Theses of Doctoral (PhD) Dissertation

Genome-wide analysis of the genetic diversity of native sheep breeds and determination of potential selection signatures for climate change adaptation

George Wanjala
Ph.D. candidate

Supervisor:
Prof. Dr. Szilvia Kusza, D.Sc.

UNIVERSITY OF DEBRECEN
Doctoral School of Animal Science

Debrecen, 2024

1. INTRODUCTION

Globally, indigenous sheep breeds have seen significant reductions in population size and are now facing a state of vulnerability. The understanding of native breed's genetic architecture and their genomic response to local environmental selection pressures remains limited, however, they (native breeds) offer an opportunity to potentially uncover genomic areas that are more valuable for adapting to climate change. Many studies have been conducted to explore the relationship between the genetic and phenotypic diversity of indigenous animal breeds and their ability to contribute to food security. The majority of these research works were conducted within limited geographical bounds, so limiting their ability to provide a full grasp of the subject matter on a global scale. Hence, the existing body of information pertaining to the genetic diversity, interrelationships, and adaptability of indigenous sheep breeds is fairly limited. There is a possibility that these breeds possess significant genetic variations that enable them to adapt to climate change, and these variations might potentially be incorporated into commercial breeds using genomic selection technology. Therefore, it is imperative to enhance the conservation efforts of endangered indigenous breeds and explore their adaptive characteristics via the use of contemporary genomic and bioinformatics methodologies.

The present doctoral research involves a comprehensive examination of the genomic architecture of indigenous sheep breeds that may have developed adaptations to diverse climatic zones. Furthermore, this study examines the current signatures of selection for climate change adaptability and the bio-climatic variables influencing the adaptation of sheep breeds to climatic change.

The results of this research are expected to be used by sheep breeders in order to optimize the utilization of gene-environment interaction for enhanced productivity and adaptation to the specific local environment. Moreover, the findings of this research will provide a chance to enhance the development of a breeding program aimed at maintaining genetic diversity, enhancing animal welfare and ecological management. The research will provide guidance to organizations involved in the conservation of sheep genetic resources, namely in the planning and prioritization of resources for the management and conservation of vital sheep genetic resources.

Therefore, to provide a comprehensive understanding of the genomic structure of native sheep breeds, this study aimed to achieve the following four specific objectives:

i. To perform genetic characterization to determine within and between populations genetic diversity.
ii. To detect patterns of run of homozygosity (ROH) and genomic based inbreeding coefficient (Froh) per breed.
iii. To identify signatures of selection underpinning adaptation in different climatic regions.
iv. To perform gene x environment association (GEA) to explore the complex interplay between genes and climatic variables in shaping adaptation traits.

2. MATERIALS AND METHODS

2.1. Ethics statement

This research did not require approval by the Scientific and Ethics Committee of the Centre for Agricultural Genomics and Biotechnology of the University of Debrecen.

2.2. Sample description, collection, and genotyping

This research included the analysis of 907 samples obtained from a total of 32 different native sheep breeds. Out of these samples, 620 were specifically genotyped for the purpose of this study, representing 22 breeds, hence forth referred to as raw samples/data. The other samples were obtained from publicly available databases. For instance, samples from the Magreb region were studied by BEN JEMAA et al (2019), those from Sweden were analyzed by ROCHUS et al (2020), and samples from South Africa, Iran, and China were published by KIJAS et al. (2009), whereas samples from Morocco were donated by a collaborator. Table 1 provides a comprehensive overview of the sample sizes for each breed, together with their respective geographic origins and the presumed climatic conditions to which they are adapted to. Three breeds, namely HUN.IldeFr, HUN.Suff, and HUN.Tetra, were included into the study as outgroups, indicating that they are not indigenous to Hungary. In order to get well represented results, within a certain breed, samples were collected from distinct flocks. All efforts were made to include populations representing non-related adult individuals from native sheep populations.

Meanwhile, raw samples were obtained by collecting blood samples from individual animals via puncturing their jugular veins. The acquired blood samples were then stored in vacutainer tubes (THOMSON et al., 1983) and afterwards placed onto the Flinders Technology Associates (FTA) cards® (DASH et al., 2020; HSIAO et al., 1999). The dried FTA cards were stored at ambient temperature prior to their submission to the genotyping company. The whole procedure of DNA isolation, genotyping, and conversion of raw signals into ovine50K genotype was delegated to Neogen company limited (https://www.neogen.com), which used the Infinium® OvineSNP50 BeadChip array which features more than 50,000 evenly distributed SNP-targeting probes, providing adequate SNP density for genetic diversity and signatures of local selection investigation. The genotype calls were determined using the ovine Oar_v3.1 assembly (CLARK et al., 2017) as a reference genome, from which the final report was derived.

Table 1: A summary of all samples used in the study before quality control

Country	Breed name	Climatic zone	Breed code	N
Hungary	Hungarian Tsigai	Continental climate	HUN.Tsig	28
	Hungarian Merino		HUN.Mer	29
	Dorper		HUN.Dor	21
	White Dorper		HUN.Dor.W	20
	Bábolna Tetra		HUN.Tetra	27
	Ile de France		HUN.IldeFr	27
	Suffolk		HUN.Suff	20
	Hortobágy Racka		HUN.Rac	28
Bosnia-Herzegovina	Kupreška Pramenka	Continental climate	BIH.KPram	31
	Hercegovačka Pramenka		BIH.HPram	27
	Dubska Pramenka		BIH.DPram	25
Greece	Boutsiko	Mediterranean	GRC.Bout	30
	Chios		GRC.Chios	32
Romania	Rusty Tsigai	Continental climate	ROU.Rtsig	30
	White head Turcana		ROU.Turc	30
Serbia	Lipe Pramenka	Continental climate	SRB.LPram	25
	Zackel Pramenka		SRB.ZPram	20
Poland	Aleksandrowice	Continental climate	POL.Alek	25
	Cakiel Podhalanski		POL.Pod	25
Sweden	Dalapäls	Nordic climate	SWE.Dal	18
	Klövsjö		SWE.Klo	19
	Gotland		SWE.Got	19
Lithuania	Coarse-wooled sheep		LTU.Bface	30
Kenya	Red Maasai	Tropical climate	KEN.RedM	50
	East African fat tail		KEN.Eaft	40
Ethiopia	Menz		ETH.Menz	34
Morocco	Morocco	North Africa’s Mediterranean	MAR	30
Magreb region	Margeb		MARG	40
South Africa	South Africa Dorper	Sub-Tropical climate	SA.Dor	21
China	Tibetan	Sub-tropical climate	CHN.Tib	37
Iran	Afshar		IR.Afs	35
	Morganii		IR.Mog	34
Total				907

N: Number of samples before QC

2.3. Quality Control and Genetic Diversity Analyses

The raw samples acquired from Neogen were processed and transformed into plink binary files (CHANG et al., 2015) for the purpose of merging using PLINK v1.9 (PURCELL et al., 2007) them with the downloaded datasets. These merged files were then used for further analysis. The process of data filtering was conducted by using the following predetermined criteria. Individuals that were closely related (first line relatives) or duplicates, variants on unknown and/ or sex chromosome, individuals with missing genotypes exceeding 10%, SNPs that were genotyped in less than 90% of the samples, SNPs with a minor allele frequency below 0.05%, and SNPs that deviated from the Hardy-Weinberg equilibrium (HWE) threshold of $p < 10^{-3}$ were all excluded from the analysis. After conducting quality control measures, a total of 41,098 SNPs were omitted from the analysis due to missingness. Additionally, 115 SNPs were removed based on the HWE threshold. Furthermore, 180 SNPs were disqualified from the study based on the MAF criterion. Moreover, 40 samples were excluded from the analysis due to missing genotypes, and an additional 35 samples were removed due to their degree of relatedness. The aforementioned procedure yielded a dataset including 832 samples and 42,140 SNPs, which was then used for downstream analysis. Henceforth, the dataset resulting from the implementation of quality control measures on the SNP data will be denoted as the working dataset.

2.4. Statistical analysis

2.4.1. Determination of within and between breeds genomic variation and between sheep breeds genomic relationship

Four statistical approaches were used to determine the genetic diversity levels of the studied breeds. The statistical approaches included: observed, expected heterozygosity and inbreeding coefficient, effective population size through generations, between population relationship by PCA and phylogenetic tree, and ancestry analysis. To achieve these, additional QC procedures were applied to the working dataset to ensure the accurate assessment of genetic diversity within and among sheep populations. The quality control techniques included the removal of SNPs that showed linkage disequilibrium (LD). This measure aimed at minimizing the potential bias resulting from a small set of genomic regions as explained by YUAN (2017). The LD pruning criteria were used to exclude single nucleotide polymorphisms (SNPs) until the correlation (r^2) between any two SNPs did not exceed 0.5 within a 200 Kilobase pair range. The calculation of genetic diversity indices was performed using Arlequin software v 3.5 (EXCOFFIER et al., 2005). The computation of genetic diversity metrics, such as expected heterozygosity (H_e) and observed heterozygosity (H_o), was conducted using haplotype frequency. Furthermore, the assessment of inbreeding was conducted using the method known as moments F coefficient estimates, employing the following formula:

$$\frac{Observed\ Homozygosity\ count - expected\ count}{Total\ observations - expected\ count}$$

Whereby the determination of expected count is based on the minor allele frequencies (MAFs; PURCELL et al., 2007). The determination of effective population size was performed using linkage disequilibrium information (SVED, 1971). The SNeP software, developed by BARBATO et al. (2015), was used to compute the effective population size over generations using non-LD pruned data set. Analysis of molecular variance (AMOVA; MEIRMANS, 2006) using the specially designed population genetics tool Arlequin v 3.5 (EXCOFFIER et al., 2005) was used to determine genetic variation among individuals within the population and among all the 32 studied sheep breeds. AMOVA is a strong statistical tool for describing factors impacting population structure (MENGONI and BAZZICALUPO, 2002).

To ascertain the genomic correlation/relationship between populations and, the genomic homogeneity of the animals collected, unsupervised multidimensional clustering statistical analysis (ANDERSON, 1962), commonly known as principal component analysis (MAĆKIEWICZ and RATAJCZAK, 1993) was computed on all genotypes using plink v1.9 (CHANG et al., 2015). To ascertain the evolutionary links between different sheep breeds, the computation of Wright's FST index was performed on the working dataset using R package StAMPP (PEMBLETON et al., 2013) and visualized in R studio. To determine ancestry and quantify population structure, ADMIXTURE v1.3.0 (ALEXANDER et al., 2015) was used. The investigation was conducted using pre-determined ancestry clusters (K), which were believed to between 2 and 36 ancestral populations. The value of K with the lowest 15-fold cross validation error was chosen as the most likely. The Q matrix that was outputted was visualized using the BITE package v2 (MILANESI et al., 2017) using membercoef.plot function which exports a circular plot.

