PHENOTYPIC CHARACTERIZATION, GENETICS AND GENOMICS OF LIVESTOCK IN LOW INPUT SYSTEMS

EDITED BY: Mohammed Ali Al Abri, Olivier Hubert Hanotte and Younes Miar

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-285-3
DOI 10.3389/978-2-83250-285-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PHENOTYPIC CHARACTERIZATION, GENETICS AND GENOMICS OF LIVESTOCK IN LOW INPUT SYSTEMS

Topic Editors:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman
Olivier Hubert Hanotte, University of Nottingham, United Kingdom
Younes Miar, Dalhousie University, Canada

Citation: Al Abri, M. A., Hanotte, O. H., Miar, Y., eds. (2022). Phenotypic Characterization, Genetics and Genomics of Livestock in Low Input Systems. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-285-3

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 08 September 2022
DOI 10.3389/fgene.2022.1013177

OPEN ACCESS

EDITED AND REVIEWED BY
Martino Cassandro,
University of Padua, Italy

*CORRESPONDENCE
Mohammed Ali Al Abri,
abri1st@squ.edu.om

SPECIALTY SECTION
This article was submitted to Livestock Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 06 August 2022
ACCEPTED 12 August 2022
PUBLISHED 08 September 2022

CITATION
Al Abri MA (2022), Editorial: Phenotypic characterization, genetics and genomics of livestock in low input systems.
Front. Genet. 13:1013177.
doi: 10.3389/fgene.2022.1013177

Editorial: Phenotypic characterization, genetics and genomics of livestock in low input systems

Mohammed Ali Al Abri*

Department of Animal and Veterinary Sciences, College of Agriculture and Marine Sciences, Muscat, Oman

KEYWORDS

genetics, genomics, low input production system, livestock, phenotypic characterization

Editorial on the Research Topic
Phenotypic characterization, genetics and genomics of livestock in low input systems

This Research Topic includes a collection of research papers that highlighted various genetic aspects of indigenous breeds in different countries around the globe. The topic discussed the recent advances, current challenges and latest developments in phenotypic characterization and genetics and genomics of livestock in low input production systems. The research conducted varied from Genome Wide Association Studies (GWAS) and selection signatures to estimation of genetic parameters and heritability to genetic background and genetic diversity of indigenous breeds.

The first manuscript by Sutera et al., implemented a GWAS based on DEBV to identify genomic regions putatively associated with mastitis using somatic cells count (SCC) as a proxy. Their research, which was carried out in local Valle del Belice sheep, highlighted eight genomic markers associated with somatic cell score that harboured candidate genes related to udder conformation and the immune system. These finding could prove very useful for the fate of this and other sheep breeds in the Mediterranean basin countries as the dairy sheep production there usually relies on well adapted local breeds. Another manuscript implementing GWAS was by Mancin et al. who implemented a single step GWAS to analyse Beef Traits in Local Alpine Breed. Their study revealed the diversity of the pathways and some novel genes impacting a number of beef traits. The discovery of these new pathways and genes shows the importance of genetic research in local breeds which can expand to global livestock populations. Habimana et al. conducted a GWAS to study of growth performance and immune response to Newcastle disease virus of indigenous chicken in Rwanda. Their results showed several genes implicated in body weight and New Castle disease antibody variation.

An alternative approach to gene discovery is using a candidate gene approach in which a number of selected genes are selected and genotyped in a population of interest.

Verma et al. implemented this approach to discover a panel of 10 reference genes in Indian cattle populations adapted to hot arid normoxia and cold arid hypoxia environments. All together, the previous studies open the door for the discovery of causative genes that could be utilized in selection programs in these local breeds in future.

Searching for Selection Signatures is another method for discovery of genes impacting traits of interest. Unlike GWAS, it relies on discovery of regions at which nucleotide variations is reduced due to the rise of a new beneficial mutation. The research article by Serranito et al. was on searching for Selection Signatures Related to Trypanosomosis Tolerance in African Goats. Trypanosomosis is a parasite that greatly impacts various livestock species including camels, cattle, sheep and goats with very little research focusing on it in goats. Therefore, examining selection signatures of innate immunity to it is of vital importance to goat production in Africa. The researchers found 33 selection signatures, 18 of which overlapped previously published research. Such findings are crucial for goat production as they could eventually lead to discovery of genes/allelic combinations of major role on immunity to trypanosome infections in goats. Another attempt to search for selection signatures and genes responsible for trypanotolerence was made by (Yougbaré et al.). In this article, the researchers reported several genes potentially involved in tolerance to trypansoma resistance in the Baoulé x Zebu crossbred cattle in Burkina Faso. In addition to the two previous articles, the selection signatures in West African indigenous cattle of Benin was investigated by (Vanvanhossou et al.). In this study, the results unraveled some genes related to production/economic traits in hybrid Beninese cattle and genes related to immunity and feed efficiency in local Somba cattle.

A key element of this article collection is research in genetic improvement of local breeds in low input systems. Such improvement is the corner stone of future utilization of these breeds and can help complement their adaptation traits with production traits of interest. The first step to any genetic improvement program is the estimation of genetic parameters of the traits to be improved by the program. Two articles in this article collection discussed genetic parameters estimation in low input systems breeds. The first article discussed the estimation of genetic parameters of type traits in first parity Slovenian Cika cows using a generalized linear model procedure (Simčič et al.). In this article, the authors reported medium to high heritabilities for body frame and autochthonous traits and low to medium heritabilities for scored udder traits. These heritabilities are already in use in the prediction of breeding values in Cika cattle which is exemplary for other breeds in other low input systems. The second article was by Hako Touko et al. in which the authors estimated the heritabilities of antibody response to Newcastle disease vaccination impacting local chicken in Cameron. Due to the lack of a large sample size or pedigree data (common to low input systems), the authors used three alternative methods to estimate heritabilities. The first was the breeder's equation method, the second method was the graphical method and the third was the full-sib/half-sib method. The authors reported low but similar estimates of antibody response to vaccination and a very low estimate of heritability for survival. Nevertheless, the authors reported a significant increase in antibody response when crossing cocks and hens under 1% and 3% selection intensity respectively.

A number of articles in this collection have explored the genetic and phenotypic background of local breeds including (Alaqeely et al. and Samaraweera et al.). In the former, the authors analysed the sequence variation in the mtDNA control region of Dromedary Camels. Their results showed that small genetic differences exist between camel types. They reported two haplogroups shared between almost all dromedary camels with one of them being more common in African dromedary camels. In addition, in order to better characterize camel populations, the authors advise future researchers to use WGS or SNP data which is the same advise provided by Piro. In the later article, the authors reported high genetic diversity but absence of population structure in local chickens of Sri Lanka using Microsatellite Markers. In contrast to the findings reported in the previous article, this research showed high genetic diversity in Sri Lanka's local chicken breeds which translates to a large potential for the development of locally adapted genetically improved chicken breeds. On the other hand, Kebede et al. used Species distribution models (SDMs) and Phenotypic distribution models (PDMs) to phenotypically characterize the relationship between environmental variation and phenotypic differentiation in Ethiopian indigenous chicken. Their results identified three major ecotypes that show phenotypic differentiation in addition to nine major environmental variables conducive to habitat suitability. In another study, Ouhrouch et al. used Whole Genome Sequencing (WGS) data on 87 sheep to study the genetic uniqueness Moroccan sheep belonging to the five major sheep breeds. Their study assessed the genetic diversity, effective population size, population structure and relationships between the breeds. In addition, they were able to find interesting intra-breed selection signatures. Their study showed that the Morocan breeds are not highly genetically differentiated but that they harbor high genetic diversity and effective population size which makes them a suitable reservoir for the development of adaptive traits against the threats of climate change. The study by Babigumira et al. aimed at characterizing the genetic ancestry and inbreeding levels in the local a population of smallholder pigs in Uganda using 422 pigs. The authors used porcine GeneSeek Genomic Profiler (GGP) 50 K SNP Chip to infer the genetic ancestry as well as inbreeding levels. The study showed a relatively low inbreeding in the breed and revealed a genetic background of the breed that is a mix of old British and modern pig ancestries. Pig production, which is managed mostly through low input systems, is one of the main sources of income for African small holder producers.

The findings in the previous research articles highlight the importance of preserving local breeds for their notable adaptation to various vector borne diseases. Furthermore, identification of genetic regions/genes can help integrate them from genomes of local lower-production breeds to genomes of other breeds with higher production abilities. Therefore, this article collection will serve as a stepping stone towards the characterization, development and genetic improvement of numerous livestock in low input systems.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 March 2021
doi: 10.3389/fgene.2021.643531

Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep

*Anna Maria Sutera[1], Angelo Moscarelli[2], Salvatore Mastrangelo[2], Maria Teresa Sardina[2], Rosalia Di Gerlando[2], Baldassare Portolano[2] and Marco Tolone[2]**

[1] Dipartimento Scienze Veterinarie, University of Messina, Messina, Italy, [2] Dipartimento di Scienze Agrarie Alimentari e Forestali, University of Palermo, Palermo, Italy

In the Mediterranean basin countries, the dairy sheep production is usually based on local breeds, which are very well-adapted to their production systems and environments and can indeed guarantee income, employment, and economic viability in areas where production alternatives are scarce or non-existent. Mastitis is still one of the greatest problems affecting commercial milk production. However, genetic evaluation of mastitis is particularly difficult because of its low heritability and the categorical nature of the trait. The aim of this study was to identify genomic regions putatively associated with somatic cells count (SCC) in the local economically important Valle del Belice sheep breed using of deregressed breeding values (DEBV) as response variables. All the samples were genotyped using the Illumina OvineSNP50K BeadChip. Genome-wide association analysis was carried out based on regression of DEBV. A total of eight markers were found to be significantly associated with log-transformed SCC. Several candidate genes associated with SCC were identified related to immunity system and udder conformation. The results can help improving the competitiveness of the local Valle del Belìce breed. Further studies considering a higher sample size or independent population will be needed to confirm our results.

Keywords: mastitis, local dairy sheep, GWAS, SNPs arrays, candidate genes

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman

Reviewed by:
Xiaozhu Wang, Auburn University, United States
Hadi Atashi, Shiraz University, Iran

***Correspondence:**
Marco Tolone
marco.tolone@unipa.it

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *18 December 2020*
Accepted: *01 March 2021*
Published: *22 March 2021*

Citation:
Sutera AM, Moscarelli A, Mastrangelo S, Sardina MT, Di Gerlando R, Portolano B and Tolone M (2021) Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. Front. Genet. 12:643531. doi: 10.3389/fgene.2021.643531

INTRODUCTION

In the Mediterranean basin countries, the dairy sheep production is usually based on local breeds, which are very well-adapted to their production systems and environments and can indeed guarantee income, employment, and economic viability in areas where production alternatives are scarce or non-existent. Mastitis is the most important problem for the milk industry due to the decrease quality of milk and increased cost of flock regeneration due to early culling of ewes. It can be induced, for example, by a lack of hygiene, by pushed manual milking or feed disorder. In dairy sheep, generally, the most important agents involved in mastitis are the bacterial infections, and the most frequently isolated pathogens are coagulase-negative staphylococci (CNS) that are present on and around the udder skin (Leitner et al., 2008) with a different pathogenicity, causing clinical, and subclinical mastitis (Contreras et al., 2007; Riggio and Portolano, 2015). The udder infection

determines the increase of the somatic cell count (SCC) in milk (Raynal-Ljutovac et al., 2007; Leitner et al., 2008) that causes significant damage of curd and cheese yields. Since the heritability of mastitis is low, genetic selection to improve mastitis by traditional selection is not very effective. SCC or log transformed SCC (i.e., somatic cell score, SCS) have relatively higher heritability compared to mastitis and is used as the first trait to improve mastitis resistance (Shook and Schutz, 1994). Kelly et al. (2000) found that an elevated SCC can alter the protein fractions distribution; decrease casein and lactose levels in milk; increase rennet clotting time, cheese moisture, and losses of fat and proteins in whey, and reduce curd firmness and cheese yielding. A study conducted by Sutera et al. (2018) confirmed that high levels of SCC in sheep milk are associated with milk yield losses and variations of fat and protein percentages. The estimated losses in milk yield ranged from 883 g for SCC $\leq$ 2,000 $\times$ 10^3 to 1,052 g for SCC $\leq$ 500 $\times$ 10^3 with an overall decrease of 16%, whereas fat and protein percentages increased to 0.06 and 0.29%, respectively. The negative effects of mastitis are provoked by a combination of animal characteristics (age, lactation stage, etc.), genetic (breed, inbreeding, etc.) and environmental factors (season, management, etc.) (Oget et al., 2019). Therefore, different individuals may have a different susceptibility to the disease, depending on their genetic heritage. In fact, there are several studies about the mastitis in dairy sheep confirming a genetic basis for mastitis resistance (Tolone et al., 2013; Oget et al., 2019), but no assumption had been made about the genes and the relative mechanisms.

The emergence of high-throughput genotyping technologies allowed routine genome-wide association studies (GWAS) to be performed in livestock populations. GWAS allows screening of the genome utilizing a large number of genetic markers spread across the entire genome to detect genetic variants associated with a particular disease or trait. The estimated breeding values (EBVs) were generally used to perform the GWAS. As an alternative, the EBVs can be "deregressed" (Garrick et al., 2009; Ostersen et al., 2011) to standardize the variance and influence of the individuals' EBVs while still accounting for information from relatives. The use of deregressed EBVs (DEBVs) as dependent variables can improve the power of GWAS (Sell-Kubiak et al., 2015; Sevillano et al., 2015). An advantage of GWAS is that we can overcome the candidate gene approach through which sometimes significant results were not obtained due to the wrong or incomplete choice of candidate genes. In the last decades, several GWASs were conducted in sheep for milk production related traits (Sutera et al., 2019; Li et al., 2020), for fatty acids profile (Rovadoscki et al., 2018), for body weight (Ghasemi et al., 2019; Tao et al., 2020), for wool production (Wang et al., 2014), for nematode resistance (Becker et al., 2020) and ovine lentivirus resistance (White et al., 2012). To date, few GWASs have been conducted for SCC or SCS in dairy sheep (Oget et al., 2019), especially in local adapted breeds.

In Sicily, dairy sheep production represents an important resource for the local economy, and the Valle del Belice is the main local breed reared on the island for the production of traditional raw milk cheeses, at farm level by small local dairies. The breed is subjected to limited breeding selection programs for milk production traits, but shows excellent adaptability to local environments, sometimes with harsh conditions (Mastrangelo et al., 2017). Therefore, the aim of this study was to identify the genomic regions putatively associated with SCC in Valle del Belice sheep breed using of DEBVs as response variables.

MATERIALS AND METHODS

Data and Estimation of Breeding Value

Between 2006 and 2016 the University of Palermo collected phenotypic data from 15 Valle del Belìce flocks, for a total of 1,813 individuals. The milk samples were collected aseptically from each individual from the two udder halves in sterile containers following an A4 recording procedure (ICAR, 2014), stored at 4°C and transferred to the laboratory to determine daily SCC using Fossomatic 6000 (Foss Electric Hillerød, Denmark) equipment. The phenotypic data set originated by these sampling works was composed of 15,008 observations. Animals with less than 3 test-day measurements within lactation were discarded. For each individual the following information were registered: order of parity, number of born lambs, lactation days, age, birth season and somatic cell count. Birth season was classified in three classes: 1 if the lambing was from August to November; 2 from December to March; 3 from April to July. SCC was normalized through a logarithmic transformation into somatic cell score (SCS) according to the formula of Ali and Shook (1980):

$$SCS = log_2\left(\frac{SCC}{100,000}\right) + 3$$

Preliminary analyses using the general linear model of ASReml R (Butler et al., 2009) were performed to determine the significance of the fixed effects where the Wald tests are implemented in the form of the ANOVA method. A single trait repeatability test day (TD) animal model was performed to estimate the breeding values (EBV) as follows:

$$y = X\beta + Z_{htd} + Z_a + Z_{pe} + e$$

where y is the observation vector for SCS TD; β is the vector of fixed effects that includes order of parity (op: 4 classes), age at first lambing (age: 4 classes, 1 when first lambing occurred at 10–14 months of age, 2 at 15–19 months of age, 3 at 20–24 months of age, and 4 at 25–29 months of age); birth season (bs: 3 classes), interaction between herd and birth season (hbs: 74 classes) and days of lactation (dim) modeled with a Legendre polynomial of order three. Htd is the vector of interaction between herd and test day random effect; a is the vector of direct additive genetic effects (breeding values); pe is the vectors of permanent environmental effect between lactations; e is the residual vector. X and Z are the corresponding incidence matrices relating records to fixed, animal, and permanent environmental between lactations effects, respectively. The pedigree file included 5,534 animals with 178

sires and 2,548 dams. The assumptions regarding the components of the model were:

$$E\begin{bmatrix} y \\ b \\ htd \\ a \\ pe \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and $V_a = A\sigma_a^2$; $V_{htd} = I\sigma_{htd}^2$; $V_{pe} = I\sigma_{pe}^2$; $V_e = I\sigma_e^2$ where A is the numerator relationship matrix based on pedigree and I are the identity matrix with orders equal to numbers of dams for htd and pe effects and equal to the records for residuals e. Variance components and breeding values for SCS were estimated based on REML method using ASReml R (Butler et al., 2009). In addition, EBVs were also deregressed according to Garrick et al. (2009) as follows:

$$DEBV = EBV/r^2$$

where, EBV is the estimated breeding value and r^2 is the reliability of that EBV.

Blood Sampling and DNA Extraction

A total of 476 sheep of Valle del Belice breed were sampled. About 10 mL of blood was collected from the jugular vein using vacutainer tubes containing EDTA as anticoagulant. The procedures involving animal sample collection followed the recommendation of directive 2010/63/EU. Sampling was carried out by trained veterinarians within the frame of vaccination Campaigns, hence no permission from the animal research ethics committee was necessary. Veterinarians adhered to standard procedures and relevant national guidelines to ensure appropriate animal care. Genomic DNA was extracted from each blood sample with a salting-out method (Miller et al., 1988). The DNA sample was quantified with a NanoDropND-1000 spectrophotometer (NanoDropTechnologies, Wilmington, DE, United States), diluted to a final concentration of 50 ng/mL (as required by the Illumina Infinium protocol), and stored at 4°C until use.

Genotyping and Quality Control

All the samples were genotyped using the Illumina OvineSNP50K BeadChip v2. Position and chromosomal coordinates for each SNP were obtained from the ovine genome sequence assembly (Oar 4.0)[1]. Quality control and association analyses were performed using GenABEL package (Aulchenko et al., 2007) in R environment[2]. Only SNPs located on autosomes were extracted and considered for further analyses. Animals and markers that fulfilled the following criteria were kept in the analysis: (i) call rate per individuals and per SNPs > 95%; (ii) minor allele frequency > 2%; (iii) no extreme deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$).

[1]https://www.ncbi.nlm.nih.gov/genome/?term=ovis+aries

[2]http://www.r-project.org

GWAS Analyses

Genome-wide association analysis was carried out based on regression of DEBV with the genotypes of animals for one SNP at a time. We used the three-step approach referred to as genomic GRAMMAR-GC (Amin et al., 2007; Aulchenko et al., 2007). The advantage of this approach, especially in livestock, is that it accounts for cryptic population structure caused by the presence of closely related animals (Aulchenko et al., 2007) inferring relationships through genomic marker data. After Bonferroni correction, significant thresholds were $P < 1.34 \times 10^{-6}$ for genome-wide ($P < 0.05$) and $P < 2.69 \times 10^{-5}$ for suggestive ($P < 0.10$) (i.e., one false positive for genome scan), corresponding to −log10(P) equal to 5.87 and 4.56, respectively. Quantile-quantile (Q-Q) plots were used to analyze the extent to which the observed distribution of the statistic test followed the expected (null) distribution, in order to assess potential systematic bias due to population structure or analytical approach. Population substructure was explored using classical multidimensional scaling (MDS) in order to verify the genetic homogeneity of the sample before analysis using PLINK v1.9 (Purcell et al., 2007). The least square means of DEBV for the three genotypes affecting somatic cell count of significant SNP were also calculated by a general linear model (GLM) using R package lsmeans (Lenth and Lenth, 2018) and the significant threshold was set at $P < 0.05$.

Annotation

Genomic regions showing significant results were further explored to identify candidate genes underlying the loci. In particular, the gene contents located at ±250 kb distances from the significant SNP were annotated using Genome Data Viewer genome browser at the National Center for Biotechnology Information Database[3]. The presence of Quantitative Traits Loci (QTLs)[4] related with the considered trait was also checked. Finally, to investigate the biological function and the phenotypes that are known to be affected by each annotated gene, we conducted a comprehensive literature search, including information from other species.

RESULTS

Genetic Parameters and Estimated Breeding Values

Descriptive statistics and genetic merit for SCS in the sampled animals are presented in **Table 1**. About 15,000 TD observations for SCS were considered to estimate EBVs then, the DEBVs of 5,534 individuals were estimated. Heritability and repeatability estimates for SCS in the studied population were 0.045 (standard error = 0.02) and 0.40 (standard error = 0.01), respectively.

[3]https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000298735.2

[4]http://www.animalgenome.org/QTLdb

TABLE 1 | Descriptive statistics for somatic cell score.

Variable	N	Mean	SD	CV	Min–Max
SCS	15,008	2.67	0.72	0.27	1–5.31
DEBV	5,534	–0.19	0.52	2.75	–5.59 to 4.16

SCS, somatic cell score; DEBV, deregressed breeding value; N, number of records; SD, standard deviation; CV, coefficient of variation; Min-Max, minimum and maximum values.

Quality Control for Genotyping Data

Among the 54,241 SNPs, 7,414 SNPs are located on sex chromosomes and thus were withdrawn from the analysis. A total of 3,999 SNPs were removed due to genotype rate <0.05, 2,037 SNPs due to minor allele frequency < 0.02 and 3,651 SNP due to Hardy-Weinberg disequilibrium ($P < 10^{-6}$). Moreover, 12 individuals were also excluded for a low (<95%) call rate. Then, after quality control, we considered a total of 37,140 SNPs and 464 individuals for further analyses.

