **37**, i142–i150. https://doi.org/10.1093/bioinformatics/btab267 (2021).
29. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967. https://doi.org/10.1371/journal.pgen.1002967 (2012).
30. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. https://doi.org/10.1007/s12686-011-9548-7 (2012).
31. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x (2005).
32. Machová, K. et al. Genetic diversity of two native sheep breeds by genome-wide analysis of single nucleotide polymorphisms. *Animal* **17**, 100690. https://doi.org/10.1016/j.Animal.2022.100690 (2023).
33. Nosrati, M., Esmailizadeh, A. & Asadollahpour Nanaei, H. Estimation of runs of homozygosity reveals moderate autozygosity in north European sheep breeds. *J. Livest. Sci. Technol.* **9**, 31–40 (2021).
34. Bayraktar, M. Analysing the genetic diversity of three sheep breeds in Turkey and nearby countries using 50 K SNPs data. *Anim. Biotechnol.* **35**, 2329106. https://doi.org/10.1080/10495398.2024.2329106 (2024).
35. Hossein, M. & Mohammad, S. Estimation of genomic inbreeding coefficient, detection of ROH Islands and related genes in different Egyptian sheep breeds adapted to different environment. *Iran. J. Anim. Sci.* **55**, 95–109 (2024).
36. Kominakis, A., Tarsani, E., Hager-Theodorides, A. L., Mastranestasis, I. & Hadjigeorgiou, I. Clustering patterns mirror the geographical distribution and genetic history of Lemnos and Lesvos sheep populations. *PLoS ONE.* **16**, e0247787. https://doi.org/10.1371/journal.pone.0247787 (2021).
37. Deniskova, T, Dotsev, A., Selionova, M., Brem, G. & Zinovieva, N. Biodiversity of Russian Local sheep breeds based on pattern of runs of homozygosity. *Diversity* **13** (2021).
38. Liu, J. et al. Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds. *J. Anim. Sci. Biotechnol.* **12**, 95. https://doi.org/10.1186/s40104-021-00608-9 (2021).
39. Nosrati, M., Nanaei, A., Javanmard, H., Esmailizadeh, A. & A. & The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. *Genomics* **113**, 1407–1415. https://doi.org/10.1016/j.ygeno.2021.03.005 (2021).
40. Ovaska, U. et al. The conservation of native domestic animal breeds in nordic countries: from genetic resources to cultural heritage and good governance. *Animals* 11 (2021).
41. Wanjala, G. et al. A review on the potential effects of environmental and economic factors on sheep genetic diversity: Consequences of climate change. *Saudi J. Biol. Sci.* **30**, 103505. https://doi.org/10.1016/j.sjbs.2022.103505 (2023).
42. Serranito, B. et al. Local adaptations of Mediterranean sheep and goats through an integrative approach. *Sci. Rep.* **11**, 21363. https://doi.org/10.1038/s41598-021-00682-z (2021).
43. Tsartsianidou, V. et al. A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. *Genet. Sel. Evol.* **53**, 90. https://doi.org/10.1186/s12711-021-00682-7 (2021).
44. Ma, L. et al. Effect of sample stratification on dairy GWAS results. *BMC Genom.* **13**, 536. https://doi.org/10.1186/1471-2164-13-536 (2012).
45. Sonstegard, T. S. et al. in *9th World Congress on Genetics Applied to Livestock Production.* (2011).
46. He, J. & Gai, J. *Plant Genotyping: Methods and Protocols* (ed. Yuri, S.) 123–146 (Springer US, 2023).
47. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354. https://doi.org/10.1038/ng.548 (2010).
48. Moghaddar, N., Swan, A. A. & van der Werf, J. H. Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. *Genet. Sel. Evol.* **46**, 58. https://doi.org/10.1186/s12711-014-0058-4 (2014).
49. Fernández, J., Toro, M. A., Gómez-Romano, F. & Villanueva, B. The use of genomic information can enhance the efficiency of conservation programs. *Anim. Front.* **6**, 59–64. https://doi.org/10.2527/af.2016-0009 (2016).
50. Brito, L. F. et al. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal* **15** (Suppl 1), 100292. https://doi.org/10.1016/j.animal.2021.100292 (2021).
51. Haile, A. et al. Community-based sheep breeding programs generated substantial genetic gains and socioeconomic benefits. *Animal* **14**, 1362–1370. https://doi.org/10.1017/S1751731120000269 (2020).
52. Appelhans, T. et al. *mapview: Interactive Viewing of Spatial Data in R*. R package version 2.11.2. (2023). https://github.com/r-spatial/mapview
53. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4** https://doi.org/10.1186/s13742-015-0047-8 (2015).
54. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
55. Jombart, T. & Ahmed, I. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521 (2011).
56. RAD tools. vcf2treemix.py (2018). https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2treemix.py.
57. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
58. Pina-Martins, F., Silva, D. N., Fino, J. & Paulo, O. S. Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. *Mol. Ecol. Resour.* **17**, e268–e274. https://doi.org/10.1111/1755-0998.12702 (2017).
59. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191. https://doi.org/10.1111/1755-0998.12387 (2015).
60. Chhatre, V. E. Distruct v2.3, A modified cluster membership plotting script (2018). http://distruct2.popgen.org
61. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. https://doi.org/10.1534/genetics.114.164350 (2014).
62. Goudet, J. HIERFSTAT, a package for r to compute and test hierarchical F-statistics. *Mol. Ecol. Notes.* **5**, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x (2005).
63. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370. https://doi.org/10.2307/2408641 (1984).
64. Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. detectRUNS: An R package to detect runs of homozygosity and heterozygosity in diploid genomes. *Front. Genet.* **9**, 747. https://doi.org/10.3389/fgene.2018.00747 (2018).

11

65. Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M. W. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* **6**, 109. https://doi.org/10.3389/fgene.2015.00109 (2015).

Acknowledgements

Author contributions

A.C.D. Conceptualization, Investigation, Writing – original draft; A.C.D., G.M. Methodology, Formal analysis; A.C.D., G.M., G.H. Data curation; G.H. Resources, Funding acquisition, Project administration; All authors contributed to writing, read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

All participating farm owners' consent to sample animals for experimentation purpose. All experimental protocols were approved by Agriculture Research Institute of Cyprus which is part of the Ministry of Agriculture, Rural Development and Environment. All methods were carried out in accordance with relevant guidelines and regulations. All methods are reported in accordance with ARRIVE guidelines.

Additional information

Correspondence and requests for materials should be addressed to A.C.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.