2.4.2. Determination and detect patterns of run of homozygosity (ROH) and genomic based inbreeding coefficient (Froh) per native sheep breed

Runs of Homozygosity are stretches of DNA within an individual's genome where all the alleles are homozygous. These patterns provide valuable insights into an individual's ancestry, evolutionary history, and the signatures of natural selection on the genome. On the other hand, genomic inbreeding, ROH-based inbreeding, or FROH (fraction of homozygosity), is a metric that quantifies the fraction of the genome occupied by extended segments of ROH. The calculation of ROHs was performed using the detectRUN package (BISCARINI et al., 2018) in the R program, using the sliding window approach. Given the absence of widely established guidelines for calculating the ROHs, we used the criteria outlined by DZOMBA et al. (2021). To summarize, the variants were not trimmed based on linkage disequilibrium (LD), i.e the dataset used in to calculate ROH was non-LD pruned. However, the minimum length necessary for a region to be termed a run of homozygosity (ROH) was raised to 1 megabases (mb) to remove any short ROH segments that may have been influenced by LD. The calculation of the sliding window was performed using the below parameters: window size of 15 mb and applying a threshold value of 0.05 to determine the eligibility of a SNP in the run. A minimal threshold of one heterozygous genotype was established, allowing for the possibility of one missing genotype. To meet the criteria for completeness, the ROH window was required to include a minimum of 30 SNPs. Furthermore, the minimum SNP density per ROH was established as one SNP per every 100 kilobases (kb), while the maximum distance between consecutive SNPs was defined as 1 mb. The identified ROHs were classified into four distinct groups based on their length: 1–6 mb, 6–12 mb, 12–24 mb, and over 48 mb. The cumulative length of the runs of homozygosity (ROH) in each individual animal was computed and then averaged across all breeds to ascertain the breed-specific average sum of ROH length. Furthermore, the numbers of runs for each category were recorded. The study included the determination of the total number of ROH, the mean number of ROH, and the average length of ROH for each breed. Additionally, the peak runs, which are the runs in which SNPs occur in more than a certain threshold of samples in a population, were determined. SNPs that occurred in runs in at least 70%, 60%, and 50% of samples in the population were used as the criteria for defining the peak runs.

2.4.3. Identification of selection signatures for climate change adaptation

To identify possible genetic selection for adaptation to climatic changes, we performed a pairwise comparison analyses of populations that were categorized depending on climate. The use of pairwise comparison was employed to mitigate the potential biases that may arise from undisclosed population structures. This investigation included the formation of four distinct groupings of populations depending on climate. (i) The population of sheep in tropical climates consisted exclusively of breeds originating from East Africa, specifically KEN.Eaft, KEN.RedM and ETH.Menz. (ii) The Mediterranean/desert climatic region included breeds from North Africa, specifically MAR and MARG. (iii) The group of sheep breeds in continental climatic regions included native breeds from Europe, with the exception of breeds from Sweden, Lithuania, HUN.Dor, HUN.Dor.W, and the outgroups. (iv) The Oceanic climate category included breeds from Lithuania and Sweden.

To compute and search for climate change adapted genomic selection sweeps, two complimentary statistical methods were used. Wright's fixation index, often known as FST statistics (BEAUMONT, 2005*)*, was computed in plink 1.9 to ascertain genetic differentiation between two paired populations and to discover variations in allele frequencies within these populations. It is postulated that all SNPs are subject to neutral selection and are expected to exhibit little heterogeneity. Consequently, SNPs that deviate from this standard may potentially account for the observed variance between populations and are hypothesized to be subject to natural selective pressures. The second approach used in this study was Cross Population Extended Haplotype Homozygosity (XP-EHH) as described by GAUTIER and VITALIS (2012). This methodology quantifies the length of extended haplotypes and evaluates the dissimilarities between the paired populations. The XP-EHH method has been extensively used in the identification of prospective signatures of local adaptation across several domestic livestock species (CHERUIYOT et al., 2018; EDEA et al., 2019; PAIM et al., 2022). A specific genomic region, referred to as the region of interest (RoI), was considered as a potential candidate for selection sweep if it achieved a log P-value or FST value of above 0.01%. The genes were subsequently annotated within a range of 250 kilobases upstream and downstream of the probable RoI using the BioMaRt software (DURINCK et al., 2005). The ensemble database (HUBBARD et al., 2002) was used for gene hunt and annotation. Review of multiple scientific literature was also useful in determining functions of identified genes. The visualization of signatures of selection was done using a combination of R packages including tidyverse (WICKHAM and WICKHAM, 2017) and qqman *(*TURNER, 2014).

2.4.4. Genome x climatic variable analysis to determine climatic variable driving adaptation of native sheep breeds in different climatic zones

The gene x environment association or landscape genomics (MANEL et al., 2003) is a prominent approach for identifying the environmental factors that are most likely to impact breed adaptation. This methodology seems to be advantageous in comprehending the effects of climate change on the agricultural sector, with a particular emphasis on livestock species. Historically, the focus of study has mostly been on human-mediated selection signatures because of their significant economic implications. In light of the increasing apprehension about climate change and its implications for genetic diversity, there has been a recent surge in interest among academics regarding environmentally induced selection signatures, which are indicative of natural selection. Consequently, the acquisition of reliable outcomes relating to the signatures of natural selection may be achieved through the use of indigenous breeds that have acclimated to diverse climatic regions. The use of global climate databases e.g., WorldClim, enables the acquisition of comprehensive global climatic information, hence facilitating an indirect methodology for investigating genetic adaptability in response to climate change. The geographical location from where the samples were obtained may serve as proxies for phenotypic adaptations to the environment. This method has been used to identify environmental variables influencing natural selection and genes likely to be under selection for adaptation (FLORI et al., 2019; GOITOM et al., 2021; WIENER et al., 2021).

In pursuit of this objective, just those samples that had GPS coordinates were subjected to this analysis. Here, we conducted an analysis on a total of 626 samples, which included samples from Kenya, Ethiopia, Morocco, as well as other European samples that were genotyped for this study. Prior to analysis, rigorous quality control measures were used to ensure the reliability and accuracy of the data.

Nineteen (19) annual means of climatic variables for each sample were obtained from the WorldClim database using the corresponding GPS coordinates. The WorldClim database has historical bioclimatic variables that provide primary quantitative data on temperatures and precipitation at yearly, weekly, and seasonal intervals. These variables serve as a surrogate to characterize the climate. The data provided provide a concise representation of climatic conditions spanning from 1970 to 2000. These variables are in the form of raster, with a spatial resolution of 30 seconds (~1km^2).

The degree of collinearity was assessed by the use of principal component analysis (PCA), Pearson correlation, and the variance inflation factor (VIF). The correlation among all variables was assessed using base R, and then, the covariates were trimmed to ensure that the correlation between any pair of variables did not surpass 0.7. The validation process included calculating the VIF and selecting a subset of variables with VIF values below 5. Here below is the description of each climatic variable obtained from the WorldClim database.

i. bio_1 = Annual mean temperature
ii. bio_2 = Mean diurnal range (mean of monthly (max temp - min temp))
iii. bio_3 = Isothermality (bio_2/ bio_7) (×100)
iv. bio_4 = Temperature seasonality (standard deviation ×100)
v. bio_5 = Max temperature of warmest month
vi. bio_6 = Min temperature of coldest month

vii. bio_7 = Temperature annual range (bio_5- bio_6)
viii. bio_8 = Mean temperature of wettest quarter
ix. bio_9 = Mean temperature of driest quarter
x. bio_10 = Mean temperature of warmest quarter
xi. bio_11 = Mean temperature of coldest quarter
xii. bio_12 = Annual precipitation
xiii. bio_13 = Precipitation of wettest month
xiv. bio_14 = Precipitation of driest month
xv. bio_15 = Precipitation seasonality (coefficient of variation)
xvi. bio_16 = Precipitation of wettest quarter
xvii. bio_17 = Precipitation of driest quarter
xviii. bio_18 = Precipitation of warmest quarter
xix. bio_19 = Precipitation of coldest quarter

To determine the spatial genetic structure of populations, the K-means on principal components was used to detect the genetic clusters that minimize within cluster genetic variance, which turned to be 7. Discriminant analysis of the principal component was performed on the genotypes where 3 discriminant functions were retained.

Two statistical approaches were used in the present landscape genomics analysis: latent factor mixed models (LFMM) implemented by LFMM R package (CAYE et al., 2019) and redundancy discriminant analysis (RDA) implemented in (DIXON, 2003). Since the genome wide association studies (GWAS) encounter challenges related to confounding latent factors, LFMM package is designed to provide novel strategies for estimating parameters in LFMM. The algorithms have been specifically developed to address the issue of unobserved confounding variables (CAYE et al., 2019; SERRANITO et al., 2021).

Prior to running LFMMs, missing genotypes were imputed by relying on the best K from sNMF analysis. Then LFMM ridge regression was performed by regressing the genotypes against the five environmental variables that remained after quality control. In this analysis, seven latent factors (or 7 clusters) were used to describe the population structure. Signatures of selection were identified by computing the p-values for all the SNP by environment associations and then calibrated them by genomic inflation factor to reduce higher chances of obtaining false positives and negatives.

In contrast, the RDA function of Vegan package allows to test the impact of specific environmental variables on the dataset. It selectively retains only those variables that have a statistically significant influence on the molecular matrix. The process of 'variable selection' may be carried out via a 'forward model building approach' (WHITLEY et al., 2000). The process of forward model selection indicated that all the 5 environmental variables were significant, hence all were included in the association analysis. The R-squared adjustment was used to account for the percentage of variance explained by the environmental predictors. Subsequently, the statistical significance of the RDA model was assessed by F-statistics, with the null hypothesis stating that there is no linear relationship between the SNP data and the environmental variables. The identification of candidate SNPs associated with local adaptation was done by assessing their loading in the ordination space. In this analysis, we calculated SNP loadings specifically from the three constrained axes although the identified were seven. Constrained axes were deliberately reduced to three since it is assumed that the first three axes have a higher impact than others. The detection of potential SNPs under divergent selection were achieved by examining the loadings that exceeded ±3 standard deviations from the average of each distribution, which corresponds to a two-tailed p-value of 0.0027 or below (FORESTER et al., 2018).