Genome-Wide Association Analyses

In total we detected eight significant SNPs for SCS, and among these, only one marker reached the genome-wide significant threshold ($P < 4.72 \times 10^{-7}$). The details of these SNPs including *P*-values, the positions on *Ovis aries* v4.0 genome assembly, the chromosomes and the closest known genes are given in **Table 2**. Manhattan plot, showing the profiles of the *P*-values [in terms of –log(P)] of all tested SNPs, is showed in **Figure 1**. The QQ-plot in **Supplementary Figure S1** shows the observed and expected *P*-values of the GWAS for SCS. The genomic inflation factor (lambda) was lower than one indicating some population stratification. However, departure from this line is also expected for a really polygenic trait, as many causal SNPs may not yet have reached genome-wide significance owing to a lack of power (Power et al., 2017). The results for the MDS showed that the bulk of the samples were not separated by the first dimension, indicating a lack of substructure (**Supplementary Figure S2**). The eight SNPs were located on five different chromosomes: three SNPs on OAR1, one SNP on OAR3, one SNP on OAR7, two SNPs on OAR8 and one SNP on OAR10. Considering the range of ±250 kb surrounding each significant SNP, a total of 34 genes (**Table 2**) were found. The most significant SNP (rs161717499) was located within the coding region of the Stress Associated Endoplasmic Reticulum Protein 1 (SERP1) gene on OAR1.

For each of the eight significant SNPs, we calculated the LSM of the DEBV for the three genotypes affecting the trait to investigate their genetic contribution (**Figure 2**). Five out of the eight above reported SNPs (rs401598547, rs403091159, rs161717499, rs422960374, rs426621433) reached the significance threshold ($P < 0.05$). Individuals with homozygous genotypes GG for rs401598547, CC for rs161717499 and AA for rs403091159, rs422960374, and rs426621433, showed lower somatic cells content among all three genotypes (**Figure 2**). After checking on SheepQTLdb tool, no one of the eight detected SNPs was located within a known QTLs related at SCS or mastitis.

TABLE 2 | Single nucleotide polymorphisms (SNPs) significantly associated with somatic cell score at genome-wide ($P < 1.34 \times 10^{-6}$) and suggestive ($P < 2.69 \times 10^{-5}$) thresholds.

OAR	SNP	Position (bp)	−log10 (*p*-value)	Genes
1	rs401598547	46,865,607	4.94	NEGR1
1	rs403091159	49,692,787	4.65	LRRIQ3, LOC105605157, FPGT, LOC101120030
1	rs161717499	235,497,703	6.33	SIAH2, ERICH6, LOC101119269, EIF2A, SERP1, TSC22D2, TRNAR-UCG
3	rs422960374	24,797,321	4.99	FAM49A, TRNAC-GCA
7	rs406841304	57,592,284	4.60	ATP8B4, LOC105607291, DTWD1, FAM227B, FGF7
8	rs420334414	67,510,451	4.63	HIVEP2, AIG1, ADAT2
8	rs426621433	82,781,340	4.90	SOD2, WTAP, ACAT2, TCP1, MRPL18, PNLDC1, MAS1, IGF2R, LOC106991323, LOC106991303, SLC22A1, SLC22A2
10	rs422370366	4,119,025	4.82	–

OAR, Ovis aries chromosome; Genes, the closest genes to the significant SNP found within ± 250 kb region surrounding it.

DISCUSSION

Mastitis is still one of the greatest problems affecting commercial milk production. However genetic evaluation of mastitis is particularly difficult because of the low heritability and the categorical nature of the trait. As a consequence, SCC has been promoted as an indirect method of predicting mammary infections due to the positive correlation between these two traits (Boettcher, 2005). It is worth to mention that collecting information on SCC is easier, cheaper, and less time demanding for farmers compared to use bacteriological status as direct measure of mastitis (Riggio, 2012). In this study we estimated the breeding value for SCS and identified the genomic regions putatively involved in mastitis resistance in the local economically important Valle del Belice dairy sheep breed.

The mean SCS (**Table 1**) was lower than those reported in previous studies (Riggio et al., 2010; Tolone et al., 2013) in the same breed, and by Ariznabarreta et al. (2002) in Churra sheep and Leitner et al. (2003) in Isaraeli-Assaf and Awassi sheep. These differences in SCS in the Valle del Belice breed were due to different sampled population. The heritability estimate for SCS in this study falls within the range (0.04–0.16) reported in literature for sheep (e.g., Barillet et al., 2001; Hamann et al., 2004; Tolone et al., 2013).

In this study, DEBVs of the SCS were used as trait scores for the association analysis. The estimated breeding values (EBVs) were generally used (like as pseudo-phenotypes) to perform the

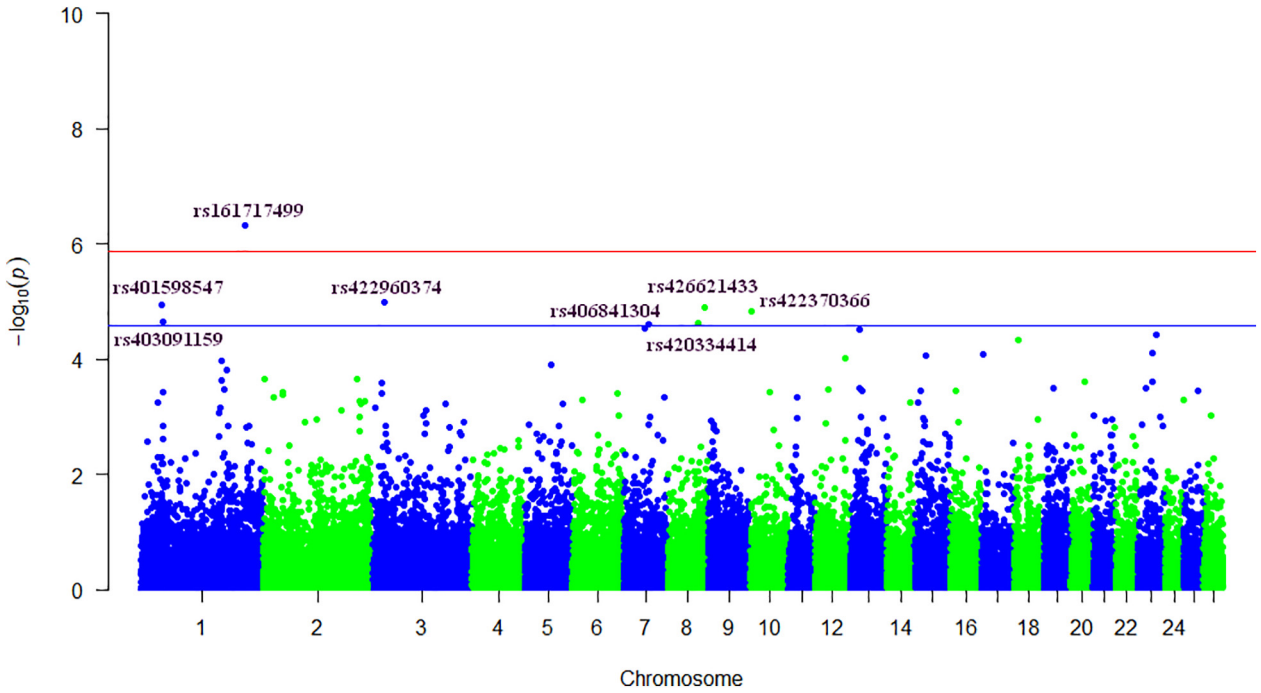

FIGURE 1 | Genome-wide plot of −log10(*P*-values) for association of SNPs with somatic cell score. Blue and red lines represent suggestive [−log10(*P*-values) > 4.57] and genome-wide thresholds [−log10(*P*-values) > 5.87], respectively.

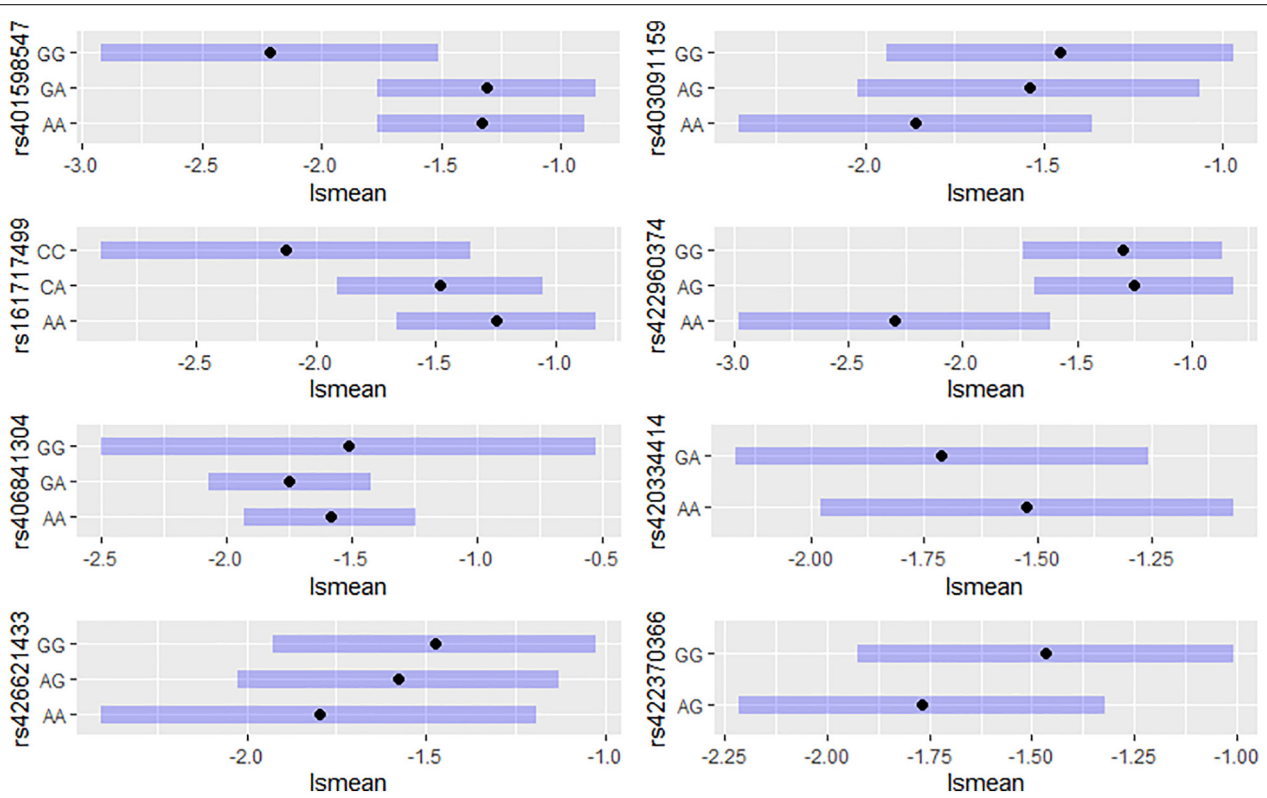

FIGURE 2 | Least squares means (at 95% CI) of DEBV for the three genotypes affecting the trait of significant SNP detected from GWAS study.

GWAS. Although EBVs have been used as dependent variables in GWAS (Johnston et al., 2011; Becker et al., 2013), this approach gave high false positive rate (Ekine et al., 2014). Consequences of using EBVs include varying levels of precision and "shrinkage effect" among the values used to represent phenotypes of different individuals, a reduction in the sample variance of the phenotypes, and double-counting of information from relatives (Garrick et al., 2009; Ostersen et al., 2011). The DEBV make good use of available information from genotyped animals as well as from their relatives, which can appropriately avoid bias introduced by simply pooling or averaging data information and account for heterogeneous variance (Garrick et al., 2009).

A total of eight SNPs were found to be significantly associated with SCS in Valle del Belice sheep. For five significant SNPs, results suggested that individuals with the GG genotypes at rs401598547, CC at rs161717499, and AA at rs403091159, rs422960374, and rs426621433, could be selected to reduce the somatic cells content in milk, although these genotypes had a low frequency in the breed. The lack of selection pressure in Valle del Belice dairy sheep may also contribute to the low frequency of the favorable alleles and genotypes. Therefore, the effect of these alleles for somatic cells content trait should be verified in a larger population or by testing them in an independent sample.

The most significant SNP associated with SCS was located in the intronic region of *SERP1*. This gene encodes the stress-associated endoplasmic reticulum protein 1 and was associated with immune system (Moravčíková et al., 2018). Another relevant gene close to the most significant SNP was *SIAH2*, involved to apoptosis and programmed cell death (Crisà et al., 2016). On the same chromosome, the other two markers were close to *NEGR1*, a gene involved with medium white blood cell count (a leukocyte trait) in Yak (Ma et al., 2019), *LRRIQ3* related to the innate immune system upon recognition of pathogens (Pablo-Maiso et al., 2018) and *FPGT*, which is part of the L-fucose pathway, a key sugar in complex carbohydrates involved in cell-to-cell recognition, inflammation, and immune processes (Becker and Lowe, 2003). As above reported, mastitis is a persistent, inflammatory response of mammary tissue attributed to intramammary invasion of a mastitis-causing pathogen. Therefore, according to their role and function, these aforementioned genes can be considered as candidate involved in mastitis resistance and SCS. The SNP rs422960374 on OAR03, was close (∼70 Kb) to the *FAM49A* gene. Marete et al. (2018), in a GWAS for milking speed in French Holstein cows, reported the *FAM49A* as candidate gene for this trait. This gene was also associated to rear udder height in Holstein cattle (Gonzalez et al., 2020). The genetic correlation between the SCS and udder attachment in sheep was observed by Casu et al. (2010); De la Fuente et al. (1996) reported the indirect selection for subclinical mastitis resistance due to the inclusion of udder morphology traits in selection objectives. Moreover, Gutiérrez-Gil et al. (2018) suggested that sheep with udders and high degree of suspension or shallow udders close to the abdominal wall should be associated to lower SCS. Despite specific functions of this gene are not known yet, all the aforementioned aspects suggested a possible involving of *FAM49A* gene in our trait. Similarly, on OAR7, near to SNP rs406841304, two close genes are related with udder conformation (*FAM227B*) (Scienski et al., 2019) and with epithelial cell proliferation and differentiation (*FGF7*) (Bazer and Slayden, 2008; Yang et al., 2020), suggesting their possible role in the epithelial mammary cell proliferation. Moreover, the *FGF7* has been reported as putative target gene in bovine mammary tissue infected with *Streptococcus uberis* (Naeem et al., 2012). Another significant SNP was located on OAR8 (rs426621433) at position 82,781,340 bp. This SNP mapped within a QTL for SCC (81.4–83.5 Mb) reported in a commercial French dairy sheep population (Rupp et al., 2015), and near a QTL for SCC (ID number 160869) on OAR 8 (80.5–80.6 Mb) in Churra sheep. Among the closest annotated genes in the region of ±250 kb surrounding it, the *SOD2* gene seems to be the most plausible candidate affecting the SCS. In fact, the expression of SOD2 at mRNA and protein levels has been reported up-regulated in the mammary glands of ewes with clinical mastitis compared to healthy ewes (Gao et al., 2019). Mitterhuemer et al. (2010) showed an increase of SOD2 gene level in mammary tissue from mastitis cows inoculated with *E. coli* 24 h after infection as compared to controls. Finally, another candidate gene mapped near SNP rs426621433 on OAR8, was *IGF2R*, with a crucial role for the regulation of cell proliferation, growth, differentiation and survival, and associated with milk production traits. In fact, Dehoff et al. (1998) showed that lactation in the bovine mammary gland is associated with increased *IGF2R* concentration.

CONCLUSION

In this study, we estimated the breeding value for SCS in Valle del Belice sheep. DEBVs of the SCS were used as trait scores for the association analysis. Several candidate genes associated with SCS were identified related to immunity system and udder conformation. These candidate genes provide valuable information for future functional characterization. Therefore, our results may contribute to increase knowledge on the role the genes play in the genetically determined mechanisms involved in mastitis in sheep. The results can help improving the competitiveness of the local Valle del Belìce breed, through the development of genetic improvement programs directed toward reducing the incidence of mastitis, also considering the udder conformation into selection objectives, with planned mating between subjects carrying favorable alleles. Anyway, further studies considering a higher sample size or independent population will be needed to confirm our results.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://figshare.com/articles/dataset/GWAS_for_somatic_cell_counts_in_sheep/13797851.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the procedures involving animal sample collection followed the recommendation of directive 2010/63/EU. Sampling was carried out by trained veterinarians within the frame of vaccination Campaigns, hence no permission from the animal research ethics committee was necessary.

AUTHOR CONTRIBUTIONS

MT and AS: conception of the work. AM, RD, and MT: contributed to the data acquisition. MT, AS, and AM: data analysis. AS, MT, SM, and BP: results interpretation. MT, SM, and AM: drafting the article. MT, SM, MS, and AS: critical revision of the article. AS, AM, SM, MS, RD, BP, and MT: final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financed by the LEO: Livestock Environment Opendata, PSRN 2014—2020—Sottomisura: 16.2, Project number PRJ—0185, CUP: J84I18000090007.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.643531/full#supplementary-material

REFERENCES

Ali, A., and Shook, G. E. (1980). An optimum transformation for somatic cell count in milk. *J. Dairy Sci.* 63, 487–490. doi: 10.3168/jds.S0022-0302(80)82959-6

Amin, N., Van Duijn, C. M., and Aulchenko, Y. S. A. (2007). Genomic background based method for association analysis in related individuals. *PLoS One* 2:e1274. doi: 10.1371/journal.pone.0001274

Ariznabarreta, A., Gonzalo, C., and San Primitivo, F. (2002). Microbiological quality and somatic cell count of ewe milk with special reference to staphylococci. *J. Dairy Sci.* 85, 1370–1375. doi: 10.3168/jds.S0022-0302(02)74203-3

Aulchenko, Y. S., de Koning, D. J., and Haley, C. (2007). Genome wide rapid association using mixed model and regression: a fast and simple method for genome wide pedigree-based quantitative trait loci as-sociation analysis. *Genetics* 177, 577–585. doi: 10.1534/genetics.107.075614

Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J. M., and Jacquin, M. (2001). Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genet. Sel. Evol.* 33, 397–415. doi: 10.1186/1297-9686-33-4-397

Bazer, F. W., and Slayden, O. D. (2008). Progesterone-induced gene expression in uterine epithelia: a myth perpetuated by conventional wisdom. *Biol. Reprod.* 79, 1008–1009. doi: 10.1095/biolreprod.108.072702

Becker, D., Wimmers, K., Luther, H., Hofer, A., and Leeb, T. (2013). A genome-wide association study to detect QTL for commercially important traits in Swiss Large White boars. *PLoS One* 8:e55951. doi: 10.1371/journal.pone.0055951

Becker, D. J., and Lowe, J. B. (2003). Fucose: biosynthesis and biological function in mammals. *Glycobiology* 13, 41R–53R. doi: 10.1093/glycob/cwg054

Becker, G. M., Davenport, K. M., Burke, J. M., Lewis, R. M., Miller, J. E., Morgan, J. L. M., et al. (2020). Genome—wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. *Anim. Genet.* 51, 330–335. doi: 10.1111/age.12895

Boettcher, P. (2005). Breeding for improvement of functional traits in dairy cattle. *Ital. J. Anim. Sci.* 4, 7–16. doi: 10.4081/ijas.2005.3s.7

Butler, D. G., Cullis, B. R., Gilmour, A. R., and Gogel, B. J. (2009). *ASReml-R Reference Manual. Technical Report*. Brisbane, QLD: Queensland Department of Primary Industries.

Casu, S., Sechi, S., Salaris, S. L., and Carta, A. (2010). Phenotypic and genetic relationships between udder morphology and udder health in dairy ewes. *Small Rumin. Res.* 88, 77–83. doi: 10.1016/j.smallrumres.2009.12.013

Contreras, A., Sierra, D., Sánchez, A., Corrales, J. C., Marco, J. C., and Paape, M. J. (2007). Mastitis in small ruminants. *Small Rumin. Res.* 68, 145–153. doi: 10.1016/j.smallrumres.2006.09.011

Crisà, A., Ferrè, F., Chillemi, G., and Moioli, B. (2016). RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC Vet. Res.* 12:264. doi: 10.1186/s12917-016-0881-7

De la Fuente, L. F., Fernandez, G., and San Primitivo, F. (1996). A linear evaluation system for udder traits of dairy ewes. *Livest. Prod. Sci.* 45, 171–178. doi: 10.1016/0301-6226(96)00003-6

Dehoff, M. H., Elgin, R. G., Collier, R. J., and Clemmons, D. R. (1998). Both type I and II insulin-like growth factor receptor binding increase during lactogenesis in bovine mammary tissue. *Endocrinology* 122, 2412–2417. doi: 10.1210/endo-122-6-2412

Ekine, C. C., Rowe, S. J., Bishop, S. C., and de Koning, D. J. (2014). Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3 Genes Genom. Genet.* 4, 341–347. doi: 10.1534/g3.113.008706

Gao, J., Li, T., Lu, Z., Wang, X., Zhao, X., and Ma, Y. (2019). Proteomic analyses of mammary glands provide insight into the immunity and metabolism pathways associated with clinical mastitis in meat sheep. *Animals* 9:309. doi: 10.3390/ani9060309

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55. doi: 10.1186/1297-9686-41-55

Ghasemi, M., Zamani, P., Vatankhah, M., and Abdoli, R. (2019). Genome-wide association study of birth weight in sheep. *Animal* 13, 1797–1803. doi: 10.1017/S1751731118003610

Gonzalez, M., Villa, R., Villa, C., Gonzalez, V., Montano, M., Medina, G., et al. (2020). Inspection of real and imputed genotypes reveled 76 SNPs associated to rear udder height in Holstein cattle. *Adv. Vet. Anim. Res.* 7, 234–241.

Gutiérrez-Gil, B., Esteban-Blanco, C., Suarez-Vega, A., and Arranz, J. J. (2018). Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50K SNP chip and whole-genome sequencing. *J. Dairy Sci.* 101, 9072–9088. doi: 10.3168/jds.2018-14736

Hamann, H., Horstick, A., Wessels, A., and Distl, O. (2004). Estimation of genetic parameters for test day milk production, somatic cell score and litter size at birth in East Friesian ewes. *Livest. Prod. Sci.* 87, 153–160. doi: 10.1016/j.livprodsci.2003.09.015

ICAR (2014). International Agreement of Recording Practices. International Committee for Animal Recording. Available online at: https://www.icar.org/index.php/icar-recording-guidelines/ (accessed September 10, 2020).

Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., et al. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* 20, 2555–2566. doi: 10.1111/j.1365-294X.2011.05076.x

Kelly, A. L., Tiernan, D., O'sullivan, C., and Joyce, P. (2000). Correlation between bovine milk somatic cell count and polymorphonuclear leukocyte level for samples of bulk milk and milk from individual cows. *J. Dairy Sci.* 83, 300–304. doi: 10.3168/jds.S0022-0302(00)74878-8

Leitner, G., Chaffer, M., Caraso, Y., Ezra, E., Kababea, D., Winkler, M., et al. (2003). Udder infection and milk somatic cell count, NAGase activity and milk

composition—fat, protein and lactose—in Israeli-Assaf and Awassi sheep. *Small Rumin. Res.* 49, 157–164. doi: 10.1016/S0921-4488(03)00079-8

Leitner, G., Silanikove, N., and Merin, U. (2008). Estimate of milk and curd yield loss of sheep and goats with intrammamary infection and its relation to somatic cell count. *Small Rumin. Res.* 74, 221–225. doi: 10.1016/j.smallrumres.2007.02.009

Lenth, R., and Lenth, M. R. (2018). Package 'lsmeans'. *Am. Stat.* 34, 216–221.

Li, H., Wu, X. L., Tait, R. G. Jr., Bauck, S., Thomas, D. L., Murphy, T. W., et al. (2020). Genome−wide association study of milk production traits in a crossbred dairy sheep population using three statistical models. *Anim. Genet.* 51, 624–628. doi: 10.1111/age.12956

Ma, X., Jia, C., Fu, D., Chu, M., Ding, X., Wu, X., et al. (2019). Analysis of hematological traits in polled yak by genome-wide association studies using individual SNPs and haplotypes. *Genes* 10:463. doi: 10.3390/genes10060463

Marete, A., Sahana, G., Fritz, S., Lefebvre, R., Barbat, A., Lund, M. S., et al. (2018). Genome-wide association study for milking speed in French Holstein cows. *J. Dairy Sci.* 101, 6205–6219. doi: 10.3168/jds.2017-14067

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., et al. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genet. Sel. Evol.* 49, 84. doi: 10.1186/s12711-017-0360-z

Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A salting out method for DNA preparation. *Nucleic Acids Res.* 16:1215. doi: 10.1093/nar/16.3.1215

Moravčíková, N., Simčič, M., Mészáros, G., Sölkner, J., Kukučková, V., Vlček, M., et al. (2018). Genomic response to natural selection within alpine cattle breeds. *Czech J. Anim. Sci.* 63, 136–143. doi: 10.17221/62/2017-CJAS

Naeem, A., Zhong, K., Moisá, S. J., Drackley, J. K., Moyes, K. M., and Loor, J. J. (2012). Bioinformatics analysis of microRNA and putative target genes in bovine mammary tissue infected with *Streptococcus uberis*. *J. Dairy Sci.* 95, 6397–6408. doi: 10.3168/jds.2011-5173

Oget, C., Tosser-Klopp, G., and Rupp, R. (2019). Genetic and genomic studies in ovine mastitis. *Small Rumin. Res.* 176, 55–64. doi: 10.1016/j.smallrumres.2019.05.011

Ostersen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G., and Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet. Sel. Evol.* 43:38. doi: 10.1186/1297-9686-43-38

Pablo-Maiso, D., Doménech, A., Echeverría, I., Gómez-Arrebola, C., De Andrés, D., Rosati, S., et al. (2018). Prospects in innate immune responses as potential control strategies against non-primate Lentiviruses. *Viruses* 10:435. doi: 10.3390/v10080435

Power, R. A., Parkhill, J., and de Oliveira, T. (2017). Microbial genome-wide association studies: lessons from human GWAS. *Nat Rev Genet* 18:41. doi: 10.1038/nrg.2016.132

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Raynal-Ljutovac, K., Pirisi, A., de Crémoux, R., and Gonzalo, C. (2007). Somatic cells of goat and sheep milk analytical, sanitary, productive and technological aspects. *Small Rumin. Res.* 68, 126–144. doi: 10.1016/j.smallrumres.2006.09.012

Riggio, V. (2012). *Genetic Aspects of Somatic Cell Count and Udder Health in the Italian Valle del Belice Dairy Sheep*. Gelderland: Wageningen University. .

Riggio, V., and Portolano, B. (2015). Genetic selection for reduced somatic cell counts in sheep milk: a review. *Small Rumin. Res.* 126, 33–42. doi: 10.1016/j.smallrumres.2015.01.020

Riggio, V., Portolano, B., Bovenhuis, H., and Bishop, S. C. (2010). Genetic parameters for somatic cell score according to udder infection status in Valle del Belice dairy sheep and impact of imperfect diagnosis of infection. *Genet. Sel. Evol.* 42:30. doi: 10.1186/1297-9686-42-30

Rovadoscki, G. A., Pertile, S. F. N., Alvarenga, A. B., Cesar, A. S. M., Pértille, F., Petrini, J., et al. (2018). Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics* 19:375. doi: 10.1186/s12864-018-4777-8

Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., et al. (2015). A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet* 11:e1005629. doi: 10.1371/journal.pgen.1005629

Scienski, K., Ialacci, A., Bagnato, A., Reginelli, D., Durán-Aguilar, M., and Strillacci, M. G. (2019). Genetic variability in a Holstein population using SNP markers and their use for monitoring mating strategies. *Rev. Mex. Cienc. Pecu.* 10, 643–663. doi: 10.22319/rmcp.v10i3.4842

Sell-Kubiak, E., Duijvesteijn, N., Lopes, M. S., Janss, L. L. G., Knol, E. F., Bijma, P., et al. (2015). Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics* 16:1049. doi: 10.1186/s12864-015-2273-y

Sevillano, C. A., Lopes, M. S., Harlizius, B., Hanenberg, E. H., Knol, E. F., and Bastiaansen, J. W. M. (2015). Genome-wide association study using deregressed breeding values for cryptorchidism and scrotal/inguinal hernia in two pig lines. *Genet. Sel. Evol.* 47:18. doi: 10.1186/s12711-015-0096-6

Shook, G. E., and Schutz, M. M. (1994). Selection on somatic cell score to improve resistance to mastitis in the United States. *J. Dairy Sci.* 77, 648–658. doi: 10.3168/jds.S0022-0302(94)76995-2

Sutera, A. M., Portolano, B., Di Gerlando, R., Sardina, M. T., Mastrangelo, S., and Tolone, M. (2018). Determination of milk production losses and variations of fat and protein percentages according to different levels of somatic cell count in Valle del BeliceBelìce dairy sheep. *Small Rumin. Res.* 162, 39–42. doi: 10.1016/j.smallrumres.2018.03.002

Sutera, A. M., Riggio, V., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Pong−Wong, R., et al. (2019). Genome−wide association studies for milk production traits in Valle del Belìce sheep using repeated measures. *Anim. Genet.* 50, 311–314. doi: 10.1111/age.12789

Tao, L., He, X. Y., Pan, L. X., Wang, J. W., Gan, S. Q., and Chu, M. X. (2020). Genome−wide association study of body weight and conformation traits in neonatal sheep. *Anim. Genet.* 51, 336–340. doi: 10.1111/age.12904

Tolone, M., Riggio, V., and Portolano, B. (2013). Estimation of genetic and phenotypic parameters for bacteriological status of the udder, somatic cell score, and milk yield in dairy sheep using a threshold animal model. *Livest. Sci.* 151, 134–139. doi: 10.1016/j.livsci.2012.11.014

Wang, Z., Zhang, H., Yang, H., Wang, S., Rong, E., Pei, W., et al. (2014). Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS One* 9:e107101. doi: 10.1371/journal.pone.0107101

White, S. N., Mousel, M. R., Herrmann-Hoesing, L. M., Reynolds, J. O., Leymaster, K. A., Neibergs, H. L., et al. (2012). Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. *PLoS One* 7:e47829. doi: 10.1371/journal.pone.0047829

Yang, C., Lan, W., Ye, S., Zhu, B., and Fu, Z. (2020). Transcriptomic analyses reveal the protective immune regulation of conjugated linoleic acids in sheep ruminal epithelial cells. *Front. Physiol.* 11:1368. doi: 10.3389/fphys.2020.588082

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 June 2021
doi: 10.3389/fgene.2021.676047

A Mix of Old British and Modern European Breeds: Genomic Prediction of Breed Composition of Smallholder Pigs in Uganda

Brian Martin Babigumira[1,2], Johann Sölkner[1], Gábor Mészáros[1], Christina Pfeiffer[1,3], Craig R. G. Lewis[4], Emily Ouma[2], Maria Wurzinger[1] and Karen Marshall[5]*

[1] Department of Sustainable Agricultural Systems, Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria, [2] International Livestock Research Institute, Kampala, Uganda, [3] PIG Austria GmbH, Steinhaus, Wels, Austria, [4] Pig Improvement Company, Barcelona, Spain, [5] International Livestock Research Institute, Nairobi, Kenya

Pig herds in Africa comprise genotypes ranging from local ecotypes to commercial breeds. Many animals are composites of these two types and the best levels of crossbreeding for particular production systems are largely unknown. These pigs are managed without structured breeding programs and inbreeding is potentially limiting. The objective of this study was to quantify ancestry contributions and inbreeding levels in a population of smallholder pigs in Uganda. The study was set in the districts of Hoima and Kamuli in Uganda and involved 422 pigs. Pig hair samples were taken from adult and growing pigs in the framework of a longitudinal study investigating productivity and profitability of smallholder pig production. The samples were genotyped using the porcine GeneSeek Genomic Profiler (GGP) 50K SNP Chip. The SNP data was analyzed to infer breed ancestry and autozygosity of the Uganda pigs. The results showed that exotic breeds (modern European and old British) contributed an average of 22.8% with a range of 2–50% while "local" blood contributed 69.2% (36.9–95.2%) to the ancestry of the pigs. Runs of homozygosity (ROH) greater than 2 megabase (Mb) quantified the average genomic inbreeding coefficient of the pigs as 0.043. The scarcity of long ROH indicated low recent inbreeding. We conclude that the genomic background of the pig population in the study is a mix of old British and modern pig ancestries. Best levels of admixture for smallholder pigs are yet to be determined, by linking genotypes and phenotypic records.

Keywords: pig, diversity, admixture, Uganda, SNP, genomic, breed

OPEN ACCESS

Edited by:
Younes Miar,
Dalhousie University, Canada

Reviewed by:
Kenneth Joseph Stalder,
Iowa State University, United States
Duy Ngoc Do,
Dalhousie University, Canada

****Correspondence:***
Johann Sölkner
johann.soelkner@boku.ac.at

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *04 March 2021*
Accepted: *25 May 2021*
Published: *23 June 2021*

Citation:
Babigumira BM, Sölkner J, Mészáros G, Pfeiffer C, Lewis CRG, Ouma E, Wurzinger M and Marshall K (2021) A Mix of Old British and Modern European Breeds: Genomic Prediction of Breed Composition of Smallholder Pigs in Uganda. Front. Genet. 12:676047. doi: 10.3389/fgene.2021.676047

INTRODUCTION

The pig (*Sus scrofa domesticus*), an even toed ungulate and a member of the genus Sus, was domesticated from its ancestor, the wild boar (*Sus scrofa scrofa*) in multiple domestication centers including the Near East, Europe, China and South-east Asia, about 9,000 years ago (Rothschild and Ruvinsky, 2011; Groenen et al., 2012). Wild boar (*Sus scrofa algira*) also inhabits North Africa (Rothschild and Ruvinsky, 2011). Since its domestication, the pig has been genetically improved into several specialized breeds through traditional and marker assisted selective breeding

(Dekkers, 2004; SanCristobal et al., 2006; Mote and Rothschild, 2020). Such work is notable for European breeds such as the Pietrain that has been intensively selected for muscle development (Amaral et al., 2011). The Landrace breed originated from British foundation stock imported to Denmark and selected for leanness and fast growth. Commercial breeds such as the Large White, Berkshire, and Hampshire were developed from crossbreeding old British and Asian pigs (White, 2011; Amills et al., 2013). Iberian pigs (Toro et al., 2008) were exported during the colonization of the Americas and contributed to development of the Duroc (Jones, 1998). Its adaptive attributes and importance as a source of meat have contributed to the global distribution of the pig (Orr and Shen, 2006). Notably, sub-Saharan Africa is not within the native range of wild boar and no archeological or genetic evidence points to a domestications event there (Ramirez et al., 2009). The origin of pigs in East Africa is traced to both the pre- and colonial eras (Blench, 2000, 2008, 2010; Ramirez et al., 2009). Indian Ocean trade and eventual European settlement have also been associated with the introduction of Asian and European pig breeds to East Africa (Boivin et al., 2013).

Pig production is an important livelihood source for smallholder farms managed under low input systems in African countries, for example Nigeria, Uganda and Malawi. Uganda is an East African inland country linked to the Indian Ocean through Kenya (east) or Tanzania (south-east). Pigs in Uganda are represented by domestic pigs (*Sus scrofa domesticus*) and the wild suids including the Giant forest hog (*Hylochoerus meinertzhageni*), Warthog (*Phacochoerus aethiopicus*), and Bush pig (*Potamochoerus porcus*) (Ghiglieri et al., 1982; Reyna-Hurtado et al., 2014). In the mid-19th century, Britain colonized Kenya and Uganda while Germany colonized Tanganyika (present Tanzania). Pig production for lard or bacon was an important consideration by the colonists and several breeding experiments were done with British pig breeds such as Large White, Yorkshire, Berkshire, Tamworth, and Large Black (Montgomery, 1921; Prosser, 1936). Pigs of the Large White breed imported from Kenya, as well as pigs distributed by the Ugandan veterinary department were kept by Ugandan farmers (Uganda, 1940). Details of the main breeds kept by the farmers are mostly lacking, but the pig populations in 1945 and 1959 were reported to be 23,158 and 15,668 (Masefield, 1962). Currently, pig production in Uganda is done by more than a million households that manage over 90% of the national herd of 4.2 million pigs (UBOS, 2019). Uganda's per capita consumption is 3.4 kg/year (FAOSTAT, 2018) and the pro-poor significance of pig farming has recently attracted policy recognition (Sentumbwe, 2017).

While one study using microsatellite data has linked the genetic background of pigs in Uganda to European and Asian ancestries (Noce et al., 2015), the breed composition of most pigs in Uganda is largely unknown and any available breed information is mostly as reported by farmers. A previous study reported local pigs on smallholder farms in Uganda (Mbuza, 1995). According to Blench (2000), African pigs are usually black, with a straight tail and popped swept back ears. Other studies have mentioned exotic breeds like Hampshire, Large White, Duroc, Landrace, and Camborough®, which is a cross Large White, Landrace and Duroc, developed by the Pig Improvement Company (PIC®), having been introduced to Uganda (Ssewannyana and Mukasa, 2004; Walugembe et al., 2014; Greve, 2015; Roesel et al., 2016). Admixture between the different breed types is common.

Since 2012, he International Livestock Research Institute (ILRI) has provided a range of technical solutions to pig production constraints in districts of Uganda where pork production is important (Ouma et al., 2015). In 2017, the ILRI led Uganda Pig Genetics Project was launched to provide technical solutions to pig breeding constraints to support previous and ongoing initiatives. A key research question of the Uganda Pig Genetics project was to determine the most-appropriate pig breed or cross-breed type for different types of smallholder pig producers, considering a variety of issues including farmer preference and profitability, amongst others. As part of this work, household pig enterprises and the pigs within them were longitudinally monitored with genomic analysis undertaken to determine the breed-type of individual pigs kept. This study draws on this genomic data to quantify the genetic background, diversity and inbreeding levels of pigs on smallholder farms in Uganda using high throughput Single Nucleotide Polymorphism (SNP) data. SNP data from international sources, publicly available or privately owned but provided for this project, was used to place the pigs of Uganda onto a global genomic map.

MATERIALS AND METHODS

Ethics Approval

This research was approved by the Institutional Research Ethics Committee (IREC), Institutional Animal Care and use Committee (IACUC) of the ILRI and Vector Control Division–Research and Ethics Review Committee (VCD-REC) of the Ministry of Health of Uganda (MOH). Prior informed consent was obtained from owners of the pigs sampled in Uganda. Research and access and benefit permits (Research Registration number: SS4550) were granted by Uganda National Council of Science and Technology (UNCST, 1990).

Site and Household Selection

This study was conducted in the districts of Hoima, Kamuli, Pallisa, Kumi, and Soroti in Uganda (**Figures 1A–D**). Hoima and Kamuli were the primary Uganda Pig Genetics project sites selected because of the importance of pig production to these districts, amongst other criteria. A purposively selected sample of 200 smallholder pig keeping households, 100 each from Hoima and Kamuli, participated in the study. The districts of Pallisa, Kumi, and Soroti were additional sampling sites for local Uganda pigs.

Animals and Genotyping Data

A sample of 422 pigs from the five districts in Uganda: Hoima (n = 163) Kamuli (n = 218), Kumi (n = 11), Pallisa (n = 12), and Soroti (n = 18) were involved in the study (see **Figure 1**). A total of 41 animals, showing the characteristics of African pigs according to Blench (2000), were sampled from 41 households

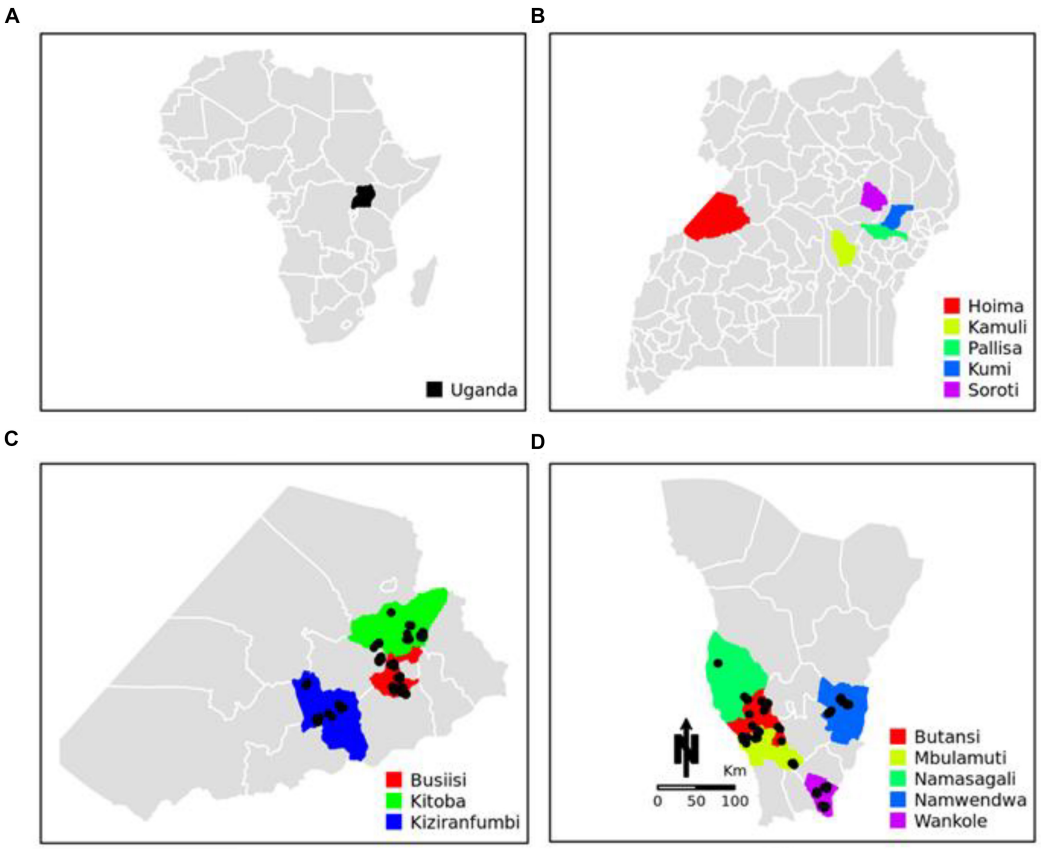

FIGURE 1 | Geographical location of sampling sites. **(A)** Map of Africa showing the location of Uganda. **(B)** Map of Uganda showing the location of Hoima, Kamuli, Pallisa, Kumi, and Soroti districts. **(C)** Map of Hoima showing the locations of Busiisi, Kitoba, and Kizaranfumbi sub-counties. **(D)** Map of Kamuli showing the locations of Butansi, Mbulamuti, Namasagali, Namwendwa, and Wankole sub-counties. The black dots in panels **(C)** and **(D)** show the sampling locations.

having been reported to keep local Uganda pigs by extension staff in the latter three districts. Prevalence of pigs with black coat color, long snout, short legs and popped ears, facing up and backward were criteria of selection of households keeping local pigs. Hair samples taken from the 422 pigs were genotyped using the Neogen GeneSeek Genomic Profiler (GGP) Porcine 50K array (Neogen Europe, 2020). Using literature on East African pigs and phenotypic characteristics of pigs owned by the smallholder farmers in the study area (**Figure 2**), we chose as putative ancestral populations, Asian, Duroc, British, Iberian, and Continental European pig breeds. We explored the ancestry of Uganda pigs in global context by incorporating publicly or privately available genotypes from the putative ancestral populations (Cleveland et al., 2012; Yang et al., 2017; Pena et al., 2019; Pfeiffer et al., 2019; Hlongwane et al., 2020).The data were merged and manipulated in PLINK1.9 (Chang et al., 2015). Prior to merging the data, the SNP positions in each dataset were updated to the sus scrofa reference 11.1 genome build (Illumina, 2013). Quality control (QC) parameters were applied to exclude closely related individuals from each dataset based on PI_HAT using – genome and – max 0.1 flags. The PIC® dataset consisted of 3359 animals. These were genotyped commercial animals – the Camborough® a first filial generation (F1) cross between the PIC® Landrace and PIC® Large White pure lines genotyped with the Illumina PorcineSNP60 chip. The sample consisted of both male and female animals born since 2000 with varying degrees of kinship. Overall, the sampling technique avoided sampling multiple individuals from full-sib families (Cleveland et al., 2012). We use the code CMB throughout this paper to refer to the Camborough® genotypes. The total genotyping rate for CMB data was around 15% lower than for other datasets, therefore we applied the –mind 0.15 flag to only this data. Data merging errors for SNPs with similar positions or on flipped strands were corrected using the – exclude or – flip flags. Samples were randomly excluded when a population exceeded 50. Also breeds without apparent interest to this study, according to literature, were excluded. The merged data (**Table 1**) was explored using Multidimensional scaling (MDS) and ADMIXTURE analysis (Alexander et al., 2009).