3. RESULTS

3.1. Within-breed genetic diversity and inter-breed genetic relationships

3.1.1. Within breed genetic diversity

Initially, a total of 907 samples and 83,533 SNPs pre-quality control were involved in the analysis. However, due to strict quality control measures that were implemented, 832 samples and 42,140 SNPs remained for downstream analysis. Overall, the diversity indices suggest that the studied breeds exhibited comparable levels of H_o and H_e, with indications of varying degrees of inbreeding seen across the breeds. The estimates of H_o for all 32 sheep breeds ranged from 0.309 to 0.392, while the estimates of H_e ranged from 0.331 to 0.389. The F coefficients, on the other hand, varied between 0.032 and 0.362, as shown in Table 2. The examination of Ne throughout several generations, spanning up to 915 generations in the past, revealed a significant and rapid decline in the effective population size (Figure 1). Several European breeds had a Ne value below 100, with the exception of BIH.Pram, GRC.Bout, HUN.IldeFr, HUN.Mer, ROU.Turc, ROU.RTsig, and POL.Pod, which showed Ne value above 100, 13 generations ago (Table 2). The effective population size (Ne) of African and Asian breeds, with the exception of SA.Dor, exceeded 100. In contrast, the SWE.Klo breed had the lowest Ne, measuring below 50. ±0.145).

Table 2*:* Means (sd) of within breed genetic diversity indices.

Breed code	***N***	**H_o ±sd**	**H_e ±sd**	**F**	**Ne_{13}**
ETH.Menz	34	0.346±0.16	0.352 ±0.14	0.174 ±0.04	211
KEN.Eaft	39	0.339±0.15	0.353±0.14	0.155±0.10	224
KEN.RedM	48	0.336±0.15	0.345±0.15	0.168±0.07	162
MAR	29	0.366±0.13	0.389±0.12	0.080±0.11	176
MARG	29	0.380±0.14	0.385±0.12	0.038±0.02	245
SA.Dor	19	0.367±0.17	0.364±0.14	0.125±0.04	71
CHN.Tib	36	0.343±0.14	0.363±0.14	0.172±0.05	164
IR.Afs	33	0.373±0.15	0.366±0.14	0.089±0.03	104
IR.Mog	34	0.376±0.15	0.376±0.13	0.067±0.02	154
BIH.DPram	23	0.383±0.13	0.375±0.13	0.034±0.05	118
BIH.HPram	23	0.383±0.16	0.376 ± 0.13	0.091±0.04	88
BIH.KPram	30	0.368±0.15	0.369±0.14	0.058±0.03	92
GRC.Bout	28	0.382±0.15	0.381±0.13	0.056±0.06	108
GRC.Chios	28	0.348±0.16	0.350±0.15	0.161±0.03	96
HUN.Dor	16	0.365±0.17	0.359±0.14	0.197±0.06	45
HUN.Dor.W	16	0.372±0.17	0.367±0.14	0.178±0.06	42
HUN.IldeFr	27	0.376±0.15	0.374±0.13	0.073±0.05	102
HUN.Mer	28	0.377±0.15	0.376±0.13	0.067±0.02	132
HUN.Rac	27	0.363±0.15	0.369±0.14	0.122±0.02	95
HUN.Suff	19	0.364±0.16	0.375±0.13	0.109±0.05	70
HUN.Tetra	23	0.392±0.16	0.384±0.13	0.033±0.02	88
HUN.Tsig	25	0.359±0.16	0.371±0.14	0.115±0.08	81
POL.Alek	23	0.368±0.16	0.364±0.14	0.115±0.05	59
POL.Pod	21	0.390±0.15	0.387±0.13	0.032±0.02	101
ROU.Rtsig	25	0.392±0.15	0.388±0.12	0.022±0.03	107
ROU.Turc	28	0.385±0.15	0.384±0.12	0.039±0.02	145
SRB.LPram	24	0.350±0.15	0.369±0.13	0.138±0.07	87
SRB.ZPram	16	0.385±0.16	0.393±0.12	0.046±0.05	85
SWE.Dal	18	0.309±0.18	0.331±0.16	0.362±0.12	40
SWE.Got	19	0.351±0.17	0.362±0.15	0.193±0.02	78
SWE.Klo	19	0.342±0.17	0.362±0.15	0.327±0.07	33
LTU.Bface	25	0.378±0.16	0.365±0.13	0.253±0.05	30

Assorted: individuals came from different breeds; N: Number of samples after QC, Ne_{13}- effective population size at 13 generations ago.

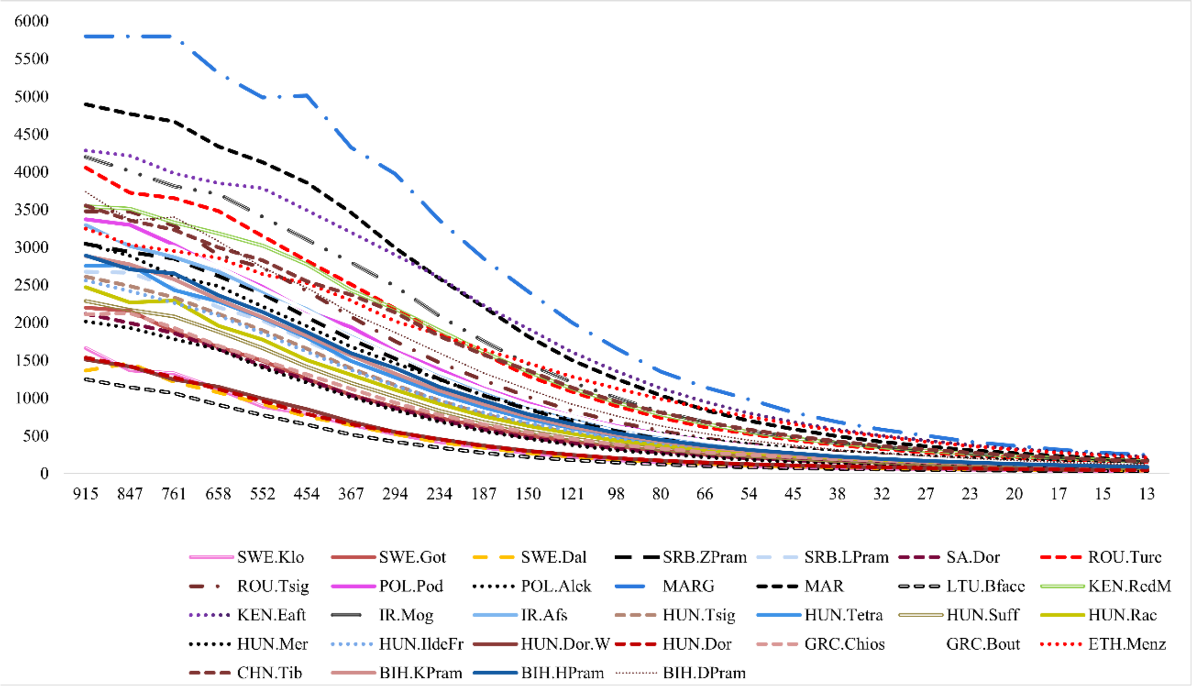


Figure 1: The trend of effective population size through generations within a population.
x-axis: number of generations; y-axis: Population size

3.1.2. Inter-breed genetic diversity and relationship

The AMOVA analysis revealed that a significant proportion of genetic variety, specifically over 91%, was found within the breed. Conversely, only a small proportion, around 8% ($p < 0.001$), was attributed to genetic variation across different breeds (Table 3).

Table 3: Genetic variation between and within the 32 sheep breeds.

Source of variation	d.f	Sum of squares	Variance components	Percentage of variation
Among populations	31	1033943.114	527.51012 Va	8.52
Within populations	1652	9357256.687	5664.19896 Vb	91.48
Total	1683	10391199.801	6191.70909	
Fixation index	0.08520, p-value < 0.001			

AMOVA was computed using distance matrix based on pairwise differences (FST)
Va- variance components among populations
Vb- Variance component between populations

The first two principal components (PCs) accounted for over 50% of the genetic variation seen in all breeds (Figure 2A), whereas the first three PCs together explained more than 64% of the variance (Figure 2B). A strong genetic proximity was seen among sheep breeds originating from the same continent, particularly those inhabiting the same climatic zone. PC’s 1 and 2 (Figure 2A) effectively differentiated breeds based on continental borders and climatological factors. The East African breeds (Tropical climate) from Kenya and

Ethiopia exhibited clustering, as did the North African breeds from Morocco and other Maghreb regions. Similarly, the Asian originating breeds formed a distinct cluster. Among the European originating breeds, all except GRC.Chios exhibited correlated genomes, with HUN.Dor.W and HUN.Dor forming a separate cluster, and HUN.Dor.W and HUN.Dor clustering together with SA.Dor. With the exception of GRC.Chios, all breeds were positioned in close proximity to or along PC3, as seen in Figure 2B. Surprisingly, Asian, North African and GRC.Chios clustered together along PCA 3 Figure 2B.

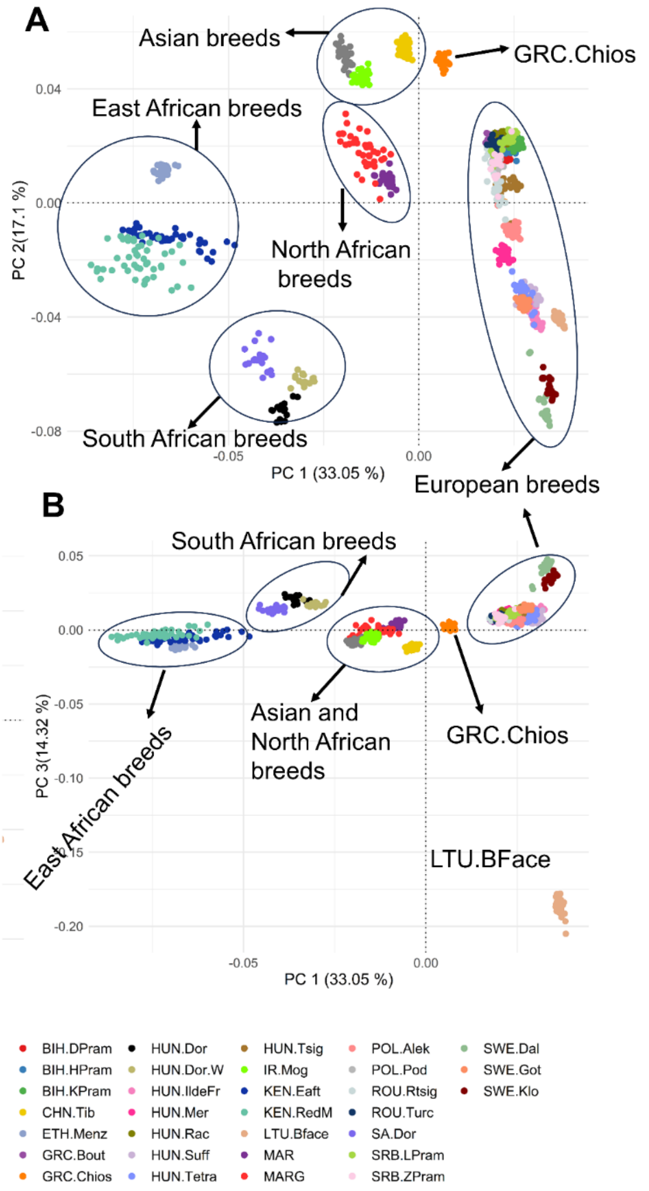


Figure 2: Principal components 1 to 3 for 32 native sheep breeds.