Multidimensional Scaling and ADMIXTURE Analysis

Following the exploratory admixture analysis outlined above, we narrowed down the list of reference populations to a panel that, to the best of our judgement, reflected the admixture seen in the

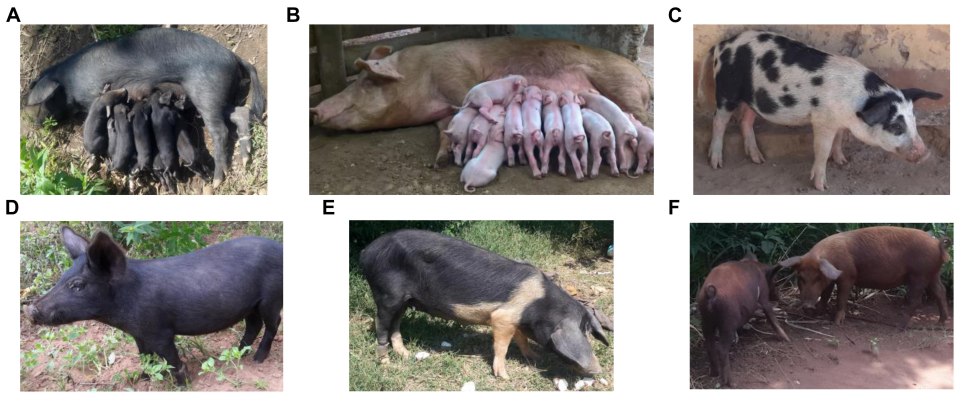

FIGURE 2 | Photographs of pigs of different breed or cross-breed types in Uganda. **(A)** A local sow with her litter **(B)** an exotic breed e.g., Camborough® sow with her litter **(C)** a spotted pig that could be old British (Gloucestershire) or a cross of exotic breeds (Landrace and/or Large White) **(D)** a local pig **(E)** a belted pig that could be a British Saddleback cross and **(F)** two red coated pigs that could be Duroc or Tamworth. Photo credit: Babigumira Brian Martin/ ILRI/ BOKU.

Uganda population. This final reference panel included American (DRC), Chinese (MS), Spanish (IB), Modern European (CMB, LR and LW), Old British (SB and LB) breeds, and Local Ugandan pigs (LOC). We run MDS analysis on the merged dataset using the – distance-matrix flag of PLINK1.9 and Classical Metric MDS and plotted the MDS results in R (Team R Core, 2020). We also run unsupervised ADMIXTURE analysis on the merged dataset for number of ancestral population (K) from two to 10 (Alexander et al., 2009).

TABLE 1 | Breeds/populations used in exploration of the ancestry of Uganda pigs.

Dataset	Breed/population	Country
Uganda Pig Genetics	Hoima, Kamuli, Pallisa, Soroti, and Kumi	Uganda
Cleveland et al., 2012	Camborough® (Pig Improvement Company)	Great Britain
Pfeiffer et al., 2019; Hlongwane et al., 2020; Yang et al., 2017	Landrace and Large White	Austria, South Africa, and Denmark
Yang et al., 2017	Jinhua, Laihuwei, Lantang, and Meishan	China
	Angler Sattelschwein, Bunte Bentheimer	Germany
	Pietrain	Germany and Netherlands
	British Lop, Saddleback, Gloucestershire, Large Black, Leicester, Middle White, Tamworth, and Welsh	Great Britain
	Berkshire and Hampshire	Great Britain and United States
	Casertana and Nera Siciliana	Italy
	Poland Pulawska Spot	Poland
	Breitov, Livni, and Murom	Russia
Yang et al., 2017; Hlongwane et al., 2020	Duroc	America and South Africa
Pena et al., 2019	Entrepelado, Retinto, Entrepelado × Retinto cross, and Retinto × Entrepelado cross	Spain

Population Structure and Admixture Analysis Using CHROMOPAINTERv2, fineSTRUCTUREv4, and GLOBETROTTER

To support the ADMIXTURE and MDS analysis, we analyzed the data using the CHROMOPAINTERv2/fineSTRUCTUREv4 pipeline supported by the Perl scripts provided with the programs (Lawson et al., 2012). The data was phased using SAHPEIT2 (Delaneau et al., 2013). First, a custom R script (Team R Core, 2020) was run to prepare the genetic maps for each chromosome, as required by SHAPEIT2 based on the Sus scrofa recombination map (Tortereau et al., 2012). We run QC (–geno 0.2) and split the data by all autosomal chromosomes using PLINK1.9 (Chang et al., 2015). To achieve a successful run with the provided QC measures (considering size of individual populations and number of variants), we included the –force flag in the SHAPEIT2 command line. We run the impute2chromopainter.pl script to transform the SHAPEIT2 files into the phase format usable by Chromopainterv2. Next, we run the convertrecfile.pl script to generate recombination files using as inputs, the phase files from the previous run and genetic maps based on the Sus scrofa recombination map (Tortereau et al., 2012). We used the default settings for both scripts and specified the HapMap format when using the latter. We then run the phase and recombination files in CHROMOPAINTERv2 (Lawson et al., 2012) twice; the first run was to estimate nuisance parameters and the second one was to generate the co-ancestry matrix using the linked model. The Estimation-Maximization (E-M) iteration was run in automatic mode ("fs") with the entire dataset for all autosomal chromosomes. Basically, each animal

was conditioned on the others in 10 E-M iterations using a sample of ten animals. The main output were two inferred nuisance parameters (Ne, somewhat similar to effective population size and mu, the mutation/switch rate) (Hellenthal, 2012). These parameters (Ne = 34.7106 and mu = 0.00500584) were fixed in the CHROMOPAINTERv2 algorithm in the second run. The main outputs were estimation of the c-factor (effective number of chunks; $c = 0.17931$) and copying vectors. These outputs were fed into the Bayesian clustering algorithm of fineSTRUCTUREv4 for all autosomes.

To further investigate the admixture in the Ugandan pig population used in this study, we exploited the analytical capabilities of GLOBETROTTER (Hellenthal et al., 2014). The Bayesian clustering algorithm of fineSTRUCTUREv4 identified 40 clusters, which, when grouped, were generally not different from our labeled data or the output from ADMIXTURE1.3. Therefore, we run GLOBETROTTER to identify, date and describe admixture in the Uganda pigs using as surrogates: MS, DRC, IB, Modern European (CMB, LR, and LW) and Old British (SB and LB) and LOC with KAM or HOI as target (recipient) populations (Hellenthal et al., 2014; Hellenthal, 2020). We ran GLOBETROTTER with default settings for all parameters except "prop.ind," "bootstrap.date.ind," and "null.ind." For the first run, we set "bootstrap.date.ind" to 0 and the other two to 1. In the second run, we set "prop.ind" to 0 and the other two to 1. For the third run, we set "null.ind" to 0 and the other two to 1 (Hellenthal, 2020). Here, we report the results from the last run.

Autozygosity Analysis of Uganda Pigs

Autozygosity is the inheritance of alleles that are identical by descent (IBD). Contiguous homozygous genotype segments of the genome are called runs of homozygosity (ROH) (Gibson et al., 2006). The ROH can be used to infer the genomic inbreeding coefficient (F_{ROH}) and distinguish ancient from recent inbreeding (Keller et al., 2011). We run this analysis using the dataset of Uganda (HOI and KAM) pigs (381 samples and 50,697 SNPs).

TABLE 2 | Ancestral and Uganda population used in ancestry analysis.

Dataset	Population	Breed	Code	Country	Samples
Cleveland et al., 2012	Modern European	Camborough®	CMB	Great Britain	30
Hlongwane et al., 2020	Duroc	Duroc	DRC	South Africa	20
Pena et al., 2019	Iberian	Iberian	IB	Spain	24
Uganda Pig Genetics	Hoima	Hoima	HOI	Uganda	161
	Kamuli	Kamuli	KAM		218
	Local	Local	LOC		38
Yang et al., 2017	Meishan	Meishan	MS	China	20
	Modern European	Landrace	LR	Denmark	20
		Large White	LW		16
	Old British	Saddleback	SB	Great Britain	20
		Large Black	LB		20

The analysis was run in the cgaTOH (Zhang et al., 2013). The ROH analysis was run using minimum run lengths of 2, 4, 8, and 16 Mb and at least 20 SNP. Heterozygous calls in ROH were not allowed up to 16 Mb while only one heterozygous call was allowed for ROH > 16 Mb. The proportions of the ROH for each of the cut-offs (F_{ROHi}, i = 2, 4, 8, and 16) were computed using as total length of autosome covered by SNPs of 2,259,445,079 bases. The genomic inbreeding coefficient (F_{ROH}) was computed (McQuillan et al., 2008) as follows:

$$F_{\mathrm{ROH}} = \sum \frac{L_{\mathrm{ROH}}}{L_{\mathrm{AUTO}}}$$

L_{ROH} is the sum of ROH per individual and L_{AUTO} is the total length of autosome covered by SNPs.

RESULTS

Exploratory Analysis of Uganda Pigs in a Global Context

The merged dataset used in the exploratory analysis had 28,894 SNP**s** and 1,198 animals from 44 populations and 31 breeds (**Table 2**). The first eigenvector of the MDS analysis separated the Chinese and Iberian from the rest of the populations. The second eigenvector separated the Duroc from the rest of the populations. Both eigenvectors explained about 17% of the variation observed (**Figure 3**). The Ugandan samples were all situated inside a large cluster, including British and Continental European breeds.

Following the exploratory analysis, we retained 30 of the 50 Camborough® (CMB) samples based on proportions of both Large White and Landrace breeds. Further, we removed three local Ugandan pigs that had high exotic proportions. Finally, the panel of ancestral breeds narrowed down to those potentially interesting based on their ancestry contribution in the Uganda pigs. The final dataset (**Table 2**) had 28,894 SNPs, 587 samples from 9 populations and 7 breeds. The populations were Meishan, Duroc, Iberian, Modern European (Landrace, Camborough®, and Large White), old British (Saddleback and Large Black), Uganda (Hoima, Kamuli, and Local).

Multidimensional Scaling and Admixture Analysis

The first eigenvector of the MDS analysis of the dataset in **Table 2** separates the Chinese and Iberian breeds from the Uganda, American, Modern European, and Old British breeds. The second eigenvector clusters some of the modern European breeds (largely comprised of Large White) closely with the Uganda pigs. It also separates the Uganda pigs from the rest of the Modern European, Old British, Duroc, Iberian and Chinese breeds (**Figure 4**).

We ran unsupervised analysis to infer ancestries of HOI and KAM pigs using various ancestral populations (K) without getting meaningful clusters at the lowest cross-validation error (CV) value. Therefore, we selected results at K = 6 and visualized the results using POPHELPER (Francis, 2017). The LOC pigs (purple) represented the main ancestry which was also

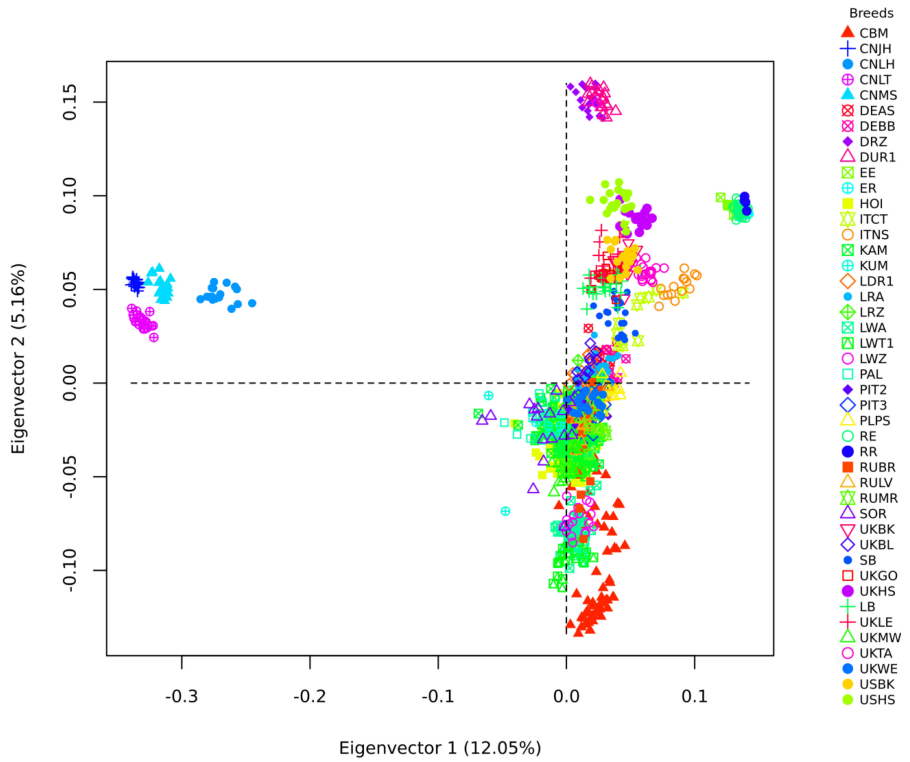

FIGURE 3 | Multidimensional Scaling analysis of Uganda pigs in a global context. CMB: Camborough®; Great Britain; CNJH: Jinhua; CNLH: Laihuwei, CNLT Lantang; CNMS: Meishan-China; DEAS Angler Sattelschwein-Germany; DEBB Bunte Bentheimer-Germany; DRZ: Duroc-South Africa; DUR1: Duroc-United States; EE: Entrepelado; ER-Spain: Entrepelado × Retinto cross-Spain; HOI: Hoima-Uganda; ITCT: Casertana-Italy; ITNS: Nera Siciliana-Italy; KAM: Kamuli-Uganda; KUM: Kumi-Uganda; LDR1: Landrace-Denmark; LRA: Landrace-Austria; LRZ: Landrace-South Africa; LWA: Large White-Austria; LWT1: Large White-Denmark; LWZ: Large White-South Africa; PAL: Pallisa-Uganda; PIT2: Pietrain-German; PIT3: Pietrain-Netherlands; RE: Retinto x Entrepelado cross-Spain; RR: Retinto-Spain; RUBR: Breitov-Russia; RULV: Livni-Russia; RUMR: Murom-Russia; SOR: Soroti-Uganda; UKBK: Berkshire-Great Britain; SB: Saddleback-Great Britain; UKGO: Gloucestershire-Great Britain; UKHS: Hampshire-Great Britain; LB: Large Black-Great Britain; UKLE: Leicester-Great Britain; UKMW: Middle White-Great Britain; UKTA: Tamworth-Great Britain; UKWE: Welsh-Great Britain; USBK: Berkshire-United Sates of America; USHS: Hampshire-United Sates of America.

shared with Old British breeds. The modern European breeds contributed most of the "exotic" ancestry in the Hoima and Kamuli pigs (**Figures 5A,B**).

Results from using ADMIXTURE1.3 showed that modern European breeds (CMB, LR and LW) contributed on average 22.8% with a range of 2–50% of the ancestry while LOC contributed 69.2% (36.9–95.2%). The other 8.0% were contributed by DRC, IB and MS. We also found higher frequency of MS ancestry in LOC than in HOI or KAM pigs (**Figures 5A,B**). Note that ADMIXTURE1.3 did not separate the Old British breeds into a uniform cluster but linked it to various populations, notably to Iberian and Ugandan types.

Population Structure and Admixture Analysis Using CHROMOPAINTERv2, fineSTRUCTUREv4, and GLOBETROTTER

The fineSTRUCTUREv4 analysis identifies three main clusters based on the empirical *c*-value – HOI, KAM, and LOC and the third cluster comprising international breeds (DRC, MS, IB, Modern European and Old British breeds). Considering that fineSTRUCTUREv4 did not identify clusters that differed much from our labeled data, we run "as is" the data in GLOBETROTTER to identify and date the admixture. GLOBETROTTER identified a one-date-multiway (1-DMW) for HOI and a one-date (1-D) admixture event with two sources for KAM pigs. The GLOBETROTER inferred date and confidence intervals (95% CI) for HOI and KAM were 6.371 (3.543-7.311) and 4.719 (2.420-5.093) generations (**Figure 6**). We converted generations to years using a generation interval of 1.9 years (Welsh et al., 2010) and the present year as 2019 in the formula (Hudjashov et al., 2017):

$$Y = y - (1 + x) \times g$$

where Y is the admixture date in years, y, the present year, x, the generations inferred by GLOBETROTTER and g, the generation interval. The admixture date and 95% CI (years) for HOI was 2004 (2003–2010) and for KAM, 2008 (2007–2012). For KAM, the best match sources of admixture were mostly Modern European (CMB, LR, and LW) and LOC pigs. The best match sources for the admixture event in HOI pigs were LOC and modern European (**Figure 7**).

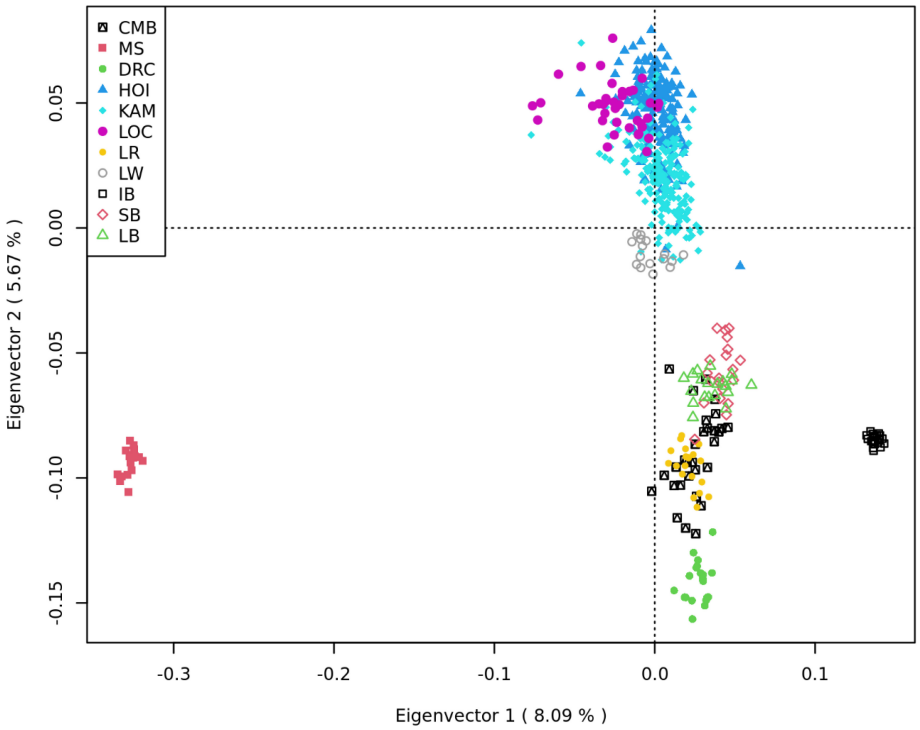

FIGURE 4 | Multidimensional scaling analysis of Camborough® (CMB); Meishan (MS); Duroc (DRC); Uganda [Hoima (HOI), Kamuli (KAM), and Local (LOC)]; Landrace (LR); Large White (LW); Iberian (IB); Saddleback (SB); and Large Black (LB). The first Eigen vector separates the MS and IB from the rest of the population. The second eigenvector closely clusters LW and Uganda pigs and separates them from the rest.

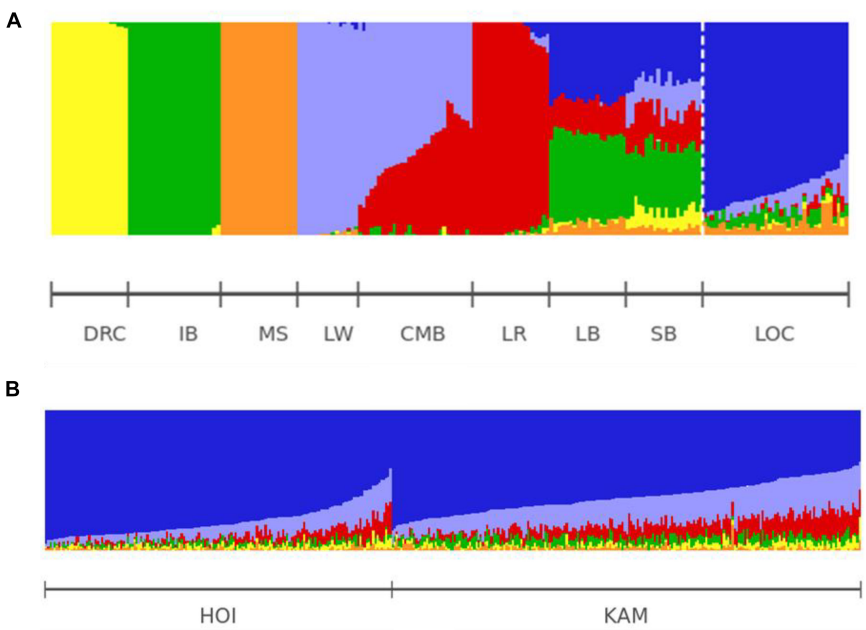

FIGURE 5 | Admixture analysis. **(A)** Ancestry pig populations for $K = 6$: the modern European consists of Large white (LW; sky blue), Landrace (LR; red) and Camborough® (CMB; nearly half sky blue/red). The old British breeds (SB: Saddleback: and LB: Large Black) share ancestries with the Iberian (IB), modern European and local Ugandan pigs (LOC). **(B)** The Uganda pigs Hoima (HOI) and Kamuli (KAM) have a dominant blue ancestry that we refer to as "local" shared with the old British breeds. The modern European breeds contribute most of the exotic ancestry.