A: PC1 vs PC2, B: PC1 vs PC3

The phylogenetic tree (Figure 3) exhibited three primary branches, signifying three distinct patterns of breed divergence. Branch A comprises mostly African breeds together with HUN. Dor.W and HUN.Dor, as well as breeds originating from Asia and Greece. Branches B and C include all other European breeds. The sub-branches emanating from main branches linked closely related breeds that either originated from nearby geographical locations or have similar names, albeit coming from different nations. The phylogenetic tree showed a close grouping of all the outgroup breeds.

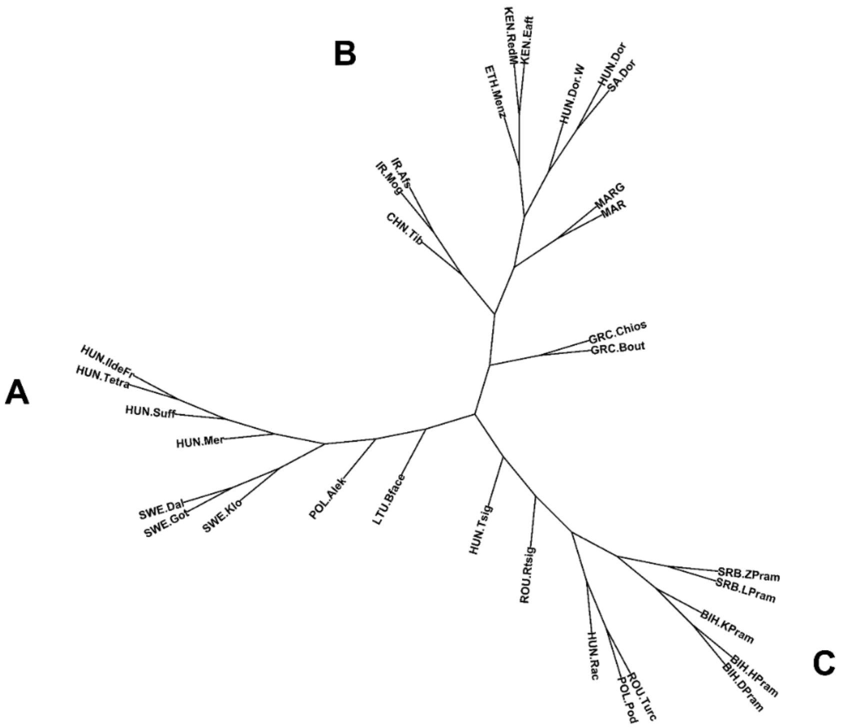


Figure 3: Phylogenetic tree showing the genetic linkages between the studied sheep breeds.

The ancestry study of the 32 breeds conducted using admixture software showed that the most probable number of clusters 25 based on the cross-validation entropy (CV; Figure 4A). It was observed that at K= 25, European breeds, which are thought to have adapted to the continental environment, had a significant level of admixture when compared to breeds from Northern Europe, Asia, and Africa (Figure 4B). Several breeds coming from Africa and Asia had discernible genetic origins at K=25, with the exception of KEN.RedM and SA. At the value of K = 32, which represents the number of breeds examined, it was seen that European breeds still had notable amounts of genetic admixture, whereas African and Asian breeds demonstrated distinct genetic ancestry.

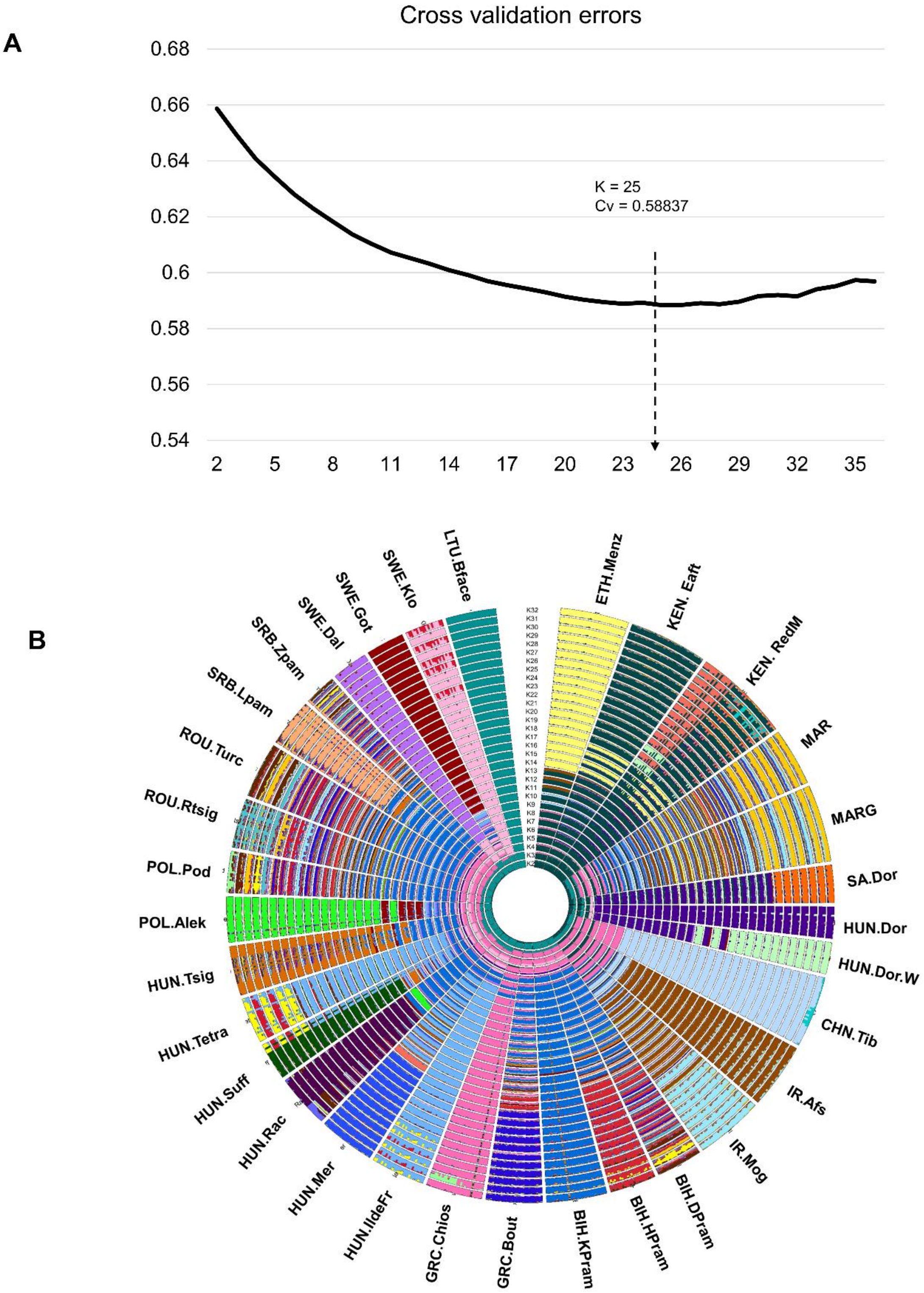


Figure 4: Ancestry member coefficient for each sheep breed samples.

A: Cross-entropy validation, B: ancestry membership from K2 to K32

3.2. Patterns of ROH and Froh genome per native sheep breed

Table 4 shows an overview of the average number of runs per class, the total counts of runs per class, and the Froh genome. In general, a discernible disparity was seen among breeds in terms of the average and count of runs in each class. The average number of runs per class within the 0-6 class varied from 1.837 for MARG to 3.448 for LTU.B face. In the class range of 12-24, the mean number of runs ranged from 7.713 for SWE.Klo to 8.9-3 for ROU.Tsig. In the class of 24-48 MB, the highest mean value was observed to be 35.540 (MAR). Conversely, in the class of >48 , the highest mean value recorded was 97.417 (GRC.Bout). The latter two classes had some breeds without mean runs. Similarly, there was significant variation in the counts of runs seen among different breeds. In the counts for class 0-6 varied from 3 (MARG) to 2244 (SWE.Dal).For class 6-12, the counts ranged from 6 (SWE.Got) to 513 (LTU.Bface). The classes 12-24, 24-48, and >48 had the highest counts of 345 (LTU.Bface), 114 (LTU.Bface), and 23 (KEN.EAFT), respectively while some breeds did not register runs in these last three classes. All the breeds registered a considerable level of Froh_genome, which ranged between 0.001 (MARG) and 0.262 (LTU.Bface).

3.3. Climate change adaptation selection signatures

For the purpose of this research, six pairings of population were selected based on the climatic zones they are adapted to. These pairs include tropical against Africa's Mediterranean populations, tropical against continental populations, tropical against Nordic populations, African Mediterranean against continental populations, African Mediterranean against Nordic populations, and continental versus Nordic adapted breeds. To clarify, the population group of "African Mediterranean" refers to populations in North Africa, "Nordic populations" pertains to sheep in North European region, "tropical population" refers specifically to East African populations, and "continental populations" include all other European populations except the Nordic ones.

Generally, the distribution of p-values and FST scores throughout the genome exhibited a similar pattern, with a few minor peaks seen in some pairs (Figures 5 and 6). However, it is worth noting that chromosome 3 had distinct deviations in the XP-EHH analysis (Figure 5), namely in the comparison between Tropical vs Nordic and Continental populations vs Nordic populations.

Table 4: Mean ROH, Froh (±sd) and counts of ROH per class.