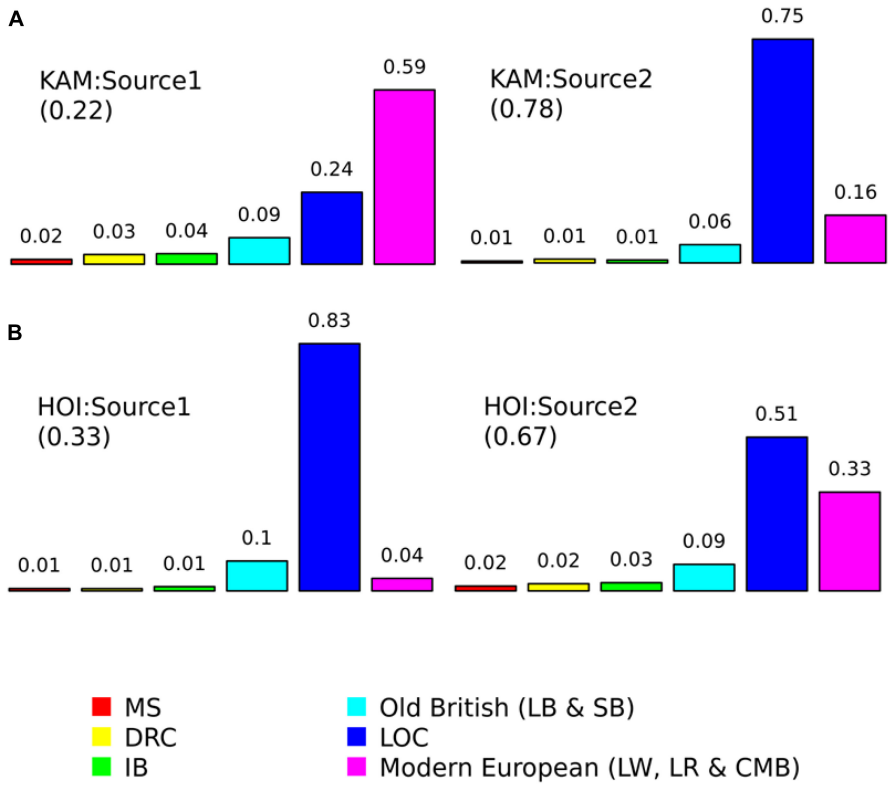

FIGURE 6 | Proportions contributed by surrogate populations to the minor and major sources of admixture for **(A)** Kamuli (KAM) and **(B)** Hoima (HOI) pigs. The surrogate populations are: Meishan (MS); Duroc (DRC); Iberian (IB); Large Black (LB); Saddleback (SB); local Ugandan (LOC); Large White (LW); Landrace (LR); and Camborough® (CMB). The numbers in brackets are the proportions each source contributes to the admixture in the target (recipient) population and they sum up to one. The numbers on top of the bars are the contributions of each surrogate population within each source and they sum up to one.

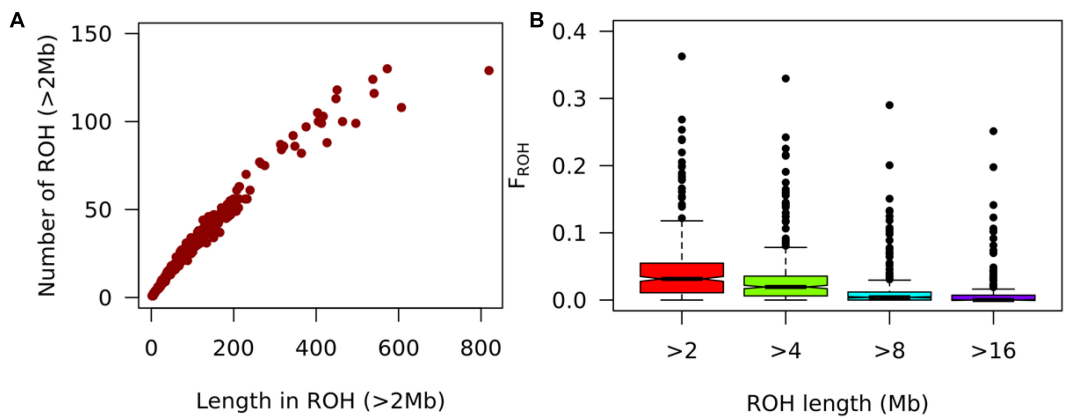

FIGURE 7 | Analysis of autozygosity states. **(A)** Number of ROH distributed and along cumulative total ROH length on the pig genome in ROH. **(B)** Boxplot of $_{FROH}$ cutoffs of 2, 4, 8, and 16 Mb for smallholder pigs from Hoima and Kamuli districts in Uganda.

Autozygosity Analysis of Uganda Pigs

For a 50K SNP Chip, ROHs of length less than 2Mb may contain undetected heterozygotes and hence prone to false positives (Ferencakovic et al., 2013). Therefore, we reported inbreeding levels ($F_{ROH\ >\ 2Mb}$) for ROH lengths greater than 2Mb. Only 348 of 381 pigs from Hoima and Kamuli districts, Uganda had at least one ROH > 2Mb. The $F_{ROH\ >\ 2Mb}$ ranged from 0.000 to 0.363 with a mean of 0.043. The average ROH > 2Mb length per animal was 3.6 ± 1.9Mb and most pigs (81.6%) had at least one ROH > 4Mb. The average F_{ROH} for ROH length of 4, 8, and 16 were 0.030, 0.013, and 0.007. The longest individual ROHs (>20 Mb) were on chromosomes 4 and 14. The most

inbred individual ($F_{ROH > 2Mb}$ = 0.363) had 129 ROHs, the longest individual ROH (28.9Mb), longest total length of ROHs (819.35 Mb) for $F_{ROH > 2Mb}$ and was from Kamuli district (**Figures 7A,B**).

DISCUSSION

Breed Composition of Uganda Pigs

While we use the term LOC (local) to refer to pigs commonly considered to have been in Uganda for some time, we also note that no pig domestication event in Sub-Saharan Africa has been reported or supported by archeological or genetic evidence (Blench, 2000; Amills et al., 2013). Therefore, technically, indigenous Ugandan (Ugandese) pigs do not exist. However, we use the term local (LOC) throughout this paper to differentiate the resident population from exotic ones. We found what appeared LOC to relate more to ancestry contributions from black or belted old British pigs (**Figures 5A,B**). We also found signatures of MS in both the Uganda pigs as previously reported (Noce et al., 2015) and old British breeds. The MS signature in Uganda pigs is likely from an introgression through the old British breeds (Ramirez et al., 2009). We also observed a higher LOC ancestry in HOI than KAM pigs. This may be because Hoima is located further from Kampala, the capital of Uganda and a source of exotic pigs, than Kamuli. The local pigs of Uganda are not characterized and are only identified phenotypically according to the definition of African pigs by Blench (2000). It was difficult to find the local pigs especially where restocking programs had been or were operational. Our results complement previous findings and advocate for characterization and conservation of local pigs in Uganda.

The GLOBETROTTER analysis identified a one-date-multiway (1-DMW) admixture event for HOI pigs. The event involved mostly LOC and modern European breeds and dated the event to 2004 (95% CI: 2003–2010). In the case of Kamuli, a one date admixture event mostly involving modern European and LOC pigs was identified and dated to 2008 (95% CI: 2007–2012). These admixture dates imply recent introductions of exotic pigs in these areas, corresponding with varied "on-the ground" activities that have been observed with the introduction of new pig breeds. Pigs have been distributed to Ugandan farmers through programs run by the government and non-government organizations (Ampaire, 2011; Tatwangire, 2013; Ouma, 2017). The inferred admixture dates coincide with the out scaling of National Agricultural Advisory Services (NAADS) programs in Uganda. The NAADS program essentially sourced and distributed farm inputs including pigs and other livestock to smallholder farmers (Benin et al., 2007; Ouma et al., 2015). Non-government organizations in Kamuli that also distributed pigs to smallholder farmers are Volunteer Efforts for Development Concerns (VEDCO) (Ampaire, 2011), and Iowa State University-Center for Sustainable Rural Livelihoods (Csrl, 2021). Additionally, smallholder pig farmers in Uganda may also purchase pigs mainly from other nearby smallholder pig keepers or local markets (Ouma et al., 2015; Lichoti et al., 2016). These programs or farmers aim to improve productivity of pig herds through crossbreeding by distributing or purchasing pigs of commercial breeds including Landrace, Large White or Camborough®. The GLOBETROTTER and ADMIXTURE results together suggest the following. First, restocking programs have the potential to change the genetic constitution of smallholder pig herds. Second, the several admixture sources observed in the HOI and KAM pigs suggest indiscriminate crossbreeding (Greve, 2015) rather than for example a two- or three-way crossbreeding program. However, they could also suggest an ongoing upgrading of local herds given the proportionately higher frequency of the Modern European breed alleles.

Inbreeding Levels of Uganda Pigs

Using the porcine GGP 50K SNP Chip, we investigated the occurrence of ROHs and quantified autozygosity in pigs in Kamuli and Hoima districts of Uganda. In this study, we found the genomic inbreeding coefficient ($F_{ROH > 2Mb}$) to be 0.043 (0–0.363) for HOI and KAM pigs. The low $F_{ROH > 2Mb}$ indicates low inbreeding in the pig population. This is contrary to what has been previously reported (Tatwangire, 2014). Because of the very small herd size, sows are typically mated with village boars. Boar keepers usually source boars from outside the local area and the piglets received as payment for boar service are sold. Additionally, farmers with sows may source village boar service from sources outside their village depending on boar availability (Ouma et al., 2014; Lichoti et al., 2016). These scenarios suggest a low likelihood of mating related individuals, thus keeping inbreeding levels low. Somewhat higher inbreeding levels could be expected for households which own boars, but this a small minority. For instance, the most inbred individual ($F_{ROH > 2Mb}$ = 0.363) could be the offspring of full sib or parent-offspring mating. Also, events necessitating stock replacement, like African Swine Fever outbreaks (Lichoti et al., 2016; Ouma et al., 2018) would also lower inbreeding levels.

CONCLUSION

Smallholder pig production in Uganda is constrained by several factors, mostly related to pig health, nutrition and genetics (Ouma et al., 2015). Coupling genetic improvement with other appropriate management interventions would enhance productivity of smallholder pig herds. The results of this study showed that the contribution of Modern European ancestry did not exceed 50% for any of the animals sampled in Uganda. This was contrary to expectation, based on breed composition reported by smallholder farmers. The terms "local," "crossbred," and "exotic" used in this context seemed to reflect farmer perception rather than actual breed history. The gradient of ancestries observed in the Hoima and Kamuli populations of this study is still big enough to investigate the effect of the proportion of Modern European ancestry on growth, health and reproduction of pigs in those areas. Longitudinal data for these traits for most of the animals in the current study is available and will be analyzed subsequently. Only then appropriate

crossbreeding levels may be determined and farmers advised about choice and sources of breeding stock.

DATA AVAILABILITY STATEMENT

The porcine 50k SNP chip data, including 50697 SNPs of 422 animals from Uganda included in this study was uploaded to DRYAD. The dataset has been assigned a unique identifier doi: 10.5061/dryad.4qrfj6q95 and is accessible via this temporary link: https://datadryad.org/stash/share/qKhv_9otEd2ivmo6TsIPuQHG30ZFg3BuSJjlg5SDj_M.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Research Ethics Committee (IREC), Institutional Animal Care and use Committee (IACUC) of the International Livestock Research Institute (ILRI) and Vector Control Division–Research and Ethics Review Committee (VCD-REC) of the Ministry of Health of Uganda (MOH). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

JS and KM conceived and designed the whole study. EO, JS, and KM provided the technical, administrative, and logistic support. BB oversaw the field-work, analyzed the data, and drafted the manuscript. CL and CP provided the genotype data through their companies. GM supported assembly of the data. BB and JS analyzed the data. GM, KM, CL, and CP contributed to the interpretation of the data. All authors critically revised and approved the manuscript for submission.

FUNDING

This study was part of the Uganda Pig Genetics project, led by the International Livestock Research Institute and University of Natural Resources and Life Sciences, Vienna (BOKU), and funded by the Austrian Development Agency and the CGIAR Research Program on Livestock (Livestock CRP).

ACKNOWLEDGMENTS

The smallholder pig farmers and district officials in the districts of Kamuli, Hoima, Soroti, Kumi, and Pallisa, as well as the Uganda Pig Genetics project staff particularly the enumerators Peter Lule and David Ahura, are gratefully thanked. Noelia Ibañez Escriche, Universitat Politècnica de València, provided data of Iberian pigs, thank you for that. Garett Hellenthal is thanked for availing CHROMOPAINTERv2 and GLOBETROTTER and the quick response to queries.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Amaral, A. J., Ferretti, L., Megens, H.-J., Crooijmans, R. P., Nie, H., Ramos-Onsins, S. E., et al. (2011). Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLoS One* 6:e14782. doi: 10.1371/journal.pone.0014782

Amills, M., Ramírez, O., Galman-Omitogun, O., and Clop, A. (2013). Domestic Pigs in Africa. *Afr. Archaeol. Rev.* 30, 73–82.

Ampaire, A. (2011). *Farmers' Experiences With Rearing Pigs, Goats And Chickens To Improve Household Nutrition And Income In Kamuli, Uganda*. Ph.D. thesis, Ames: Iowa State University.

Benin, S., Nkonya, E., Okecho, G., Pender, J., Nahdy, S., and Mugarura, S. (2007). *Assessing the impact of the National Agricultural Advisory Services (NAADS) in the Uganda rural livelihoods*. Washington: IFPRI.

Blench, R. (2008). The Austronesians in Madagascar and their interaction with the Bantu of the East African coast: Surveying the linguistic evidence for domestic and translocated animals. *Stud. Philippine Lang. Cult.* 18, 18–43.

Blench, R. M. (2000). "A history of pigs in Africa," in *Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography*, eds R. M. Blench and K. Mac Donald (Abingdon: Routledge Books).

Blench, R. M. (2010). "New evidence for the Austronesian impact on the East African coast," in *The Global Origins and Development of Seafaring (Proceedings of the Conference held in Cambridge, 19th-21st september 2007)*, eds C. Anderson, J. Barrett, and K. Boyle (Cambridge: McDonald Institute Monographs), 239–248.

Boivin, N., Crowther, A., Helm, R., and Fuller, D. Q. (2013). East Africa and Madagascar in the Indian Ocean world. *J. World Prehist.* 26, 213–281. doi: 10.1007/s10963-013-9067-4

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

Cleveland, M. A., Hickey, J. M., and Forni, S. (2012). A common dataset for genomic analysis of livestock populations. *G3 (Bethesda)* 2, 429–435. doi: 10.1534/g3.111.001453

Csrl, C. F. S. R. L. (2021). *Livestock Integration*. United States: Iowa State University.

Dekkers, J. C. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82, E313–E328.

Delaneau, O., Zagury, J. F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6. doi: 10.1038/nmeth.2307

FAOSTAT. (2018). *Pigmeat supply quantity (kg/capita/yr) in Uganda. in: Food Supply-Livestock and Fish Primary Equivalent. Food and Agriculture Organization of the United Nations (FAO)*. Available online at: http://www.fao.org/faostat/en/#data/CL/visualize> (Accessed May 23, 2018).

Ferencakovic, M., Hamzic, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., et al. (2013). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J. Anim. Breed. Genet.* 130, 286–293. doi: 10.1111/jbg.12012

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17, 27–32. doi: 10.1111/1755-0998.12509

Ghiglieri, M. P., Butynski, T. M., Struhsaker, T. T., Leland, L., Wallis, S. J., and Waser, P. (1982). Bush pig (*Potamochoerus porcus*) polychromatism and ecology in Kibale Forest, Uganda. *Afr. J. Ecol.* 20, 231–236. doi: 10.1111/j.1365-2028.1982.tb00298.x

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Greve, D. (2015). Analysis of performance, management practices and challenges to intensive pig farming in peri-urban Kampala, Uganda. *Int. J. Livest. Prod.* 6, 1–7. doi: 10.5897/ijlp2014.0223

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

Hellenthal, G. (2012). Instruction Manual for "ChromoPainter: a copying model for exploring admixture in population data". Available online at: http://gensoft.pasteur.fr/docs/chromopainter/0.0.4/ChromoPainterInstructions.pdf

Hellenthal, G. (2020). Instruction Manual for "GLOBETROTTER: a program for identifying, dating and describing admixture events in population data". Available online at: https://people.maths.bris.ac.uk/~madjl/finestructure/globetrotter.html

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., et al. (2014). A genetic atlas of human admixture history. *Science* 343, 747–751. doi: 10.1126/science.1243518

Hlongwane, N. L., Hadebe, K., Soma, P., Dzomba, E. F., and Muchadeyi, F. C. (2020). Genome wide assessment of genetic variation and population distinctiveness of the pig family in South Africa. *Front. Genet.* 11:344. doi: 10.3389/fgene.2020.00344

Hudjashov, G., Karafet, T. M., Lawson, D. J., Downey, S., Savina, O., Sudoyo, H., et al. (2017). Complex Patterns of Admixture across the Indonesian Archipelago. *Mol. Biol. Evol.* 34, 2439–2452. doi: 10.1093/molbev/msx196

Illumina. (2013). *PorcineSNP60 v2.0 Manifest File*. San Diego: Illumina

Jones, G. (1998). "Genetic aspects of domestication, common breeds and their origin," in *The Genetics of the Pig*, eds M. F. Rothschild and A. Ruvinsky (Oxon: CAB International), 17–50.

Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249. doi: 10.1534/genetics.111.130922

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet.* 8:e1002453. doi: 10.1371/journal.pgen.1002453

Lichoti, J. K., Davies, J., Kitala, P. M., Githigia, S. M., Okoth, E., Maru, Y., et al. (2016). Social network analysis provides insights into African swine fever epidemiology. *Prev. Vet. Med.* 126, 1–10. doi: 10.1016/j.prevetmed.2016.01.019

Masefield, G. B. (1962). Agricultural change in Uganda: 1945-1960. *Food Res. Inst. Stud.* 3, 87–124.

Mbuza, F. (1995). The indigenous domestic animal genetic resources of Uganda. *Anim. Genet. Res. Inf.* 15, 23–42. doi: 10.1017/s1014233900000407

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372.

Montgomery, R. E. (1921). On a form of swine fever occurring in British East Africa (Kenya Colony). *J. Comp. Pathol. Ther.* 34, 159–191.

Mote, B. E., and Rothschild, M. F. (2020). "Modern genetic and genomic improvement of the pig," in *Animal Agriculture*, eds F. W. Bazer, G. C. Lamb, and G. Wu (Cambridge: Academic Press), 249–262. doi: 10.1016/b978-0-12-817052-6.00014-8

Noce, A., Amills, M., Manunza, A., Muwanika, V., Muhangi, D., Aliro, T., et al. (2015). East African pigs have a complex Indian, Far Eastern and Western ancestry. *Anim. Genet.* 46, 433–436. doi: 10.1111/age.12305

Orr, D., and Shen, Y. (2006). "World pig production, opportunity or threat," in *Indiana Farm Bureau's Midwest Swine Nutrition Conference*, Sheridan: .JBS United, Inc.

Ouma, E. (2017). "Overview of ILRI's smallholder pig value chain efforts in Hoima district," in *Presented at the Stakeholders Meeting on Ugandan Pigs, Hoima, Uganda*, (Nairobi, Kenya).

Ouma, E., Dione, M., Birungi, R., Lule, P., Mayega, L., and Dizyee, K. (2018). African swine fever control and market integration in Ugandan peri-urban smallholder pig value chains: an ex-ante impact assessment of interventions and their interaction. *Prevent. Vet. Med.* 151, 29–39. doi: 10.1016/j.prevetmed.2017.12.010

Ouma, E., Dione, M., Lule, P., Rosel, K., and Pezo, D. (2014). Characterization of smallholder pig production systems in Uganda: constraints and opportunities for engaging with market systems. *Livest. Res. Rural Dev.* 26:56.

Ouma, E. A., Dione, M. M., Lule, P. M., Pezo, D. A., Marshall, K., Roesel, K., et al. (2015). *Smallholder pig value chain assessment in Uganda: results from producer focus group discussions and key informant interviews*. Kenya: ILRI.

Pena, R. N., Noguera, J. L., Garcia-Santana, M. J., Gonzalez, E., Tejeda, J. F., Ros-Freixedes, R., et al. (2019). Five genomic regions have a major impact on fat composition in Iberian pigs. *Sci. Rep.* 9:2031.

Pfeiffer, C., Fuerst-Waltl, B., Schodl, K., and Knapp, P. (2019). Genetic Analysis of Feet and Leg Conformation and Proportion of Crushed Piglets in Austrian Large White and Landrace Sows. *Acta Univ. Agric. Silvic. Mendelianae Brun.* 67, 1213–1219. doi: 10.11118/actaun201967051213

Prosser, J. (1936). Prospects of pig farming in British East Africa. *East Afri. Agric. J.* 2, 488– 491.

Ramirez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L. S., Folch, J. M., et al. (2009). Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol. Biol. Evol.* 26, 2061–2072. doi: 10.1093/molbev/msp118

Reyna-Hurtado, R., Tumukunde, A., Chapman, C. A., Rojas, E., Sanvicente, M., Sengupta, R., et al. (2014). On the track of the Giant Forest Hog in Kibale National Park, Uganda: a preliminary report on studying the species. *Suiform Sound* 12:38.

Roesel, K., Nockler, K., Baumann, M. P., Fries, R., Dione, M. M., Clausen, P. H., et al. (2016). First Report of the Occurrence of Trichinella-Specific Antibodies in Domestic Pigs in Central and Eastern Uganda. *PLoS One* 11:e0166258. doi: 10.1371/journal.pone.0166258

Rothschild, M. F., and Ruvinsky, A. (eds). (2011). *The genetics of the pig*.Wallingford: Cabi Publishing.

SanCristobal, M., Chevalet, C., Haley, C., Joosten, R., Rattink, A., Harlizius, B., et al. (2006). Genetic diversity within and between European pig breeds using microsatellite markers. *Anim. Genet.* 37, 189–198.

Sentumbwe, J. (2017). *Uganda Government Livestock Sector Prioritization in Uganda under the Agriculture Sector Strategic Plan (ASSP)*.Uganda: Ministry of Agriculture, Animal Industry and Fisheries.

Ssewannyana, E., and Mukasa, B. (2004). Assessment of the potential productivity of pigs in the Teso and Lango farming systems, Uganda: A case study. *Uganda J. Agric. Sci.* 9, 549–553.

Tatwangire, A. (2013). *Successes and Failures of Institutional Innovations to Improve Access to Services, Input and Output Markets for Smallholder Pig Production Systems and Value Chains in Uganda*.

Tatwangire, A. (2014). *Uganda smallholder pigs value chain development: Situation analysis and trends. ILRI (aka ILCA and ILRAD)*. Nairobi: ILRI

Team R Core. (2020). *R: A Language and Environment for Statistical Computing*. R: A Language and Environment for Statistical Computing.

Toro, M. A., Rodrigañez, J., Silio, L., and Rodriguez, C. (2008). Genealogical Analysis of a Closed Herd of Black Hairless Iberian Pigs. *Conserv. Biol.* 14, 1843–1851. doi: 10.1111/j.1523-1739.2000.99322.x

Tortereau, F., Servin, B., Frantz, L., Megens, H. J., Milan, D., Rohrer, G., et al. (2012). A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* 13:586. doi: 10.1186/1471-2164-13-586

UBOS. (2019). *Statistical abstract*.Kampala: Uganda Bureau of Statistics.

Uganda, P. (1940). *Annual Reports of the Veterinary Department for the years ended 31st December, 1939, 1940, 1941, 1942, 1943, and 1944"*. Entebbe : Govt. Printer.

UNCST. (1990). *Uganda National Council for Science and Technology (UNCST) Act (CAP 209)*. Kampala: UPPC.

Walugembe, M., Nadiope, G., Stock, J., Stalder, K. J., Pezo, D., and Rothschild, M. F. (2014). Prediction of live body weight using various body measurements in Ugandan village pigs. *Livest. Res. Rural Dev.* 26:1. doi: 10.5455/ijlr.20170102113257

Welsh, C. S., Stewart, T. S., Schwab, C., and Blackburn, H. D. (2010). Pedigree analysis of 5 swine breeds in the United States and the implications for genetic conservation. *J. Anim. Sci.* 88, 1610–1618. doi: 10.2527/jas.2009-2537

White, S. (2011). From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. *Environ. Hist.* 16, 94–120. doi: 10.1093/envhis/emq143

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R. P., Zinovieva, N., et al. (2017). Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet. Sel. Evol.* 49:71.

Zhang, L., Orloff, M. S., Reber, S., Li, S., Zhao, Y., and Eng, C. (2013). cgaTOH: extended approach for identifying tracts of homozygosity. *PLoS One* 8:e57772. doi: 10.1371/journal.pone.0057772

Conflict of Interest: CP was employed by company PIG Austria GmbH. CL was employed by company Pig Improvement Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 August 2021
doi: 10.3389/fgene.2021.715732

Search for Selection Signatures Related to Trypanosomosis Tolerance in African Goats

*Bruno Serranito[1,2], Dominique Taurisson-Mouret[3], Sahraoui Harkat[4], Abbas Laoun[5], Nadjet-Amina Ouchene-Khelifi[4], François Pompanon[6], Badr Benjelloun[7], Giuliano Cecchi[8], Sophie Thevenon[9,10], Johannes A. Lenstra[11] and Anne Da Silva[2]**

[1] Museum National d'Histoire Naturelle, CRESCO, Dinard, France, [2] University of Limoges, PEREINE, E2LIM, Limoges, France, [3] University of Limoges, GEOLAB, UMR 6042, Limoges, France, [4] Science Veterinary Institute, University of Blida, Blida, Algeria, [5] University of Djelfa, Djelfa, Algeria, [6] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Grenoble, France, [7] National Institute of Agronomic Research, Regional Centre of Agronomic Research, Beni-Mellal, Morocco, [8] Food and Agriculture Organization of the United Nations, Animal Production and Health Division, Rome, Italy, [9] CIRAD, UMR INTERTRYP, Montpellier, France, [10] INTERTRYP, University of Montpellier, CIRAD, IRD, Montpellier, France, [11] Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Olivier Hubert Hanotte, University of Nottingham, United Kingdom

Reviewed by:
Qianjun Zhao, Institute of Animal Sciences (CAS), China
Guillermo Giovambattista, CONICET Institute of Veterinary Genetics (IGEVET), Argentina

***Correspondence:**
Anne Da Silva
anne.blondeau@unilim.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *27 May 2021*
Accepted: *28 June 2021*
Published: *03 August 2021*

Citation:
Serranito B, Taurisson-Mouret D, Harkat S, Laoun A, Ouchene-Khelifi N-A, Pompanon F, Benjelloun B, Cecchi G, Thevenon S, Lenstra JA and Da Silva A (2021) Search for Selection Signatures Related to Trypanosomosis Tolerance in African Goats. Front. Genet. 12:715732. doi: 10.3389/fgene.2021.715732

Livestock is heavily affected by trypanosomosis in Africa. Through strong selective pressure, several African indigenous breeds of cattle and small ruminants have acquired varying degrees of tolerance against this disease. In this study, we combined LFMM and PCAdapt for analyzing two datasets of goats from West-Central Africa and East Africa, respectively, both comprising breeds with different assumed levels of trypanotolerance. The objectives were (i) to identify molecular signatures of selection related to trypanotolerance; and (ii) to guide an optimal sampling for subsequent studies. From 33 identified signatures, 18 had been detected previously in the literature as being mainly associated with climatic adaptations. The most plausible signatures of trypanotolerance indicate the genes *DIS3L2*, *COPS7B*, *PD5A*, *UBE2K*, and *UBR1*. The last gene is of particular interest since previous literature has already identified E3-ubiquitin ligases as playing a decisive role in the immune response. For following-up on these findings, the West-Central African area appears particularly relevant because of (i) a clear parasitic load gradient related to a humidity gradient, and (ii) still restricted admixture levels between goat breeds. This study illustrates the importance of protecting local breeds, which have retained unique allelic combinations conferring their remarkable adaptations.

Keywords: molecular adaptations, local breeds, Sub-Saharan Africa, crossbreeding, tsetse flies

INTRODUCTION

The productivity of the livestock sector in Africa is heavily affected by trypanosomosis, *via* morbidity, mortality, impact on growth and reproduction (Budd, 1999; Kristjanson et al., 1999; Swallow, 2000; Shaw et al., 2017). This vector-borne parasitic infection, also called "Nagana" in domestic animals, is caused by extracellular protozoa that are mainly cyclically transmitted by several species of tsetse flies (*Glossina* spp.). The tsetse distribution area, or tsetse fly "belt," covers 39

African countries, whose economic development is thus significantly impacted (Alsan, 2015). While many studies have examined trypanosomosis in cattle, this disease has been poorly investigated in sheep and even less in goats (Geerts et al., 2009).

For African goats, the most pathogenic trypanosomes are *Trypanosoma vivax* and *Trypanosoma congolense* (Kusiluka and Kambarage, 1996; Gutierrez et al., 2006). An estimated 173 million of Africa's 228 million goats live in areas of the continent infested by tsetse flies (Gutierrez et al., 2006). In addition to the economic consequences of trypanosomosis on caprine production (Kanyari et al., 1986), goats could act as a reservoir of trypanosomes for other species, including humans (Informal Expert Group on Gambiense HAT Reservoirs et al., 2018).

The high parasitic pressure in humid to sub-humid dry climates may have played a major role in the historical southward movement of domestic stocks. This movement started at 4,000 to 5,000 YBP on the occasion of major and more favorable environmental changes (Smith, 1992; Newman, 1995). It was probably slowed down by endemic parasitic diseases, so that it reached Southern Africa only around 2000 YBP (Pereira and Amorim, 2010). Under the strong pressure of natural selection that occurred throughout the millennia, indigenous breeds have acquired adaptations that enable them to tolerate various diseases and parasites. In the same way as in cattle and sheep, a few goat breeds are likely to show qualities of tolerance to trypanosomosis. Well defined in cattle, trypanotolerance is a multigenic traits that refers to the ability of some breeds to live and be productive in trypanosomosis endemic areas while susceptible animals usually die without treatments (Murray et al., 1984). Phenotypically, trypanotolerant breeds display lower parasitic loads, maintain higher packed cell volume and are less affected by weight loss during infection (Hanotte et al., 2003; Berthier et al., 2016). In West and Central Africa, about 47% of the goats, the Djallonke or West African Darf breeds that live in tsetse infested areas, are estimated to be in some way trypanotolerant (Mawuena, 1986; Bengaly et al., 1993; Osaer et al., 1994; Agyemang, 2005; Geerts et al., 2009). Trypanotolerance in East Africa populations has been less well studied (Geerts et al., 2009), but some Small East African goat breeds exhibit various trypanotolerance traits (Griffin and Allonby, 1979a,b; Monirei et al., 1982; Murray et al., 1982; Kanyari et al., 1986; Mutayoba et al., 1989; Katunguka-Rwakishaya et al., 1997).

In this study, we used published 50K SNP datasets for sub-Saharan goat populations, including West-Central as well as East African populations, with the objective of identifying molecular signatures of selection related to trypanosomosis tolerance. First, we selected native goats from the Adaptmap dataset, choosing breeds that had been living in tsetse-infested areas for a long time, and of neighborhood breeds living under low tsetse pressure on the other hand. Second, we mapped their distribution area and extracted the data allowing to infer the level of pressure linked to trypanosomes but also the corresponding climatic and topographic variables. Then we searched for selection signatures using PCAdapt (Luu et al., 2017) and LFMM (Frichot et al., 2013) methods. Briefly, PCAdapt uses principal component analysis to describe the population structure and identifies candidate markers as outliers in terms of inferred population structure. The LFMM approach searches for significant associations with environmental factors whilst controlling for the neutral population structure. As complementary methods, we included the Hapflk (Fariello et al., 2013) and the Bayescan (Foll et al., 2010) programs. Finally, (i) we identified plausible candidate genes for trypanosomosis tolerance by analyzing our results in relation to the literature. In addition, (ii) our data may guide an optimization of African goats sampling for follow-up studies.

MATERIALS AND METHODS

Datasets Building

We used the AdaptMap goat dataset (Bertolini et al., 2018), which represents a worldwide coverage of original breeds. Out of the 75 African breeds initially present in the dataset, we selected the 40 indigenous local breeds[1], to which we added four Algerian breeds (Ouchene-Khelifi et al., 2018; **Supplementary Table 1**).

On the basis of a NeighborNet visualization (Huson, 1998) of the Reynolds' distances between these breeds (**Supplementary Figure 1**) we selected two groups of related breeds in the West-Central Africa and in East Africa where trypanotolerant breeds are located. We retained (i) 190 West-Central African goats from eight local breeds in Burkina Faso, Cameroon, Mali, and Nigeria, including the trypanotolerant West African Dwarf type breeds (**Table 1**); and (ii) 242 East African goats from 11 local breeds in Burundi, Kenya, Tanzania, and Uganda, including the trypanotolerant Small East African type breeds (**Table 2**).

Environmental Characterization

We defined the environmental characteristics at the level of the distribution area of the breeds. In West-Central Africa we find in the arid zone, the Sahelian type which is progressively replaced by the West African Dwarf type as it moves down toward the equatorial humid zone. In East Africa, most of the breeds are derived from the Small East African type. From their cradle each type has spread, defining populations or breeds that differ from the basic type by their specificities at the local level (see **Tables 1**, **2**). The breed's distribution was mapped (**Figure 1**) *via* ArcGIS (ESRI, 2011). We used the predicted distribution of tsetse flies (*Glossina* species) from the FAO Programme Against African Trypanosomosis (PAAT) to extract the proportion of infestation in each breed's distribution area (Wint and Rodgers, 2000). The breeds were classified into groups according to the intensity of selection pressure due to trypanosomes transmitted by tsetse flies, detected by this method in their distribution area (**Tables 1**, **2**).

We also used the WorldClim database (v.1.4; Hijmans et al., 2005) to collect bioclimatic data covering the period from 1960 to 1990, with a spatial resolution of 30 arc-seconds in the WGS84 datum. We considered the relevant bioclimatic variables (BIO) to highlight the major temperature and humidity contrasts between the breed distribution areas (see **Supplementary Table 2**). Altitude information was collected from the SRTM 90 m Digital Elevation Database (v.4.1) (Jarvis et al., 2008). Finally, the latitude

[1] www.fao.org/dad-is

TABLE 1 | West-Central African goat dataset.

Breed name and acronym (type)	Nb. of heads	Country	Area of the breed in the country	Climatic characteristics of the distribution area	Sampling location* (longitude/ latitude)	Proportion of the distribution area infested by the tsetse flies and ranking for LFMM analysis
Djallonke: DJA (WAD type)	12	Burkina Faso	Sudan area covering the southern part of Burkina-Faso (from latitude 13°5′N to 11°3′N approximately) Traoré et al., 2009	Mean altitude: 318.1 m Annual mean temperature: 27.6°C Annual precipitation: 959.5 mm	−3.53/10.56	90.01% Rank: 1
Cameroonian: CAM (WAD type)	40	Cameroon	North area of the country MINEPIA, 2000	Mean altitude: 392.2 m Annual mean temperature: 26.8°C Annual precipitation: 959.5 mm	14.39/10.11	27.32% Rank: 2
West African Dwarf: WAD (WAD type)	34	Cameroon	South coastal area MINEPIA, 2000	Mean altitude: 379.7 m Annual mean temperature: 24.9°C Annual precipitation: 2,704.7 mm	10.27/5.9	93.98% Rank: 1
Naine: NAI (WAD type)	17	Mali	Southern part of the country LUKNOW, 1950	Mean altitude: 326.6 m Annual mean temperature: 27.3°C Annual precipitation: 972.3 mm	−7.48/11.42	95.14% Rank: 1
Soudanaise: SDN (Sahelian type)	24	Mali	Central part of the country LUKNOW, 1950	Mean altitude: 287.5 m Annual mean temperature: 28.4°C Annual precipitation: 378.1 mm	−6.27/13.45	6.62% Rank: 3
Red Sokoto: RSK (Sahelian type)	21	Nigeria	Central and northern part of the country Blench, 1999; Meyer, 2020	Mean altitude: 441.7 m Annual mean temperature: 26.0°C Annual precipitation: 1,027.1 mm	8.17/11.89	45.75% Rank: 2
Sahelian: SHL (Sahelian type)	21	Nigeria	Extreme northern part of the country Blench, 1999	Mean altitude: 347.2 m Annual mean temperature: 26.7°C Annual precipitation: 538.9 mm	8.73/11.25	0.26% Rank: 3
West African Dwarf: WADn (WAD type)	21	Nigeria	Southern part of the country Blench, 1999	Mean altitude: 164.3 m Annual mean temperature: 26.7°C Annual precipitation: 1,531.4 mm	3.74/7.59	98.30% Rank: 1

**According to Bertolini et al. (2018): mean altitude in meters, annual mean temperature in degrees Celsius, and annual precipitation in millimeters.*

was considered as a proxy providing information on luminosity and seasonality. This procedure was performed in R 3.5.2 (R Core Team, 2018) using the R package "RSAGA" (Brenning, 2008).

A PCA analysis was followed by a hierarchical clustering on principal components (HCPC) using Euclidian distances and Ward's method. Multivariate analyses were performed with R software, using the FactoMineR package (Lê et al., 2008).

Different LFMM analyses were conducted considering (i) the classification of the breeds based on the level of tsetse infestation in their distribution range, and (ii) the classification of the distribution areas obtained through the PCA/HPCP procedure for the climatic variables.

Genetic Characterization of the Breeds

We used the AdaptMap goat dataset, genotyped with the CaprineSNP50 BeadChip (Dryad[2]) for 53,547 SNPs. SNPs and animals were pruned with PLINK v1.07 (Purcell et al., 2007) using the following parameters: (i) SNP call rate ≤97%, (ii) SNP minor allele frequency (MAF) ≤1%, (iii) animals displaying ≥10% of missing genotypes and retaining 50,796 genotypes for 242 East African goats and 50,459 for 190 West-Central African goats.

Admixture Analyses

For ADMIXTURE analysis (Alexander et al., 2009), LD-based SNP pruning was carried out using the –indep option of PLINK with the following parameters: 50 SNPs per window, a shift of five SNPs between windows, and a variation inflation factor's threshold of two (corresponding to $r^2 > 0.5$). ADMIXTURE was run with K = 2–8 for the West-Central goat data-set and K = 2–12, for the East goat dataset, which are the respective numbers of breeds in our datasets. For each value of K, 10 independent runs were performed. The entropy criterion was calculated *via* the sNMF function implemented in the R package LEA to assess the number of ancestral populations that best explains the genotypic data (Alexander and Lange, 2011; Frichot et al., 2014). The program CLUMPAK (Kopelman et al., 2015[3]) was used to analyze the multiple independent runs at a single K and visualize the results.

Search for Molecular Selection Signature

We used different methods to identify loci possibly involved in local adaptation:

The PCAdapt approach, through the R pcadapt package, was used to scan the genome and identify outliers considering population structure. Plotting the number of principal components K from 1 to 15 using the PCAdapt plot function (Luu et al., 2017), revealed K = 5 as optimal number of genetic groups for both data-sets (see **Supplementary Figure 2**). Candidate SNPs were identified by calculating the false discovery rate (FDR; $\alpha = 0.05$) of the p-values

[2]https://doi.org/10.5061/dryad.v8g21pt

[3]http://clumpak.tau.ac.il

TABLE 2 | East African goat dataset.

Breed name and acronym	Nb. of heads	Country	Area of the breed in the country	Climatic characteristics of the distribution area	Sample location* (longitude/ latitude)	Proportion of the distribution area infested by the tsetse flies and ranking for LFMM analysis
Burundi: BUR	40	Burundi	Highlands Ndayishimiye, 1986	Mean altitude: 1,633.9 m Annual mean temperature: 18.9°C Annual precipitation: 1,279.7 mm	29.83/−2.91	10.92% Rank: 2
Galla: GAL	23	Kenya	Northern arid and semi-arid areas Porter et al., 2016	Mean altitude: 590.6 m Annual mean temperature: 26.7°C Annual precipitation: 337.1 mm	37.66/2.01	18.54% Rank: 2
Small East African: SEAK	31	Kenya	Northern arid and semi-arid areas Porter et al., 2016	Mean altitude: 590.6 m Annual mean temperature: 26.7°C Annual precipitation: 337.1 mm	36.97/0.61	18.54% Rank: 2
Gogo (SEA type): GOG	13	Tanzania	Dodoma region of central Tanzania Nguluma et al., 2016	Mean altitude: 1,254.9 m Annual mean temperature: 21.0°C Annual precipitation: 655.7 mm	36.68/−6.28	18.87% Rank: 2
Maasai: MAA	20	Tanzania	Northern Tanzania along the Great Rift Valley on semi-arid and arid lands Nguluma et al., 2016	Mean altitude: 1,071.3 m Annual mean temperature: 21.4°C Annual precipitation: 663.2	37.23/−4.53	54.44% Rank: 1
Sonjo (SEA type): SNJ	22	Tanzania	Ngorongoro district Nguluma et al., 2016	Mean altitude: 1,520.8 m Annual mean temperature: 20.1°C Annual precipitation: 750.5 mm	36.32/−2.7	55.47% Rank: 1
Small East African: SEAU	15	Uganda	Northern savannah ecological areas and northern part of Buganda Nsubuga, 1994; Onzima et al., 2018	Mean altitude: 1,062.4 m Annual mean temperature: 23.4°C Annual precipitation: 1,283.6 mm	32.07/0.58	47.44% Rank: 1
Karamoja (SEA type): KAR	20	Uganda	Karamoja region Nsubuga, 1994; Onzima et al., 2018	Mean altitude: 1,190.6 m Annual mean temperature: 23.1°C Annual precipitation: 833.0 mm	34.67/2.53	9.47% Rank: 2
Sebei (SEA type): SEB	24	Uganda	Kapchorwa district on the slopes of Mount Elgon Nsubuga, 1994; Onzima et al., 2018	Mean altitude: 1,079.7 m Annual mean temperature: 23.1°C Annual precipitation: 1,163.9 mm	34.45/1.4	13.58% Rank: 2
Mubende: MUB	23	Uganda	Mubende district Nsubuga, 1994; Onzima et al., 2018	Mean altitude: 1,224.0 m Annual mean temperature: 21.5°C Annual precipitation: 1,003.2 mm	32.29/0.44	13.36% Rank: 2
Nganda: NGD	11	Uganda	East of the central province Nsubuga, 1994; Onzima et al., 2018	Mean altitude: 1,139.2 m Annual mean temperature: 21.9°C Annual precipitation: 1,272.4 mm	32.58/0.32	50.87% Rank: 1

**According to Bertolini et al. (2018): mean altitude in meters, annual mean temperature in degrees Celsius, and annual precipitation in millimeters.*

associated with Mahalanobis distance estimated by PCAdapt, using the *q*-value function of the R package *q*-value (Storey and Tibshirani, 2003).

The LFMM latent factor mixed models, developed in the LEA R package (Frichot and François, 2015) was used to test the association genotype-environment, using a linear mixed model. We chose $K = 1$, for the number of latent factors, which take into account for neutral genetic structure, in accordance with the entropy criterion (Lee and Seung, 1999) assessed in the admixture analyses. We performed LFMM using 1,000 sweeps for burn-in and 9,000 additional sweeps, and 10 runs with different seeds. We then chose significant associations based on FDR ($\alpha = 0.05$) using the R package *q*-value.

BayeScan software V 2.1 (Foll and Gaggiotti, 2008) estimates the posterior probability that a given SNP is the target of selection on the basis on the allele frequencies in each population and using a Bayesian approach *via* Markov chain Monte Carlo (MCMC). The software was set up with 5,000 burn-in interactions, followed by 10,000 interactions with thinning interval of 10. The top candidate SNPs potentially under selection was identified with a false discovery rate of 0.05.