Breed	Mean ROH per class					Froh_genome	Number of ROHs per class				
	0-6	6-12	12-24	24-48	> 48		0-6	6-12	12-24	24-48	> 48
BIH.DPram	2.652	8.238	16.222	34.213	51.078	0.026±0.03	265	25	20	9	1
BIH.HPram	2.672	8.199	18.163	30.060	77.469	0.036±0.03	348	38	21	12	3
BIH.KPram	3.083	8.482	16.598	31.317	58.405	0.078±0.04	522	171	103	41	2
CHN.Tib	2.767	7.479	NA	NA	NA	0.052±0.04	1460	122	NA	NA	NA
ETH.Menz	2.277	8.037	17.439	33.449	62.490	0.051±0.04	1490	30	17	10	5
GRC.Bout	2.969	8.609	16.247	32.091	97.417	0.047±0.06	369	84	49	19	3
GRC.Chios	3.004	8.179	15.584	33.868	NA	0.146±0.03	2113	324	77	18	NA
HUN.Dor	3.107	8.154	16.060	30.268	60.866	0.198±0.06	1277	274	79	18	6
HUN.Dor.W	3.113	8.294	16.180	32.912	65.658	0.186±0.06	1047	199	94	24	10
HUN.IldeFr	2.897	8.244	16.533	33.247	55.285	0.088±0.05	1153	179	50	17	1
HUN.Mer	2.811	7.950	16.196	30.454	NA	0.062±0.02	1118	115	19	7	NA
HUN.Rac	3.029	8.257	16.029	32.575	52.763	0.105±0.02	1148	246	88	16	2
HUN.Suff	3.003	8.185	16.917	31.346	50.786	0.114±0.05	755	136	76	31	2
HUN.Tetra	2.873	8.323	16.284	28.947	NA	0.044±0.02	503	66	29	7	NA
HUN.Tsig	3.214	8.465	16.607	32.352	61.885	0.117±0.08	690	247	104	37	8
IR.Afs	2.916	7.626	NA	NA	NA	0.027±0.02	653	58	NA	NA	NA
IR.Mog	2.769	7.627	NA	NA	NA	0.012±0.02	323	22	NA	NA	NA
KEN.Eaft	2.336	8.509	17.473	34.756	63.707	0.064±0.10	968	45	54	44	23
KEN.RedM	2.426	8.556	16.763	32.750	57.930	0.078±0.08	1331	194	137	60	13
LTU.Bface	3.448	8.578	16.624	32.037	59.080	0.262±0.05	789	513	346	114	14
MAR	2.758	8.634	16.991	35.540	62.400	0.064±0.11	214	55	65	45	18
MARG	1.837	NA	NA	NA	NA	0.001±0.00	3	NA	NA	NA	NA
POL.Alek	3.316	8.620	16.571	30.925	54.637	0.121±0.05	605	233	109	40	6
POL.Pod	2.666	8.381	15.578	24.712	53.070	0.023±0.02	294	36	7	2	1
ROU.Rtsig	2.816	8.903	18.276	29.609	50.988	0.025±0.03	247	45	17	5	2
ROU.Turc	2.579	8.805	16.556	29.347	53.601	0.022±0.02	309	47	16	3	1

SA.Dor	3.003	8.176	15.874	29.377	66.443	0.118±0.04	1032	187	57	9	2
SRB.LPram	3.170	8.544	16.656	30.753	56.252	0.125±0.08	550	190	133	62	8
SRB.ZPram	2.781	8.780	14.903	32.286	55.578	0.036±0.05	175	46	20	7	2
SWE.Dal	2.633	7.884	NA	NA	NA	0.130±0.05	2244	39	NA	NA	NA
SWE.Got	2.436	7.713	NA	NA	NA	0.047±0.01	958	6	NA	NA	NA
SWE.Klo	2.653	7.821	NA	NA	NA	0.118±0.03	2130	39	NA	NA	NA

NA= not available.

A threshold of the top 0.01% of the log p-values or FST scores yielded a substantial number of SNPs that were indicative of selection for every pair of populations. The majority of signatures were specific to each population pair. The FST-based strategy detected a greater number of signatures compared to the XP-EHH method for each population pair. Both methodologies effectively identified a limited number of shared characteristics for each pair.

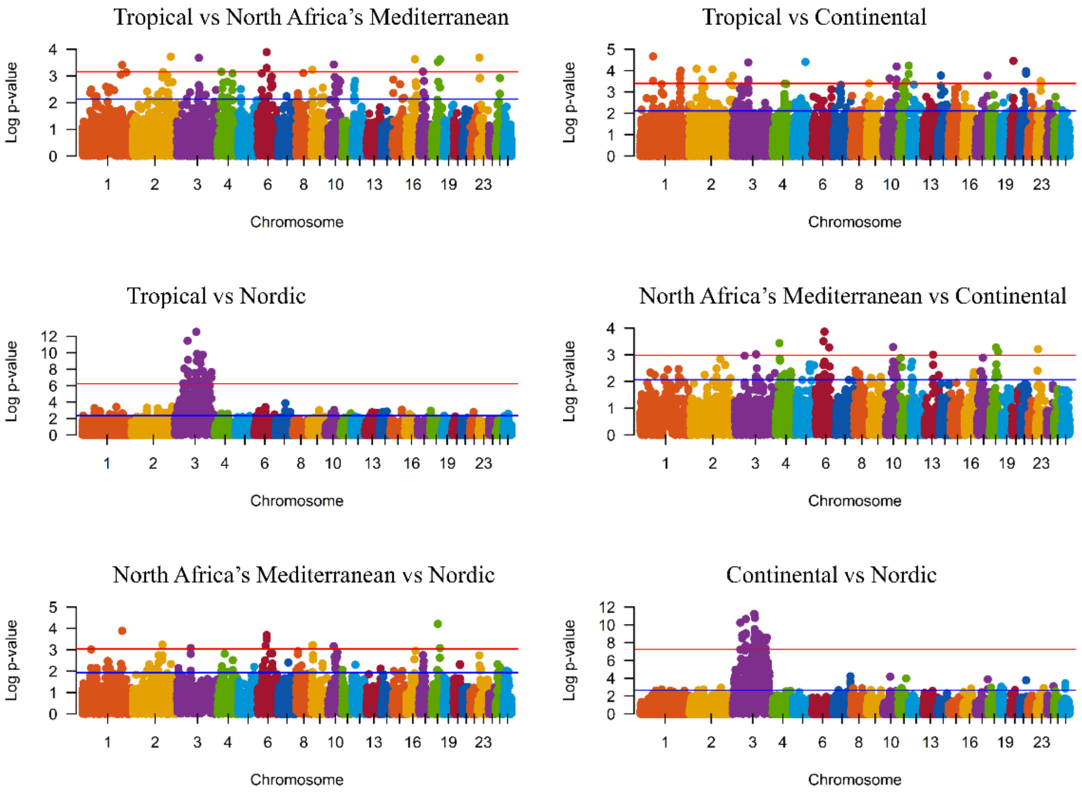


Figure 5: XP-EHH Manhattan plot showing log p-value and SNP distribution for each climate-based population pair.

Red line- suggestive threshold line at 0.01% while blue line is a genome wide line at 0.1%.

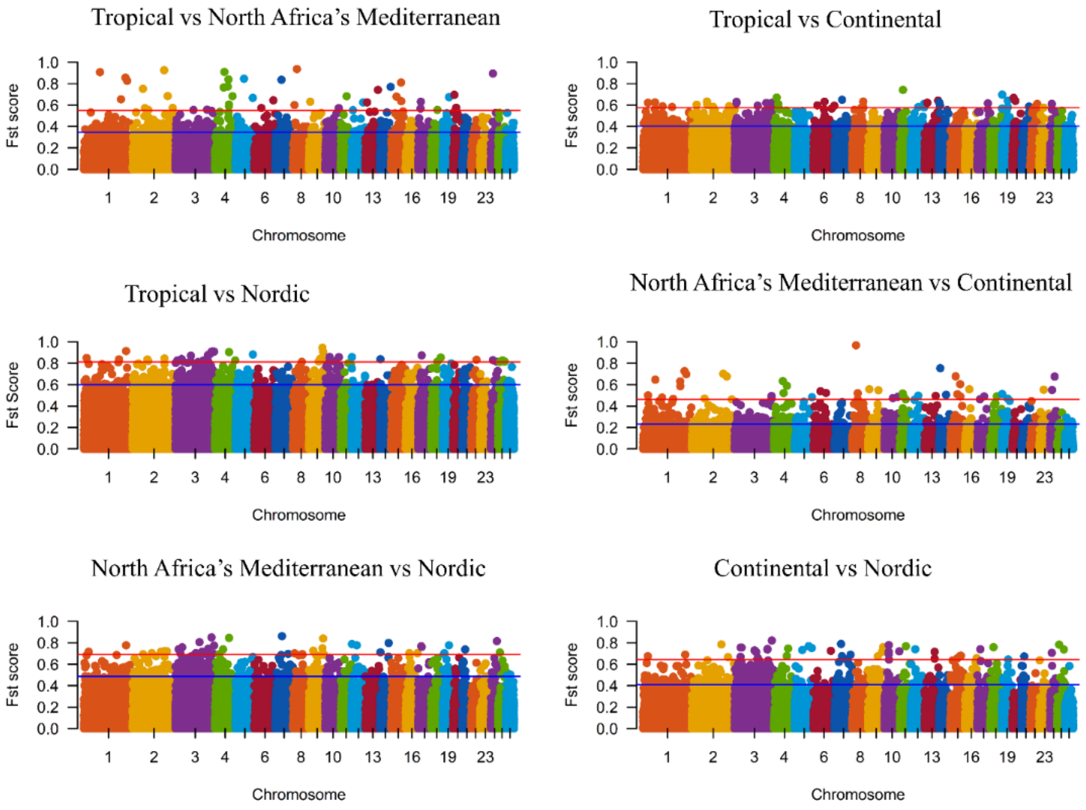


Figure 6: Manhattan plot showing FST score and SNP distribution for each climate-based population pair.

Red line- suggestive threshold line at 0.01% while blue line is a genome wide line at 0.1%.

3.4. Gene x environment association (GEA)

As a result of quality control steps of the environmental variables, only six climatic variables remained for the association analysis. The six variables that were included in the analysis were bio_2, bio_7, bio_8, bio_13, bio_18, and bio_19. Hence, this part will give the post-processing outcomes of both the LFMM (ridge) and RDA statistical methodologies in relation to the five bio climatic variables. The LFMM method resulted in the identification of 133 SNPs, while the RDA approach provided a larger set of 360 SNPs. The analysis revealed a strong association between bio_2 and many SNPs using the RDA method. Conversely, the LFMM approach identified a greater number of SNPs in bio_8 and bio_19 compared to RDA (Table 5). Both the LFMM and RDA techniques yielded a total of 28 common outlier SNPs cutting across all climatic variables.