The hapFLK 1.3.0 program[4] was used to identify signatures of selection accounting for both the hierarchical structure and haplotype information, as described by Fariello et al. (2013). The software was run on a per chromosome basis, using a kinship matrix based on the matrix of Reynolds' genetic distances between breeds (Bonhomme et al., 2010). No outgroups were defined. The number of haplotype clusters per chromosome was determined in fastPHASE using cross-validation (Scheet and Stephens, 2006) based estimation and was set at 25 clusters (-K 25). The hapFLK statistic was computed for 20 EM runs to fit the LD model (–nfit = 20). *p*-values were generated for each SNP using a chi-square distribution with a python script provided on the hapFLK webpage. A *q*-value threshold of 0.05 was applied to limit the number of false positives. We used the

[4]https://forge-dga.jouy.inra.fr/projects/hapflk

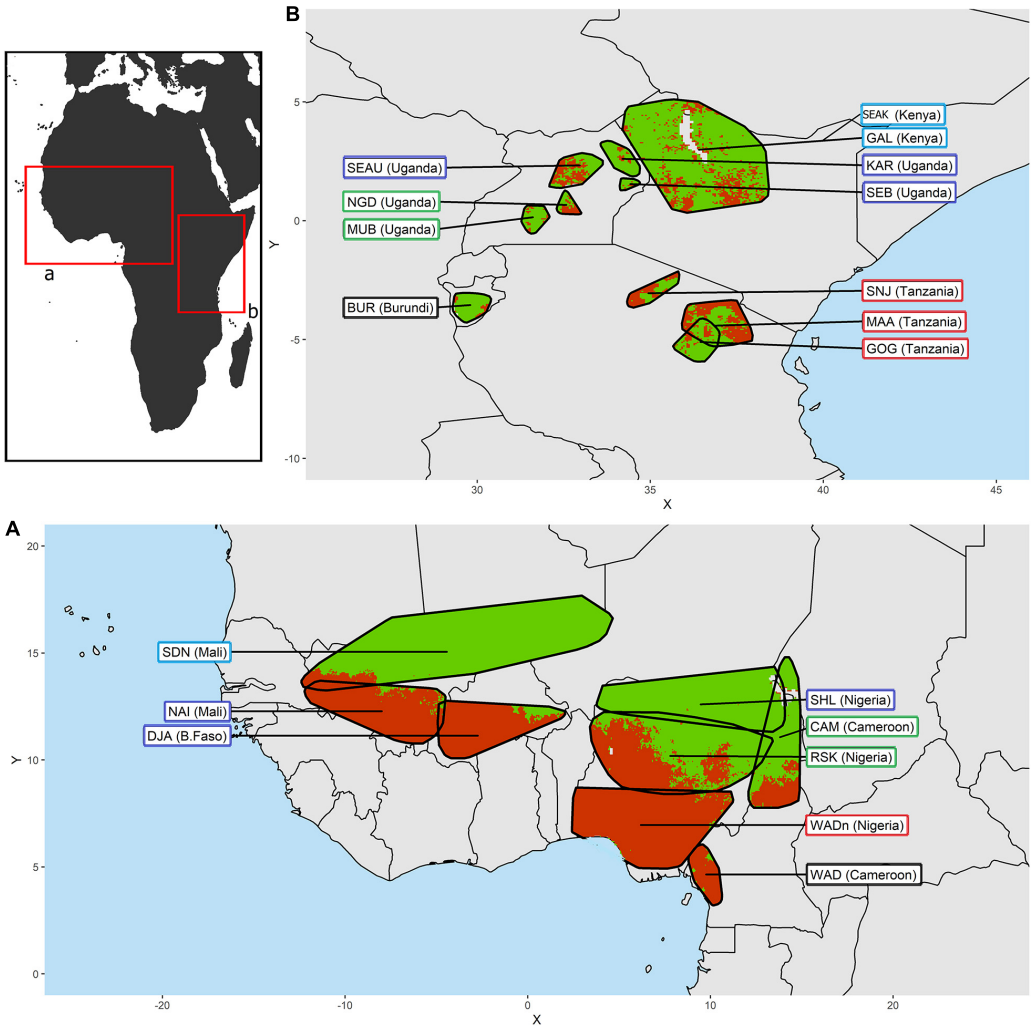

FIGURE 1 | Distribution areas for **(A)** the West-Central African goat breeds and **(B)** the East African goat breeds. In each area, the green color reflects an absence of tsetse flies while the red color identifies their presence according to Wint and Rodgers (2000). The frame of the breed name labels is colored according to the climatic clustering (see details in **Figure 2**).

CAVIAR program (Hormozdiari et al., 2014), to identify the best SNPs signals.

For the different analyses, we considered the identified outliers and applied stringent screen to determine selection signatures: we required candidate-selected regions to have at least three outliers SNPs ≤500 kb apart. For LFMM and PCAdapt we added a constraint, imposing that the three outliers (at least) showed p-value $\leq 10^{-9}$. The window chosen was informed by previous evidence that the LD in small ruminant do not exceed 500 kb (see Serranito et al., 2021).

For each analysis, genes within a region spanning 100 kb upstream and downstream of the candidate selection regions were annotated. The chromosomal regions under selection pressure were inspected using NCBI Genome Data Viewer ARS1[5].

[5]https://www.ncbi.nlm.nih.gov/genome/gdv/

RESULTS

Description of the Datasets

West-Central African Goats

In the West-Central area we observe three zones with clearly different tsetse infestation patterns (see **Figure 1A** and **Table 1**): (i) in the north, the arid zones of the Sahel are not suitable for tsetse flies, so that the Sudanese (SDN) and Sahelian (SHL) distribution areas are almost free of tsetse flies. (ii) The Red Sokoto (RSK) and the Cameroonian (CAM) inhabit areas with intermediate levels of infestations (27 and 45%, respectively). (iii) Naine (NAI), Djallonke (DJA), and WAD breeds from Nigeria and Cameroon inhabit areas with high levels of infestation (i.e., over 90%).

The PCA pattern of climate variables (**Figure 2.1**) was dominated by the first component (PC1), which accounts for 77.7% of the total variance. The distribution areas appeared

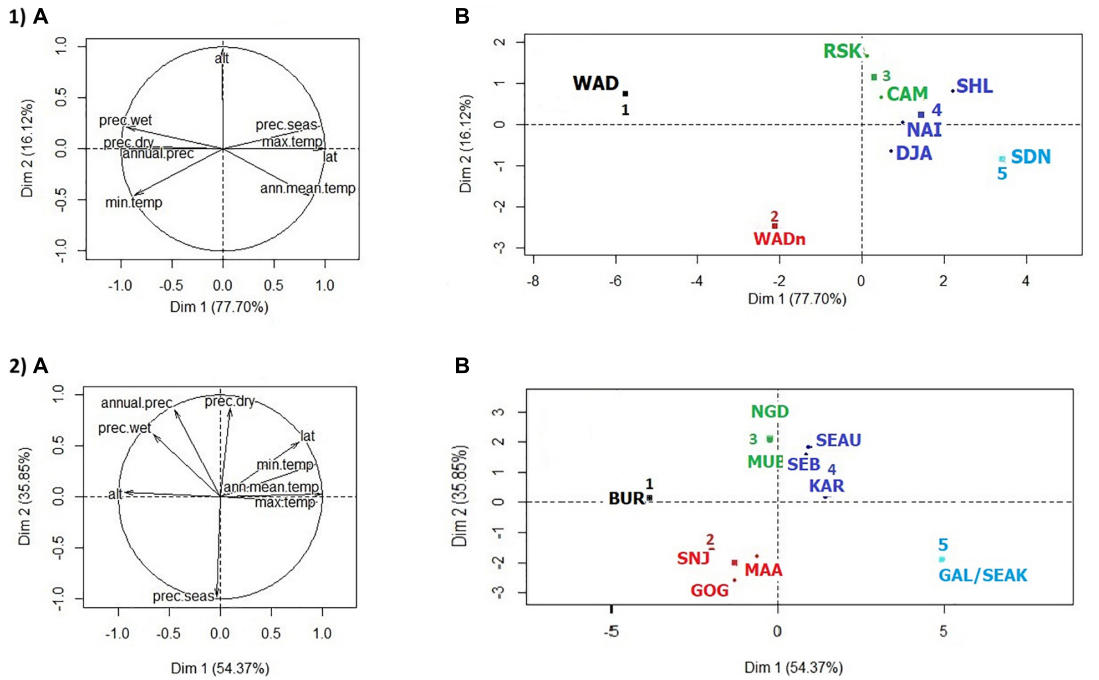

FIGURE 2 | Goat distribution areas analyses as a function of environmental variables. **(1)** HCPC analysis for the West-Central African dataset. **(A)** PCA correlation circle. **(B)** PCA score plot by breed. **(2)** HCPC analysis for the East African dataset. **(A)** PCA correlation circle. **(B)** PCA score plot by breed. Colors represent the main clusters obtained by HCPC (K = 5). PC 1 and PC 2 represent respectively principal components 1 and 2. alt., altitude in meters; temp., temperature in °C × 10; prec., precipitation in millimeters; min., minimal; max., maximal; seas., seasonality; lat., latitude; ann. mean, annual mean.

thus mainly distributed along a latitudinal South-North axis, following increasing temperature and aridity gradients. PC2 (16.1% of variance) differentiated the zones according to the average altitude. HCPC clustering at K = 5 yielded the groups of (1, 2) breeds from southern Nigeria (WADn) and Cameroon (WAD) regions with high rainfall, even during the dry season; (3) breeds at an intermediate latitudinal position and also with an intermediate level of aridity; and (4, 5) breeds subject to low rainfall levels decreasing further close to the Sahelian zone. The levels of tsetse infestation did not match exactly the climatic clustering obtained on the basis of the aridity gradient, but they were strongly related to it.

For the Admixture analysis of genetic structure (**Figure 3**), the cross-entropy criterion minimum occurred at K = 3 and K = 4. The Admixture plot distinguished at K = 3 between (i) the Cameroon breed (CAM), (ii) its more southerly compatriot (WAD), (iii) and other breeds from Nigeria, Mali and Burkina-Faso. At K = 4, the Nigerian West African Dwarf (WADn) formed a separate cluster. For K = 8, levels of admixture were observed between the DJA, NAI, and SDN group on the one hand, and the SHL and RSK group on the other; in such a way that breeds could not be individualized in each group. Finally, several Nigerian SHL individuals clustered with the West African Dwarf of Cameroon (WAD), strongly suggesting introgression.

East African Goats

The levels of tsetse infestation showed little contrast between the breed distribution areas of East Africa (see **Figure 1B** and **Table 2**). We distinguished two groups: (i) breeds in areas with low to moderate infestation rates (below 20%), and (ii) breeds for which infestation rates of living areas were around 50% (i.e., Massai MAA, Sonjo SNJ, Small East Africa from Uganda SEAU, and Nganda NGD).

In the climatic PCA analysis (**Figure 2.2**), PC1 accounted for 54.4% of the variation and PC2 for 35.9%. PC1 differentiated the areas according to altitudinal and temperature gradient. On the second axis, the areas were distinguished according to the seasonality of rainfall. HCPC clustering, considering K = 5 as the number of clusters, was as follows: group 1 concerned Burundi (BUR) with a distribution area characterized by a high average altitude and consequently low temperatures. Average temperatures and altitudes and a pronounced seasonality of rainfall characterized living areas of group 2, which included Tanzanian breeds; while group 3, including two Ugandan breeds, differed from the previous one mainly by the occurrence of rain even during the dry season. Group 4, also in Uganda, concerned areas with low seasonality as in the previous group but with higher average temperatures and correlatively lower average altitudes. Finally, group 5 included the Kenyan breeds and was characterized by high average temperatures and seasonal rainfall.

Concerning the analysis of the genetic structure (**Figure 4**), the cross-entropy criterion minimum occurred at K = 2, indicating a major divergence of the Burundi breed (BUR) and the other breeds. For K = 3 and K = 4 sub-structuring was shared by (i) breeds from Tanzania and Kenya KAR and SEB from Uganda; (ii) the three other breeds from Uganda (SEAU, MUB, and NGD).

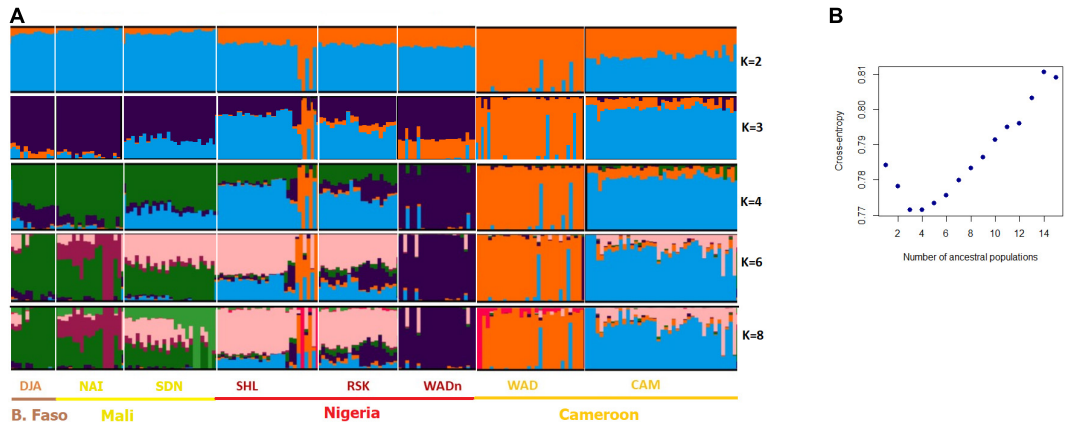

FIGURE 3 | Bayesian clustering performed with ADMIXTURE software on the West/Central African goats. **(A)** Display for different values of *K*, with *K* = number of clusters; **(B)** cross-entropy plot for the number of cluster *K* = 1–15.

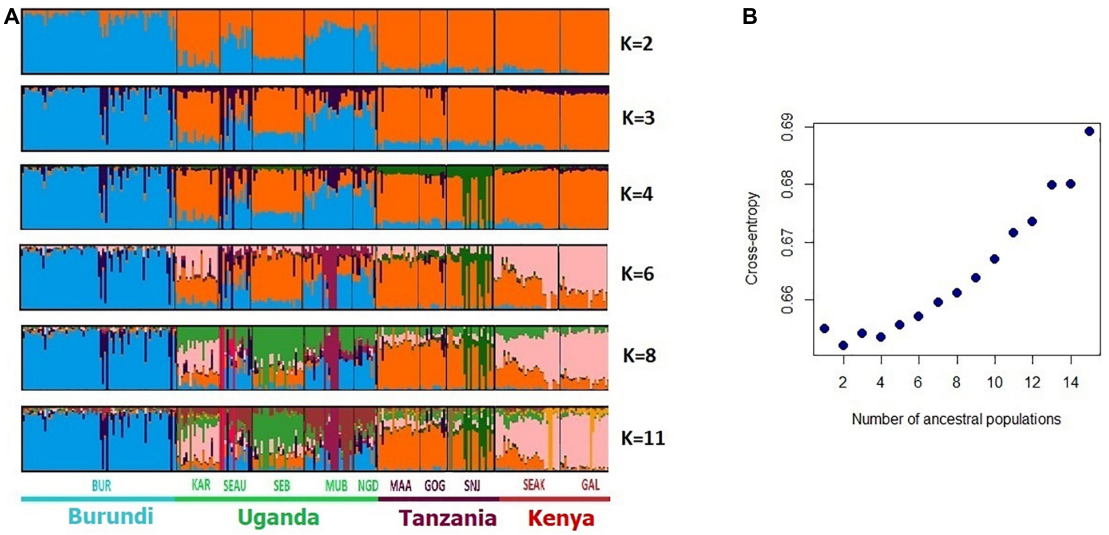

FIGURE 4 | Bayesian clustering performed with ADMIXTURE software on the East African goats. **(A)** Display for different values of *K*, with *K* = number of clusters; **(B)** cross-entropy plot for the number of cluster *K* = 1–15.

At *K* = 11, the breeds of Uganda, Tanzania, and Kenya still did not form separate clusters. High levels of admixture were observed between (i) SEAK, GAL, and KAR breeds, combining breeds from Kenya and Uganda, (ii) between the Ugandan SEAU, MUB, and NGD breeds, and (iii) between the Tanzanian, MAA, GOG, and SNJ breeds.

Selection Signatures in West and Central African Goats

The LFMM analysis (**Table 3**, see details in **Supplementary Figure 3** and **Supplementary Table 3**) based on the correlation with tsetse flies' infestations ranking (**Table 1**) allowed the detection of five selection signatures. Among them, the signature near the *Neurobeachin* (*NBEA*) gene was also identified by Bayescan, Hapflk, and the LFMM analysis based on the correlation with the PCA/HCPC climatic ranking (**Figure 2.1**, see details in **Supplementary Table 4** and **Supplementary Figure 4**). Moreover, it was found to be associated with temperature adaptation in sheep and goats around the world (see details in section "Discussion"). From the other four signatures, the signature near the *SLC34A2* gene was previously associated with high altitude conditions in cattle (Verma et al., 2018), and identified by Serranito et al. (2021) as associated with climatic adaptation in Mediterranean sheep and goat. *SOCS2* was found associated with hypoxia condition in goat and sheep (Wang et al., 2016; Yang et al., 2016). This leaves two signatures, potentially related to trypanosomosis tolerance and targeting the genes: *DIS3L2, COPS7B, RAPGEF6*, and *MEIKIN*, respectively. The latter signature was also detected by LFMM analysis correlating genomes and PCA/HCPC climatic ranking (see details in **Supplementary Table 4** and **Supplementary Figure 4**). This

TABLE 3 | Selection signatures identified in the West-Central African goat and East-African goat by LFMM following the infestation ranking and PCAdapt.

West-Central African goat dataset					East African goat dataset				
Chr	**Signature range**	**PCAdapt**	**LFMM climatic ranking**	**Gene(s) most likely under selection**	**Chr**	**Signature range**	**PCAdapt**	**LFMM climatic ranking**	**Gene(s) most likely under selection**
2	15,786,051–16,322,293			**DIS3L2**/COPS7B	2	None			
5	22,987,063–23,578,233			**CRADD**/SOCS2	5	None			
6	45,415,286–45,625,340			**SLC34A2**	6	None			
	59,436,318–59,715,618			**PDS5A**/UBE2K		None			
	None					85,978,469–86,198,470			**CSN1S1/CSN1S2**
7	59,624,626–60,848,321			**RAPGEF6/**MEIKIN[3]	7	None			
10	64,438,812–65,229,754			**UBR1**[1,2]	10	None			
11	None				11	94,156,441–94,498,221			**DENND1A**
12	44,050,377–44,471,339			**KLHL1/PCDH9**	12	43,632,105–45,137,234			**KLHL1/PCDH9**
	47,433,756–47,769,497			**PCDH9**		None			
	50,152,165–50,676,638			**RNF17**/PARP4/GJB2/ZMYM2[1]		50,047,466–50,515,766			**RNF17**
	60,710,546–60,950,668			**NBEA**[1,2,3]		None			
13	None				13	46,458,560–46,473,488			**PRNP**
16	None				16	43,284,994–43,405,682			**RERE**
17	None				17	36,231,019–36,715,246			**SPATA5**
18	None				18	36,786,664–37,006,519			**NFATC3**
19	32,969,117–33,307,741			**NCOR1**/LOC102188626/PIGL	19	32,913,594–33,307,741			**NCOR1/LOC102188626**/PIGL
20	None				20	14,021,075–14,132,568			**ADAMTS6**
21	42,315,959–42,638,483			**AKAP6**	21	42,193,046–42,466,574			**AKAP6**
	None					57,303,973–57,544,623			**BTBD7/UNC79/UBR7**
22	None				22	33,429,708–34,062,919			**SUCLG2**
24	None				24	48,936,361–49,308,788			**DYM**
25	12,862,316–13,062,061			**MKL2**	25	12,862,316–13,062,061			**MKL2**
26	28,957,011–29,548,698			**FBXW4/BTRC**/C26H10orf76	26	28,717,972–29,548,698			**FBXW4/BTRC**/C26H10orf76
	None					51,077,585–51,356,112			**UBE2D1**

Each signature was named according to the gene(s) most likely targeted by the selection process, taking into account the distance of the SNPs from the genes, the associated p-values, and the number of SNPs near or in the gene(s); in bold, the most likely targeted genes when there are several. All annotated genes associated with each selection signature are highlighted in ***Supplementary Table 3****.*

[1]*Identified by Bayescan.*

[2]*Identified by HAPFLK.*

[3]*Identified by LFMM considering the climatic gradient.*

Signatures underlined were not identified in the literature as linked to climatic adaptation. Chr.: chromosome number.

analysis detected a total of 10 selection signatures potentially correlated to climatic conditions, including four signatures on chromosome 7.

PCAdapt (**Table 3**, see details in **Supplementary Table 3**) identified 10 selection signatures, which include *NBEA* already found by LFMM. From the other nine signatures, those including genes *PCDH9, KLH1, RNF17, NCOR1,* and *BTRC* have been found in the literature as associated with climatic conditions (Wang et al., 2015; Kim et al., 2016; Mdladla, 2016; Yang et al., 2016; Mwacharo et al., 2018; Serranito et al., 2021). This leaves four signatures, targeting *PDS5A* and *UBE2K, AKAP6, MKL2*, and *UBR1*, that are potentially related to trypanotolerance. This last signature was also identified by Bayescan and Hapflk/Caviar. The Hapflk population tree constructed in the *UBR1* region under selection (**Supplementary Figure 4**) indicated that Cameroonian breeds (WAD and CAM) were particularly targeted by the selected haplotypes.

Selection Signatures in East African Goats

The analyses for this dataset did not detect any selection signatures *via* the LFMM analyses either, by the correlation with the infestation ranking (**Table 2**) or by the correlation with the PCA/HCPC climatic ranking (**Figure 2.2**). PCAdapt (**Table 3**, see details in **Supplementary Table 3** and **Supplementary Figure 5**) allowed the detection of 17 selection signatures, nine of which were previously detected in the literature as associated with agronomic traits, *CSN1S1* and *CNS1S2* (see Martin et al., 2002 for a review), disease/immunity, *PRNP* (see Greenlee, 2019 for a review), and *NFATC3* (Minematsu et al., 2011) or climate variations. In detail, signatures targeting *KLH1, PCDH9, RNF17, NCOR1,* and *BTRC* and potentially implicated in climate adaptation were also found in the West-Central dataset, while *ADAMT6* (Wu et al., 2019; Serranito et al., 2021), and *SUCLG2* (see details in section "Discussion") were only found in the East dataset. This leaves the signatures near *DENND1A, SPATA5, RERE, BTBD7* with *UNC79* and *UBR7, DYM, UBE2D1, AKAP6*, and *MKL2*, respectively, as candidates for trypanotolerance; the last two were also identified in the West-Central African dataset.

DISCUSSION

We studied two datasets including goat breeds with different degrees of trypanotolerance from West-Central and East Africa, respectively, in order to identify selection signatures potentially related to trypanosome tolerance.

Relevance of Datasets

The datasets appeared to differ in several aspects in terms of environment and molecular-genetic patterns. West-Central African breeds showed contrasting habitats in terms of tsetse flies infestation correlating with the PCA/HCPC climatic ranking, and following mainly the humidity and temperature gradients. Indeed, this is the main determinant of land cover that provides suitable habitat for tsetse flies (Cecchi et al., 2008), even if modified by for instance riverine vegetation and local habitat destruction or fragmentation (Bouyer et al., 2015).