Table 5*:* Summary of number of SNPs significantly associated with each bio climatic variable.

Bio-climatic variable	LFMM	RDA
bio_2	21	127
bio_7	2	66
bio_8	18	3
bio_13	48	59
bio_18	17	102
bio_19	27	3
Total	133	360

The Gene Ontology primarily prioritized the analysis of shared SNPs identified by both statistical approaches. This analysis resulted in the mapping of around 11 genes (Table 6). In general, both methods exhibited associations between SNPs and environmental variables that were largely similar, with a few exceptions. For instance, SNP DU464270_354.1 showed a discrepancy in its associations, as RDA linked it to bio_19 while LFMM associated it with bio_13. Similarly, SNP OAR3_231138606.1 also showed divergent associations, with RDA associating it with bio_7 and LFMM associating it with bio_18, similar to SNP s10898.1. A total of 11 distinct genes that are strongly believed to be responsible for driving adaptation were identified.

Table 6*:* Common outlier SNPs identified by both RDA and LFMM, their p values for each method, bioclimatic variables linked to the SNPs, and genes mapped near the SNPs.

SNP	Chr	BP	pRDA	pLFMM	bioRDA	bioLFMM	Gene name
s40694.1	1	285932646	0.010297	0.054294	bio_2	bio_2	-
OAR2_192491503.1	2	192491503	-0.01881	0.022217	bio_18	bio_18	STAT1
DU464270_354.1	3	132614430	-0.0125	0.042746	bio_19	bio_13	ATP5MC2
OAR3_138331159.1	3	138331159	-0.06545	-0.03431	bio_18	bio_18	TARBP2
OAR3_139371620.1	3	139371620	-0.00538	-0.02582	bio_18	bio_18	-
OAR3_231138606.1	3	231138606	-0.03065	-0.01825	bio_7	bio_18	-
OAR3_43427639.1	3	43427639	0.00301	-0.00955	bio_13	bio_13	-
s09582.1	3	85831354	0.012189	0.049759	bio_13	bio_13	-
s51845.1	3	225905945	-0.00599	0.022061	bio_13	bio_13	-
OAR6_97162589.1	6	97162589	0.014064	0.006435	bio_13	bio_13	SCD5
s21552.1	6	103818054	0.02642	0.031675	bio_13	bio_13	-
OAR7_97127242.1	7	97127242	-0.00446	0.064122	bio_13	bio_13	GSTA1
OAR7_97378846.1	7	97378846	0.060344	0.047132	bio_13	bio_13	-
s40182.1	7	88607719	0.008254	-0.00795	bio_2	bio_2	-
s52058.1	7	18561594	0.023605	0.012786	bio_13	bio_13	-
OAR11_4962381.1	11	4962381	0.005373	0.008149	bio_2	bio_2	HLF
OAR13_51727898.1	13	51727898	-0.00554	0.044412	bio_13	bio_13	-
OAR13_67246527.1	13	67246527	-0.01797	-0.03236	bio_18	bio_18,bio_2	ADIG
s66432.1	13	67952785	0.021583	0.010252	bio_18	bio_18	-
OAR16_18252778.1	16	18252778	0.005143	-0.0086	bio_18	bio_13,bio_18	-
s10898.1	18	70382783	-0.04211	-0.01173	bio_7	bio_18	-
s11142.1	18	19970738	0.024238	-0.0116	bio_13	bio_13	RHCG
s31152.1	18	19734572	-0.05197	4.98E-05	bio_18	bio_18	PLIN1
DU177621_558.1	19	56974832	0.000762	-0.02124	bio_2	bio_2	-
OAR19_15621948.1	19	15621948	-0.01489	0.013746	bio_18	bio_18	ABHD5
s08031.1	19	59803760	0.034551	-0.00581	bio_2	bio_2	-
s28575.1	21	46857187	-0.03101	-0.0279	bio_2	bio_2	-
OARX_63571789.1	27	63571789	0.069982	0.002946	bio_18	bio_18	-

4. CONCLUSION

The present comprehensive investigation has shown that indigenous sheep breeds have a high degree of genetic diversity, but with a rabid decline in diversity levels. The observed decline in effective population size across all analyzed breeds is a cause for concern. Notably, none of the breeds exhibited an Ne above 250, indicating that all breeds are possibly at risk of extinction in the face of climate change. The first two principal components explained more than 50% of the genomic variance seen across all the breeds, indicating the presence of unique genomic structures in indigenous sheep breeds. This differentiation might be attributed to factors such as geographical isolation or the influence of climatic zones on the genomes. The ancestral clustering indicated that despite being genetically diverse, European originating native breeds showed high levels of admixture at both K=25 and K=32. Consistent admixture erodes the uniqueness and integrity of specific native breeds. The categorization of local breeds based on climate resulted in the identification of over 400 distinct genes, which are mostly exclusive to each climate-based pair of populations. This finding suggests that each breed group has a unique capability for adaptability, hence can't replace each other. The majority of genes exhibit immune-related activities, so emphasizing the need to selectively breed animals that possess disease resistance traits in response to the current climate change circumstances. In addition, the distinctive pattern and distribution of ROHs further underscored the distinctiveness of the genome of each breed. Furthermore, studies on gene x environment associations have shown that precipitation and temperature are significant climatic factors that play a crucial role in shaping the genetic makeup of a breed, leading to adaptive changes. All the genes identified under GEA are either responsible for immunity, metabolism, or productivity. This study supports the assertion that indigenous sheep breeds continue to possess significant genetic diversity, but with a rapid decline which may be attributed to inadequate genetic management practices or haphazard crossbreeding. The study findings also indicate that indigenous sheep breeds possess a distinctive capacity to adapt to climatic climate change, making them valuable for the purpose of selectively breeding for climate change adaptation. This, in turn, would enhance the resilience and durability of sheep breeds. In addition, results of this study can be used to enhance breed conservation programs for mitigation of climate change effects.

5. NEW SCIENTIFIC RESULTS

Genetic characterization of indigenous sheep breeds to determine within- and between populations genetic diversity.

i. Despite high levels of within-breed genetic diversity, the 32 native sheep breeds are at risk of extinction due to a decline in effective population size. Furthermore, genetic clustering patterns suggest continental demarcations refined by climatic zones whereas European breeds are highly admixed.

Detection of patterns of run of homozygosity (ROH) and genomic based inbreeding coefficient (Froh).

i. The distribution of ROH segments across all breeds revealed a predominance of short segments (0-6 and 6-24 Mb) accompanied by a scarcity of long segments (>24 Mb). This pattern is indicative of genomic inbreeding primarily originating from inbreeding events that transpired several generations ago.

Identification of signatures of selection underpinning adaptation in different climatic regions.

i. A comprehensive analysis of climate-based pairwise groups revealed over 400 distinct signatures of climate change adaptation, with only three genes (*CFTR, CTTNBP2,* and *CBX5*) consistently appearing across multiple pairings.

Gene x environment association (GEA) to explore the complex interplay between genes and climatic variables in shaping adaptation traits.

i. A gene-environment association study revealed that temperature and rainfall are crucial climatic factors driving native sheep adaptation, with ten specific genes (*STAT1, ATP5MC2, TARBP2, SCD5, GSTA1, HLF, ADIG, RHCG, PLIN1*, and *ABHD5*) exhibiting significant associations with these climatic factors.

6. PRACTICAL USABILITY OF RESULTS

The findings of this study provide valuable insights that can be utilized to enhance conservation efforts for endangered breeds and inform decisions regarding conservation priorities for certain breeds. According to the current findings, it may be inferred that African and Asian breeds, whose risk status is currently unclear, may now be classified as endangered due to the fact that all of them exhibited a Ne value below 250.

Other more specific practical usability of these results are as follows:

i. High within breed genetic diversity can be exploited to enhance within breed selection.
ii. High inbreeding information can be used to genetically manage native sheep breeds to minimize chances of mating closely related individuals.
iii. The findings that demonstrate a rapid decrease in Ne, as well as the persistently low Ne numbers, have valuable implications for the development of a conservation program aimed at preserving sheep genetic resources.
iv. The results indicating a significant degree of breed admixture may be used by breeding and/or conservation farms to intentionally facilitate the restoration of pure breeds. This can be achieved via the rigorous selection of individuals that are either purebred or exhibit characteristics consistent with the breed's original traits, for the purpose of breeding future generations.
v. Information on ROH pattern may be used to improve breed molecular-based management to enhance genetic diversity, adaptation to specific climatic zones as well as reduce Froh_genome.
vi. The unique capacity for adaptability shown by paired population groupings in different climatic zones suggests that each breed has distinct adaptive characteristics, making them irreplaceable. Consequently, it is imperative to undertake conservation measures for endangered breeds. Nevertheless, these indigenous breeds may be used to enhance the adaptability of specialized breeds by introducing genes associated with adaptation into them.
vii. Understanding the climatic variables driving sheep adaptability offers an opportunity to develop proper breeding programs to genetically counter the effects of climate change to sheep industry.

viii. The research has identified many genes, including *STAT1, ATP5MC2, TARBP2, SCD5, GSTA1, HLF, ADIG, RHCG, PLIN1, ABHD5*, and others, which have the potential to be included in marker assisted selection which holds a promise for improving breeding strategies aimed at enhancing the resilience of sheep breeds to climate change.

7. REFERENCES LIST

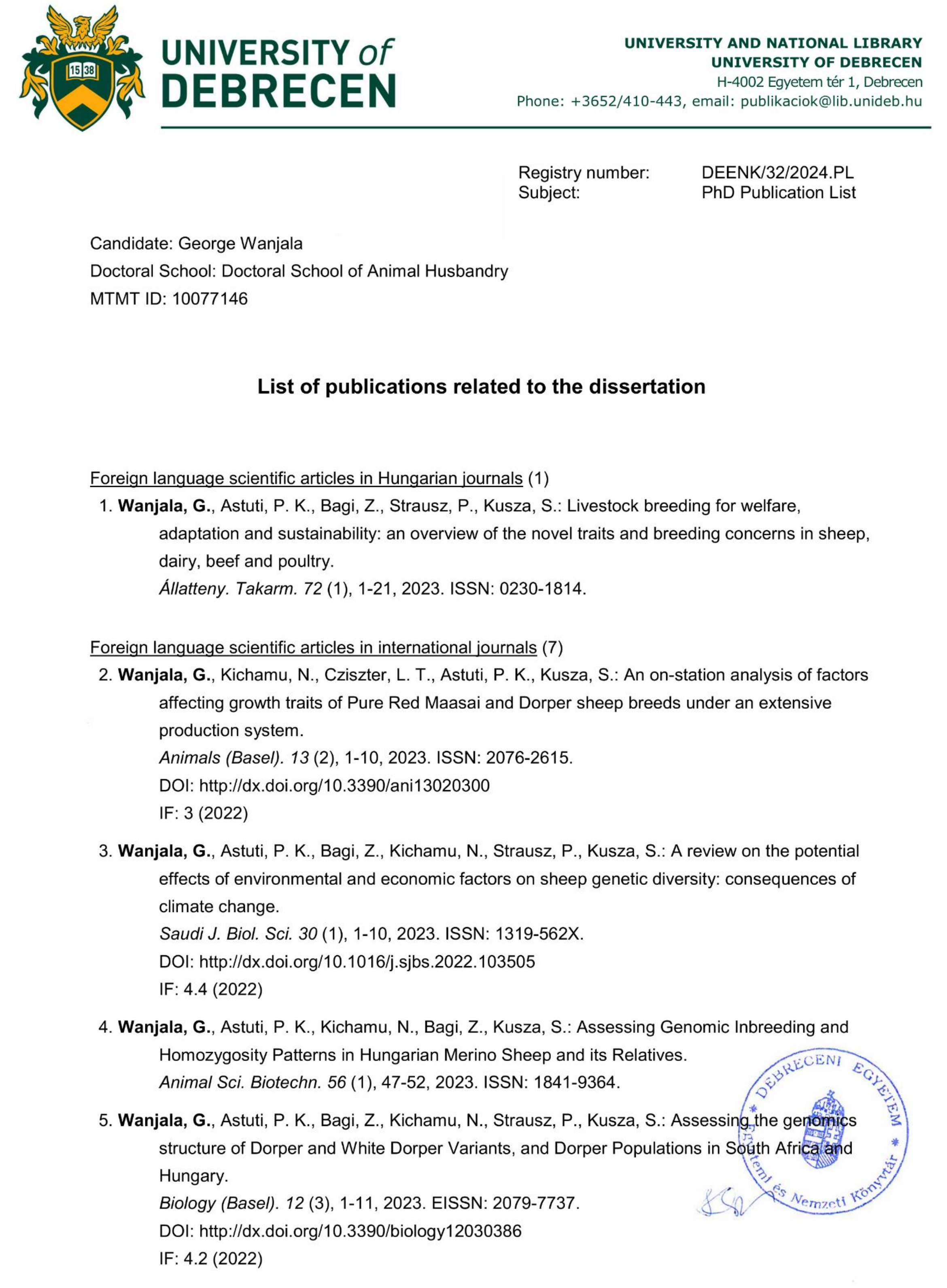
1. ALEXANDER, D. H.–SHRINGARPURE, S. S.– NOVEMBRE, J.– LANGE, K. (2015): Admixture 1.3 software manual. Los Angeles: UCLA Human Genetics Software Distribution.
2. ANDERSON, T. W. (1962). An introduction to multivariate statistical analysis. Wiley New York.
3. BARBATO, M. – OROZCO-TERWENGEL, P. – TAPIO, M. – BRUFORD, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics, 6, 109.
4. BEAUMONT, M. A. (2005). Adaptation and speciation: what can Fst tell us? Trends in Ecology & Evolution, 20 (8), 435–440.
5. BISCARINI, F. – COZZI, P. – GASPA, G. – MARRAS, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes.
6. CAYE, K. – JUMENTIER, B. – LEPEULE, J. – FRANÇOIS, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. Molecular Biology and Evolution, 36 (4), 852–860.
7. CHANG, C. C. – CHOW, C. C. – TELLIER, L. C. A. M. – VATTIKUTI, S. – PURCELL, S. M. – LEE, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4(1), s13742-015-0047–0048.
8. CHERUIYOT, E. K. – BETT, R. C. – AMIMO, J. O. – ZHANG, Y. – MRODE, R. – MUJIBI, F. D. N. (2018). Signatures of selection in admixed dairy cattle in Tanzania. Frontiers in Genetics, 9, 607.
9. CLARK, E. – BUSH, S. – SUMMERS, K. – HUME, D. – WHITELAW, C. – WATSON, M. – FARQUHAR, I. – ARCHIBALD, A. (2017). Functional Annotation of the Sheep (Ovis aries) Genome. Plant and Animal Genomes Conference XXV.
10. DASH, H. R. – SHRIVASTAVA, P. – DAS, S. (2020). Reliable Use of Whatman^{TM} FTA^{TM} Cards for One-Step Collection and Isolation of DNA. In Principles and practices of DNA analysis: a laboratory manual for forensic DNA typing (109–115). Springer.
11. DIXON, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14 (6), 927–930.
12. DURINCK, S. – MOREAU, Y. –KASPRZYK, A. – DAVIS, S. – DE MOOR, B. –BRAZMA, A. – HUBER, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics, 21 (16), 3439–3440.

13. DZOMBA, E. F. – CHIMONYO, M. – PIERNEEF, R. – MUCHADEYI, F. C. (2021). Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50k data. BMC Genomics, 22 (1).
14. EDEA, Z. – DADI, H. – DESSIE, T. – KIM, K.-S. (2019). Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. Genes & Genomics, 41, 973–981.
15. EXCOFFIER, L. – LAVAL, G. – SCHNEIDER, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 1, 117693430500100000.
16. FLORI, L. – MOAZAMI-GOUDARZI, K. – ALARY, V. – ARABA, A. – BOUJENANE, I. – BOUSHABA, N. – CASABIANCA, F. – CASU, S. – CIAMPOLINI, R. – COEUR D'ACIER, A. – COQUELLE, C. – DELGADO, J.V. – EL-BELTAGI, A. – HADJIPAVLOU, G. – JOUSSELIN, E. – LANDI, V. – LAUVIE, A. – LECOMTE, P. – LIGDA, C. et al. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. Molecular Ecology, 28 (5), 1009–1029.
17. FORESTER, B. R. – LASKY, J. R. – WAGNER, H. H. – URBAN, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology, 27 (9), 2215–2233.
18. GAUTIER, M. – VITALIS, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics, 28 (8), 1176–1177.
19. GOITOM, S – GICHEHA, M. G. – NJONGE, F. K. – KIPLANGAT, N. (2021). Landscape genomics and selection signatures of local adaptation of Eritrean indigenous cattle along environmental gradients. Tropical Animal Health and Production, 53 (4), 1–8.
20. HSIAO, K. – LIN, H. – PAN, H. – LI, T. – CHEN, S. – JOU, S. – CHIU, Y. – WU, M. – LIN, C. – LI, S. (1999). Application of FTA® sample collection and DNA purification system on the determination of CTG trinucleotide repeat size by PCR-based southern blotting. Journal of Clinical Laboratory Analysis, 13 (4), 188–193.
21. NEOGEN. https://www.neogen.com. (2023 July 20)
22. HUBBARD, T. – BARKER, D. – BIRNEY, E. – CAMERON, G. – CHEN, Y. – CLARK, L. – COX, T. – CUFF, J. – CURWEN, V. – DOWN, T. – DURBIN, R. – EYRAS, E. – GILBERT, J. – HAMMOND, M. – HUMINIECKI, L. – KASPRZYK, A. – LEHVASLAIHO, H. – LIJNZAAD, P. – MELSOPP, C. – … CLAMP, M. (2002). The Ensembl genome database project. Nucleic Acids Research, 30 (1), 38–41.
23. MAĆKIEWICZ, A. – RATAJCZAK, W. (1993). Principal components analysis (PCA). Computers & Geosciences, 19(3), 303–342.

24. MANEL, S. – SCHWARTZ, M. K. – LUIKART, G. – TABERLET, P. (2003). Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution, 18 (4), 189–197.
25. MEIRMANS, P. G. (2006). USING THE AMOVA FRAMEWORK TO ESTIMATE A STANDARDIZED GENETIC DIFFERENTIATION MEASURE. Evolution, 60(11), 2399–2402.
26. MENGONI, A. – BAZZICALUPO, M. (2002). The statistical treatment of data and the analysis of molecular variance (AMOVA) in molecular microbial ecology. Annals of Microbiology, 52 (2), 95–102.
27. MILANESI, M. – CAPOMACCIO, S. – VAJANA, E. – BOMBA, L. – GARCIA, J. F. – AJMONE-MARSAN, P. – COLLI, L. (2017). BITE: an R package for biodiversity analyses. BioRxiv, 181610.
28. PAIM, T. DO P. – ALVES DOS SANTOS, C. – FARIA, D. A. – DE, PAIVA, S. R. – MCMANUS, C. (2022). Genomic selection signatures in Brazilian sheep breeds reared in a tropical environment. Livestock Science, 258, 104865.
29. PEMBLETON, L. W. – COGAN, N. O. I. – FORSTER, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13 (5), 946–952.
30. PURCELL, S. – NEALE, B. – TODD-BROWN, K. – THOMAS, L. – FERREIRA, M. A. R. – BENDER, D. – MALLER, J. – SKLAR, P. – DE BAKKER, P. I. W. – DALY, M. J. – SHAM, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics, 81 (3), 559–575.
31. SERRANITO, B. – CAVALAZZI, M. – VIDAL, P. – TAURISSON-MOURET, D. – CIANI, E. – BAL, M. – ROUVELLAC, E. – SERVIN, B. – MORENO-ROMIEUX, C. – TOSSER-KLOPP, G. – HALL, S. J. G. – LENSTRA, J. A. – POMPANON, F. – BENJELLOUN, B. – DA SILVA, A. (2021). Local adaptations of Mediterranean sheep and goats through an integrative approach. Scientific Reports, 11(1), 21363.
32. SVED, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical Population Biology, *2* (2), 125–141.
33. THOMSON, J. M. – EASTON, A. N. N. C. – Faragher, E. B. (1983). The use of Vacutainer tubes for collection and storage of blood for coagulation testing. Clinical & Laboratory Haematology, 5 (4), 413–421.
34. TURNER, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. BioRxiv, 5165.

35. WHITLEY, D. C. – FORD, M. G. – LIVINGSTONE, D. J. (2000). Unsupervised forward selection: a method for eliminating redundant variables. Journal of Chemical Information and Computer Sciences, 40 (5), 1160–1168.
36. WICKHAM, H. – WICKHAM, M. H. (2017). Package tidyverse. Easily Install and Load the 'Tidyverse.
37. WIENER, P. – ROBERT, C. – AHBARA, A. – SALAVATI, M. – ABEBE, A. – KEBEDE, A., WRAGG, D. – FRIEDRICH, J. – VASOYA, D. – HUME, D. A. (2021). Whole-Genome Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations. Genome Biology and Evolution, 13 (3), evab014.
38. YUAN, Z. – LIU, E. – LIU, Z. – KIJAS, J. W. – ZHU, C. – HU, S. – MA, X. – ZHANG, L. – DU, L. – WANG, H. – WEI, C. (2017). Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. Animal Genetics, 48 (1), 55–66.

8. PUBLICATION LIST



UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number:	DEENK/32/2024.PL
Subject:	PhD Publication List

Candidate: George Wanjala
Doctoral School: Doctoral School of Animal Husbandry
MTMT ID: 10077146

List of publications related to the dissertation

Foreign language scientific articles in Hungarian journals (1)

1. **Wanjala, G.**, Astuti, P. K., Bagi, Z., Strausz, P., Kusza, S.: Livestock breeding for welfare, adaptation and sustainability: an overview of the novel traits and breeding concerns in sheep, dairy, beef and poultry.
   Állatteny. Takarm. 72 (1), 1-21, 2023. ISSN: 0230-1814.

Foreign language scientific articles in international journals (7)

2. **Wanjala, G.**, Kichamu, N., Cziszter, L. T., Astuti, P. K., Kusza, S.: An on-station analysis of factors affecting growth traits of Pure Red Maasai and Dorper sheep breeds under an extensive production system.
   Animals (Basel). 13 (2), 1-10, 2023. ISSN: 2076-2615.
   DOI: http://dx.doi.org/10.3390/ani13020300
   IF: 3 (2022)
3. **Wanjala, G.**, Astuti, P. K., Bagi, Z., Kichamu, N., Strausz, P., Kusza, S.: A review on the potential effects of environmental and economic factors on sheep genetic diversity: consequences of climate change.
   Saudi J. Biol. Sci. 30 (1), 1-10, 2023. ISSN: 1319-562X.
   DOI: http://dx.doi.org/10.1016/j.sjbs.2022.103505
   IF: 4.4 (2022)
4. **Wanjala, G.**, Astuti, P. K., Kichamu, N., Bagi, Z., Kusza, S.: Assessing Genomic Inbreeding and Homozygosity Patterns in Hungarian Merino Sheep and its Relatives.
   Animal Sci. Biotechn. 56 (1), 47-52, 2023. ISSN: 1841-9364.
5. **Wanjala, G.**, Astuti, P. K., Bagi, Z., Kichamu, N., Strausz, P., Kusza, S.: Assessing the genomics structure of Dorper and White Dorper Variants, and Dorper Populations in South Africa and Hungary.
   Biology (Basel). 12 (3), 1-11, 2023. EISSN: 2079-7737.
   DOI: http://dx.doi.org/10.3390/biology12030386
   IF: 4.2 (2022)

6. **Wanjala, G.**, Kichamu, N., Strausz, P., Astuti, P. K., Kusza, S.: On-station comparative analysis of reproductive and survival performance between Red Maasai, Dorper, and Merino sheep breeds.
   Animal. 17 (3), 1-6, 2023. ISSN: 1751-7311.
   DOI: http://dx.doi.org/10.1016/j.animal.2023.100715
   IF: 3.6 (2022)

7. **Wanjala, G.**, Bagi, Z., Kusza, S.: Meta-Analysis of Mitochondrial DNA Control Region Diversity to Shed Light on Phylogenetic Relationship and Demographic History of African Sheep (Ovis aries) Breeds.
   Biology (Basel). 10 (8), 1-12, 2021. EISSN: 2079-7737.
   DOI: http://dx.doi.org/10.3390/biology10080762
   IF: 5.168

8. **Wanjala, G.**, Bagi, Z., Jávor, A., Kusza, S.: Impacts of climate change on sheep genetic diversity: a review.
   Nat. Res. Sustain. Develop. 10 (2), 240-261, 2020. ISSN: 2066-6276.
   DOI: http://dx.doi.org/10.31924/nrsd.v10i2.058

<u>Foreign language abstracts</u> (2)

9. **Wanjala, G.**, Ilisiu, E., Gavojdian, D., Bagi, Z., Astuti, P. K., Kusza, S.: Genomic diversity and population structure of Hungarian and Romanian native sheep breeds.
   In: Biotechnology at the University of Debrecen - 2022 International Symposium : Abstract Book, Debreceni Egyetem, Printart-Press Kft., Debrecen, 68-69, 2022.

10. **Wanjala, G.**, Astuti, P. K., Kichamu, N., Kusza, S.: Eco-regional influence on genetic diversity, population structure, and signatures of selection of indigenous sheep in Kenya.
    In: Biotechnológia a Debreceni Egyetemen. Fókuszban az agrár- és élelmiszerbiotechnológia : Tudományos Szimpózium elődás összefoglalók, Debreceni Egyetem MÉK Agrár Genomikai és Biotechnológiai Központ, Debrecen, 46-47, 2021. ISBN: 9789634903512

List of other publications

<u>Foreign language scientific articles in Hungarian journals</u> (2)

11. Astuti, P. K., **Wanjala, G.**, Bagi, Z., Kusza, S.: Coping with climate change; is white sheep more favorable than black? = Szembenézni az éghajlatváltozással; kedvezőbb a fehér bárány a feketénél? : a review = irodalmi áttekintés.
    Állatteny. takarm. 71 (4), 270-282, 2022. ISSN: 0230-1814.

12. **Wanjala, G.**, Mihalik, B., Stéger, V., Kusza, S.: Lake Balaton as a geographical barrier for gene flow between wild boar (Sus scrofa) populations in Hungary.
    Acta Agronom. Óváriens. 58 (2), 36-55, 2017. ISSN: 1416-647X.

Foreign language scientific articles in international journals (4)

13. Balog, K., Mizeranschi, A. E., **Wanjala, G.**, Sipos, B., Kusza, S., Bagi, Z.: Application potential of chicken DNA chip in domestic pigeon species - preliminary results.
    Saudi J. Biol. Sci. 30 (3), 1-9, 2023. ISSN: 1319-562X.
    DOI: http://dx.doi.org/10.1016/j.sjbs.2023.103594
    IF: 4.4 (2022)

14. Astuti, P. K., Gavojdian, D., Ilie, D. E., **Wanjala, G.**, Monori, I., Bagi, Z., Kusza, S.: Genetic polymorphism in European and African sheep breeds reared in Hungary based on 48 SNPs associated with resistance to gastrointestinal parasite infection using KASP-PCR technique.
    Trop. Anim. Health Prod.. 55 (3), 1-12, 2023. ISSN: 0049-4747.
    IF: 1.7 (2022)

15. Astuti, P. K., Ilie, D. E., Gavojdian, D., **Wanjala, G.**, Badaoui, B., Ohran, H., Pasic-Juhas, E., Bagi, Z., Jávor, A., Kusza, S.: Validation of SNP markers for thermotolerance adaptation in Ovis aries adapted to different climatic regions using KASP-PCR technique.
    Sci. Rep. 12, 1-9, 2022. EISSN: 2045-2322.
    DOI: http://dx.doi.org/https://doi.org/10.1038/s41598-022-26909-1
    IF: 4.6

16. Mihalik, B., Kusza, S., Stéger, V., **Wanjala, G.**, Németh, Z.: Effects of natural and artificial barriers on the genetic diversity of game species: a review.
    Balk. J. Wildl. Res. 5 (1), 1-18, 2020. EISSN: 2335-0113.
    DOI: http://dx.doi.org/10.15679/bjwr.v5i1.61

Foreign language conference proceedings (3)

17. Fajardo, R., Astuti, P. K., Bagi, Z., Bodrogi, L., Pintér, T., Skoda, G., Ohran, H., **Wanjala, G.**, Kusza, S.: Continental vs. tropical breed: Immunity comparison under heat stress conditions utilizing qRT-PCR technique.
    Bio Web Conf. 8, 1-8, 2023. ISSN: 2273-1709.
    DOI: http://dx.doi.org/https://doi.org/10.1051/bioconf/20238005001

18. Astuti, P. K., Sárkány, P., **Wanjala, G.**, Kichamu, N., Kusza, S.: Genetics of coat colors and its role in climate change resiliency in sheep.
    In: Genetics and Applications: An Aspiring Interdisciplinary Journal of Genetic Research, Speciasl edition, Book of abstracts, Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Sarajevo, Bosnia and Herzegovina, 40, 2023, (eISSN 2566-431x)

19. Astuti, P. K., Fajardo, R., **Wanjala, G.**, Kichamu, N., Ohran, H., Badaoui, B., Bagi, Z., Kusza, S.: Heat Stress: Can Animals be Stressed but still be Healthy?
    Animal Sci. Biotechn. 56 (1), 194-202, 2023. ISSN: 1841-9364.

Foreign language abstracts (4)

20. Strausz, P., Latorcai, C., Triatmojo, A., Guntoro, B., **Wanjala, G.**, Kusza, S.: Actors in sharing adaptable results of genetics and management for developing sustainable livestock.
In: Book of abstract : Multidisciplinary Conference on Sustainable Development : 25-26 May 2023 : Section : Animal Resources Bioengineering, Faculty of Bioengineering of Animal Resources, Tinmisoara, 19-20, 2023, (ISSN 2821-4293)

21. Astuti, P. K., Fajardo, R., **Wanjala, G.**, Ohran, H., Badaoui, B., Bagi, Z., Kusza, S.: Heat stress: Can animals be stressed but still be healthy?
In: Book of abstract : Multidisciplinary Conference on Sustainable Development : 25-26 May 2023 : Section : Animal Resources Bioengineering, Faculty of Bioengineering of Animal Resources, Tinmisoara, 17, 2023, (ISSN 2821-4293)

22. Pius, L. O., **Wanjala, G.**, Strausz, P., Kusza, S.: A review on genetic diversity of local chicken population and their adaptation potential to climate change in tropical climate: a case study in East African countries.
In: ASD 2022 : Book of Abstracts. Eds.: Vlatka Čubrić Čurik et al, University of Zagreb Faculty of Agriculture, Zadar, 50, 2022. ISBN: 9789538276361

23. Astuti, P. K., **Wanjala, G.**, Kukovics, S., Kusza, S.: Polymorphism of genes associated with immune response in French Alpine and Saanen in Hungary.
In: 13th International Conference on Goats "ICG 2022": Book of Abstracts - Goats for the Future - Goat production in a changing environment. Eds.: Lucia Sepe, Carina Visser, Noemí Castro, Paula Menzies, Anastasio Argüello, Sándor Kukovics, Magyar Juh- és Kecsketejgazdasági Közhasznú Egyesület, Herceghalom, 28, 2022. ISBN: 9786150134192

Total IF of journals (all publications): 31,068
Total IF of journals (publications related to the dissertation): 20,368

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

01 February, 2024