In contrast, the East African breeds inhabit areas with more moderate or patchier levels of tsetse flies' infestation. The PCA climatic analysis highlighted areas of goat breeds distribution tangled between a temperature gradient on the one hand, largely influenced by altitude, and the seasonality of rainfall on the other hand. Areas of high altitudes that are found in East Africa (i.e., above roughly 2,000 m) are unsuitable for tsetse flies' survival (Slingenbergh, 1992; Cecchi et al., 2015). In addition, climatic zones are globally drier than in West and Central Africa and harbor different tsetse species with also different ecological requirements (Cecchi et al., 2008).

Our molecular analysis showed for the West-Central African dataset breeds that in spite of varying degrees of admixtures have kept their genetic identity. In East Africa only goats in Burundi were clearly distinct from other goats, while goats from Uganda, Tanzania, and Kenya were largely admixed.

The history of pastoralism has been documented in Kenya and northern Tanzania from ∼3300 BP (Marshall et al., 2011; Gifford-Gonzalez, 2017; Grillo et al., 2018). For several millennia the Bantu, Nilotic, and Khoisan-speaking tribes have evolved in these environmentally contrasting regions of East Africa, allowing the emergence of a wide diversity of goat breeds adapted to the environmental conditions (Mason, 1969; Ahuya, 1997). The colonial period induced profound shifts in traditional pastoral patterns (Masefield, 1962; Lwanga-Lunyiigo, 1987; Chacha, 1999). In particular, it has been tried to improve the productivity of the goats, essentially made up of small breeds of the Small East African type. In the 1980s, a turning point was observed in East Africa with a strong encouragement to crossbreed with imported exotic breeds, including Boer, Kamorai, Toggenburg, Saanen, Norwegian, Alpine, and Anglo-Nubian (Das and Sendalo, 1991). Major programs were set up at various levels and involved: religious organizations, government institutions, non-governmental organizations, such as Heifer International, Bill and Melinda Gates Foundation, British-Farm Africa, Livestock Production Programme, German-GTZ, etc., and academic research institutes (Wilson et al., 1990; Bill and Melinda Gates Foundation (BMFG), 2013; Mruttu et al., 2016). For instance, the "Small Ruminant – Collaborative Research Support Programme" was established in 1980 by the government of Kenya and the United States of America International Development Agency. The objectives were to establish a dual purpose goat (DPG) by a four-way cross of Toggenburg, Galla, Anglo-Nubian, and Small East Africa (Ruvuna, 1986), and associated production systems, in order to increase milk and meat production in western Kenya (Rewe et al., 2002; Bett et al., 2007). In Tanzania, in 1996, another DPG project used "Blended goats" (Kamorai 55%, Boer 30%, and indigenous Tanzania Goats 15%) × Anglo-Nubian which were transferred to smallholder farmers (Shirima, 2005).

Upgrading local goats by cross-breeding is still the prevailing strategy in East Africa, in spite of the limited results obtained in terms of productivity and the genetic erosion caused to the native gene pool (Ahuya, 1997; Gichohi, 1998; Ayalew et al., 2003;

Onzima, 2014; FAO, 2015; Mruttu et al., 2016; Wilson, 2018). These considerations provide an understanding of the levels of admixture observed, which are corroborated by additional admixture analyses (see **Supplementary Text 1**).

In the East African dataset, the non-identification of selection signatures related to the climatic environment by the LFMM method was unexpected, especially in view of the strongly contrasting environments, in particular in terms of altitude and the finding of 10 environmental selection signatures in the West-Central African dataset. The East-African dataset differs from the West-Central one, in terms of admixture levels. It can be hypothesized that cross-breeding with exotic breeds populations has obscured the link between genome and environment, but this remains to be tested with larger datasets.

Moreover, the East African dataset did not reveal, *via* LFMM, any selection signatures related to the infestation rate by tsetse flies. Given the low contrasting levels of infestation across the breeds, this may be explained by a lack of statistical power. However, the high level of admixture may also have played key role. It has been reported that trypanosome tolerance levels were reduced in the descendants of crosses between trypanotolerant East-African goat and exotic breeds (Griffin and Allonby, 1979a,b), although this was not observed with Kenyan breeds (Whitelaw et al., 1985).

For the West-Central Africa region, the oldest goat remains date from about 2300 BC to 1900 BC (Blench et al., 2006). The West African Dwarf type has diversified into heterogeneous populations in tropical and equatorial areas and has developed trypanotolerance, in contrast with the Sahelian type, which has evolved in dry to arid environments (Adah et al., 1993; Bengaly et al., 1993; Osaer et al., 1994; Geerts et al., 2009). Several authors reported that crosses between West African Dwarf type and Sahelian goats are common over the last few decades (Trail et al., 1980; Blench, 1999; Hoeven et al., 2006). Our admixture analysis indicated that this has particularly affected the dwarf breeds of Burkina-Faso and Mali. Goossens et al. (1999) showed a decrease in trypanotolerance for crosses between West African Dwarf type and Sahelian type. Introgression of non-trypanotolerant genes in West African Dwarf types may explain the limited power of our analysis, which identified *via* the LFMM method only five selection signatures related to the level of infestation.

Finally, we note that the Bayescan and Hapflk softwares, which are sensitive to admixture, only showed a weak number of signatures and were not particularly suitable to these datasets, with local heterogeneous breeds. Moreover, it is interesting to note that for the East African dataset with the higher admixture levels and less outspoken trypanotolerance contrast, only PCAdapt retained enough power to identify selection signatures.

Selection Signatures

In all, considering the two datasets, with the PCAdapt and LFMM methods according to the climatic ranking and the level of infestation ranking, 41 selection signatures were identified, 33 of which were common to both datasets or were highlighted by different methods. Of these signatures, 18 have been detected previously in the literature, as mostly linked to climatic adaptation; and in particular, they were all identified in Serranito et al. (2021), except the signature targeting *SOCS2*.

Selection Signatures Unlikely to Be Trypanosome-Related

Considering the 18 signatures in detail, we identified *via* all four method, one selection signature near the *NBEA* gene in West and Central African goats. *NBEA* was reported to be associated with high altitude in Ethiopian sheep (Edea et al., 2019), Chinese sheep (Yang et al., 2016), cattle at high altitude (Zheng and Shabek, 2017), and yaks (Qi et al., 2019). Furthermore, it was found under selection in cattle (Howard et al., 2013), in Ugandan and Moroccan goats (Onzima et al., 2018; Benjelloun, 2015) and in Mediterranean sheep and goats (Serranito et al., 2021). This latest study postulates that this gene may play a predominant role in climate adaptation.

Four of the other 17 selection signature were found in both datasets: (i) *FBXW4* and *BTRC* involved in lipid metabolism (Ishimoto et al., 2017); (ii) *NCOR1* belonging to clock circadian gene network in cattle (Wang et al., 2015); (iii) *RNF17* associated with fatty acid composition (Lemos et al., 2016), growth traits (Puig-Oliveras et al., 2014; Edea et al., 2018) and adaptation to climate variables in South African goats (Mdladla, 2016); and (iv) *KLH1* and *PCDH9*, a major signature related to adaptation to aridity (Kim et al., 2016; Serranito et al., 2021).

Most of the remaining signatures were found only in the East African dataset: *PRNP* linked to susceptibility to prion disease in small ruminants (see Greenlee, 2019 for a review), *NFATC3* involved in immunity (Minematsu et al., 2011), the milk genes *CSN1S1* and *CSN1S2* (see Martin et al., 2002), and two genes implicated in altitude adaptation, *ADAMTS6* and *SUCLG2*. This is not surprising as the East African breeds under consideration live at mean altitudes varying from 590 meters in Kenya to 1,633 m in Burundi (**Supplementary Table 5**). *ADAMTS6* is involved in porcine growth traits (Wu et al., 2019) and is associated with high altitude adaptation in Chinese sheep; whereas *SUCLG2* was found to be associated with climatic conditions in Egyptian, Chinese, and Mediterranean sheep (Kim et al., 2016; Yang et al., 2016; Serranito et al., 2021), but also in yaks (Qi et al., 2019). Finally, *SLC34A2* implicated in production and recycling of breath surfactant, in humans (Ma et al., 2018), was identified in the West-Central goat dataset, together with *CRADD* and *SOCS2*. The *SOCS2* protein was found associated with hypoxia condition, in Chinese goat and Tibetan sheep (Wang et al., 2016; Yang et al., 2016), but has also been shown to modulate the innate and adaptive immune response in various experimental models of infection, including *Toxoplasma gondii*, *Trypanosoma cruzi*, and *Plasmodium berghei ANKA* (Machado and Aliberti, 2006; Esper et al., 2012; Machado et al., 2012; Brant et al., 2016).

Selection Signatures Potentially Linked to Trypanosomosis Tolerance

The most promising signatures have all been identified in the West and Central African dataset and included the following genes: *DIS3L2, COPS7B, PD5A, UBE2K*, and *UBR1*. There is no

literature on *AKAP6* and *MKL2*, but these signatures found in both datasets could indicate a significant role for these genes.

The selection signature potentially targeting *UBR1* (*E3 ubiquitin-protein ligase component n-recognition*) was highlighted by PCAdapt, Bayescan, and Hapflk/Caviar. E3 ubiquitin-protein ligases are involved in many cellular processes in eukaryotes (Zheng and Shabek, 2017). Moreover, the E3-Ubiquitin ligase MARCH1 was shown to regulate type I interferon signaling, T cell activation, and IFN-γ production during malaria infections and was proposed to be a key molecule in immune responses and a potential target for immunotherapies (Wu et al., 2020). Furthermore, Kim et al. (2017) identified a selection signature targeting *EPB42*, which is close to *UBR1* and is involved in pathways controlling anemia, in the trypanotolerant N'Dama cattle.

Interestingly, in addition to *UBE2K* and *UBR1*, also the ubiquitins *UBE2D1* and *UBR7* have been indicated by our study, as being subject to selection. Moreover, the degradation of *NFATC3*, also identified by this study, which plays key role in immunity, is ubiquitin-dependent. Indeed, ubiquitination plays a central role in the regulation of various biological functions including immune responses (Hu and Sun, 2016).

The human *COPS7B* protein level has been linked to the invasion efficiency of *T. gondii* (Ready, 2013). *DIS3L2*, located near *COPS7B*, is an RNA-binding protein with 3′-5′ exoribonuclease activity and plays an important role in cytoplasmic RNA surveillance and decay. Becker et al. (2020) suggested a potential link between the gene *DIS3L2* and the gastrointestinal nematode resistance in sheep and postulated that *DIS3L2* polymorphisms influence its degradation functions at the oligoU tail and hence modulate immune response to parasite infection. *DIS3L2* has also been identified, among genes involved in cancer, cellular function and maintenance, and neurological disease, in West African cattle (Gautier et al., 2009). In Brazilian sheep, De Simoni Gouveia et al. (2017) found it associated with height variation.

A selection signature near the genes *PD5A* and *UBE2K* has previously been associated with the bovine resistance to endoparasites and more specifically with the antibody response to *Fasciola hepatica* (Kim et al., 2015; Twomey et al., 2019). Capalbo et al. (2012) suggested a role of *PDS5A* in HIV-1-induced cellular pathogenesis. The gene *UBE2K* was found associated with the immune response to a malaria-like parasite in a wild primate (Trujillo and Bergey, 2020).

We were able to identify selection signatures associated with climatic variables and the presence of tsetse flies, the cyclical vector of trypanosomes. Like tsetse flies whose geographic distribution depends on land cover and climate, other selective pressures are also associated with environmental factors and could thus overlap at least partly with nagana. Especially, numerous parasites and pathogenic agents are differentially distributed according to geographic and climatic patterns in Sub-Saharan Africa, like gastro-intestinal parasites (Bolajoko and Morgan, 2012), ticks (Walker et al., 2003; Silatsa et al., 2019) and *Cowdria ruminantium* (Esemu et al., 2011) transmitted by *Amblyomma* ticks. Genetic tolerance of goats to haemonchosis (Chiejina et al., 2015; Silva et al., 2018) and heartwater (Bambou et al., 2010; Donkin et al., 2015) has been reported. Therefore, the highlighted signatures cannot yet be associated exclusively with any of these diseases. Whereas co-infections between different pathogens are highly prevalent in Sub-Saharan Africa (Kusiluka and Kambarage, 1996), infection by a pathogen can modulate, in a positive or negative way, the host's response to subsequent infections by other pathogens, suggesting complex interacting processes (Faye et al., 2002; Gonçalves et al., 2014). Moreover, some trade-offs could exist, since contrary to response to trypanosomosis, WAD goats could be more susceptible to Peste des Petits Ruminants, caused by a morbillivirus, than Sahelian populations (Diop et al., 2005). In any case, several genes targeted by selection signatures have been reported throughout the literature to be involved in immune mechanisms concerning different pathogens, mainly nematodes, platyhelminthes, or protozoan parasites.

CONCLUSION

We observed a number of selection signatures that could be potentially related to trypanosomosis tolerance in goats. The most consistent signature appears to involve the ubiquitination by the *UBR1* gene, highlighted by all the methods used. Furthermore, the analysis of these two populations highlights that (i) the environmental conditions in West and Central Africa were characterized by a clear gradient related to the climate and land cover; and (ii) that West and Central African goats have remained relatively protected from inter-breeding crosses. These elements make this area of Africa a region of choice to study selection signatures linked to trypanosomosis. It will therefore be of particular interest to deepen these analyses by genotyping other breeds living in the different areas of contrasting parasitic pressures and to document epidemiological and environmental factors and breeding systems as well. All this emphasizes the importance of preserving indigenous stocks and their adaptation as legacy of thousands of years of evolution, instead of pursuing programs with short-term productive objectives. Finally, extending these analyses to sheep breeds will be most enlightening.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: genotyping data collected by AdaptMap (http://www.goatadaptmap.org/) and shared on Dryad (https://doi.org/10.5061/dryad.v8g21pt).

AUTHOR CONTRIBUTIONS

BS, SH, AL, N-AO-K, and BB contributed to the data analyses. DT-M performed the historical analyses. GC, ST, FP, and JAL contributed to the review and editing. AD designed the research, analyzed the data, and wrote the manuscript. All authors contributed to the final manuscript.

FUNDING

This work was supported by the AAP CNRS 2020: "Adaptation du Vivant à son Environnement" (EMPREINTES project, 2020-2022). FAO support to this study was provided in the framework of the Programme against African Trypanosomosis (PAAT).

ACKNOWLEDGMENTS

We would like to thank the CALI calculator of the University of Limoges, funded by the Limousin region, the European Union, the XLIM institutes, IPAM, GEIST, and the University of Limoges.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.715732/full#supplementary-material

Supplementary Figure 1 | NeighborNet graph considering African local goat breeds, from a matrix of Reynolds' distances.

Supplementary Figure 2 | PCAdapt screen plots, **(a)** considering the West-Central African goat dataset, **(b)** considering the East African goat dataset.

Supplementary Figure 3 | Display of the selection signatures identified in the West-Central African goat dataset. The figures have been modified from outputs obtained *via* NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/): ARS1. SNPs under selection (see their name and position in **Supplementary Table 3**) are symbolized by stars; red stars: SNPs identified by PCAdapt (and plots associated), green stars: SNPs identified by LFMM following the infestation ranking (and plots associated); Hapflk plot and tree displays.

Supplementary Figure 4 | Display of the selection signatures identified in the West-Central African goat dataset *via* LFMM following the climatic PCA/HCPC ranking. The figures have been modified from outputs obtained *via* NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/), assembly ARS1. SNPs under selection (see their name and position in **Supplementary Table 4**) are symbolized by stars.

Supplementary Figure 5 | Display of the selection signatures identified in the East African goat dataset. The figures have been modified from outputs obtained *via* NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/): ARS1. SNPs under selection (see their name and position in **Supplementary Table 3**) are symbolized by stars; red stars: SNPs identified by PCAdapt (and plots associated); green stars: SNPs identified by LFMM following the infestation ranking (and plots associated).

Supplementary Table 1 | Initial list of breeds for the constitution of the datasets.

Supplementary Table 2 | Presentation of the bioclimatic variables used.

Supplementary Table 3 | Details concerning the selection signatures identified in West-Central African and East African goats: name and position of the SNPs, genes identified according to PCAdapt, LFMM following the infestation ranking, Bayesca, and Hapflk/Caviar.

Supplementary Table 4 | Details concerning the selection signatures identified in West-Central African goats: name and position of the SNPs, genes identified according to LFMM following the climatic PCA/HCPC ranking.

Supplementary Table 5 | Summary of environmental data extracted for the distribution area of the West-Central African and East-African goat breeds.

Supplementary Text 1 | East African admixture analysis, including exotic breeds and crossbred individuals.

REFERENCES

Adah, M. I., Otesile, E. B., and Joshua, R. A. (1993). Susceptibility of Nigerian west African dwarf and red Sokoto goats to a strain of Trypanosoma congolense. *Vet. Parasitol.* 47, 177–188. doi: 10.1016/0304-4017(93)90020-n

Agyemang, K. (2005). Trypanotolerant livestock in the context of trypanosomosis intervention strategies. *PAAT Sci. Techn. Ser.* 7, 1–66.

Ahuya, C. O. (1997). "Community based goat improvement project in Meru Central and South districts: The FARM-Africa experiences," in *Proceedings of Goat Development in Eastern Africa, Workshop, 8-11th December, 1997, Izaak Walton Inn, Embu, Kenya*, eds C. O. Ahuya and H. Van Houton (FARM-Africa, Nairobi, Kenya).

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

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Alsan, M. (2015). The effect of the tsetse fly on African development. *Am. Economic Rev.* 105, 382–410. doi: 10.1257/aer.20130604

Ayalew, W., Rischkowsky, B., King, J. M., and Bruns, E. (2003). Cross-breds did not generate more net benefits than indigenous goats in Ethiopian smallholdings. *Agricult. Syst.* 76, 1137–1156. doi: 10.1016/s0308-521x(02)00033-1

Bambou, J. C., Vachiery, N., Despois, P., Giraud-Girard, K., Arquet, R., Pinarello, V., et al. (2010). Assessment of genetic variability of resistance to heartwater in Creole goats. *Adv. Anim. Biosci.* 1, 408–409. doi: 10.1017/S2040470010000361

Becker, G. M., Davenport, K. M., Burke, J. M., Lewis, R. M., Miller, J. E., Morgan, J. L. M., et al. (2020). Genome-wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. *Animal Genetics* 51, 330–335. doi: 10.1111/age.12895

Bengaly, Z., Clausen, P.-H., Boly, H., Kanwé, A., and Duvallet, G. (1993). *Comparaison de la trypanosomose expérimentale chez certaines races de petits ruminants au Burkina Faso. Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux. (Burkina Faso)*. France: Université Grenoble.

Benjelloun, B. (2015). *Diversité Des Génomes Et Adaptation Locale Des Petits Ruminants D'un Pays Méditerranéen: Le Maroc*. Ph.D thesis. France: Université Grenoble

Berthier, D., Brenière, S. F., Bras-Gonçalves, R., Lemesre, J.-L., Jamonneau, V., Solano, P., et al. (2016). Tolerance to trypanosomatids: A threat, or a key for disease elimination? *Trends Parasitol.* 32, 157–168. doi: 10.1016/j.pt.2015.11.001

Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., et al. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* 50:57. doi: 10.1186/s12711-018-0421-y

Bett, R. C., Kosgey, I. S., Bebe, B. O., and Kahi, A. K. (2007). Breeding goals for the Kenya dual purpose goat. I. Model development and application to smallholder production systems. *Trop. Anim. Health Product.* 39, 477–492. doi: 10.1007/s11250-007-9015-3

Bill and Melinda Gates Foundation (BMFG) (2013). *Livestock Genetics Investment Strategy Development; Tanzania Livestock Genetics Convening Workshop, Hosted By The Bmgf, 18–20 September 2013*. Arusha: Mount Meru Hotel.

Blench, R. (1999). *Traditional Livestock Breeds: Geographical Distribution And Dynamics In Relation To The Ecology Of West Africa. Working Paper 122*. London: Overseas Development Institute.

Blench, R., MacDonald, K., and MacDonald, K. (2006). *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography*. London: Routledge.

Bolajoko, M. B., and Morgan, E. R. (2012). Relevance of improved epidemiological knowledge to sustainable control of Haemonchus contortus in Nigeria. *Animal Health Res. Rev.* 13, 196–208. doi: 10.1017/S1466252312000163

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* 186, 241–262. doi: 10.1534/genetics.104.117275

Bouyer, J., Dicko, A. H., Cecchi, G., Ravel, S., Guerrini, L., Solano, P., et al. (2015). Mapping landscape friction to locate isolated tsetse populations that are candidates for elimination. *Proc. Natl. Acad. Sci.U. S A.* 112, 14575–14580. doi: 10.1073/pnas.1516778112

Brant, F., Miranda, A. S., Esper, L., Gualdrón-López, M., Cisalpino, D., de Souza, D., et al. (2016). Suppressor of cytokine signaling 2 modulates the immune response profile and development of experimental cerebral malaria. *Brain Behav. Immunity* 54, 73–85. doi: 10.1016/j.bbi.2016.01.002

Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. *Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie* 19, 23–32.

Budd, L. (1999). *Dfid-Funded Tsetse And Trypanosome Research And Development Since 1980: Economic Analysis*, Vol. Vol. 2. Aylesford: DFID Livestock Production, 123.

Capalbo, G., Müller-Kuller, T., Ottmann, O. G., Hoelzer, D., and Scheuring, U. J. (2012). HIV-1 infection suppresses expression of host cell cycle-associated gene PDS5A. *Intervirology* 55, 263–275. doi: 10.1159/000328323

Cecchi, G., Mattioli, R. C., Slingenbergh, J., and Rocque, S. D. L. (2008). Land cover and tsetse fly distributions in sub-Saharan Africa. *Med. Vet. Entomol.* 22, 364–373. doi: 10.1111/j.1365-2915.2008.00747.x

Cecchi, G., Paone, M., Argilés Herrero, R., Vreysen, M. J. B., and Mattioli, R. C. (2015). Developing a continental atlas of the distribution and trypanosomal infection of tsetse flies (Glossina species). *Parasites Vect.* 8:284. doi: 10.1186/s13071-015-0898-y

Chacha, B. K. (1999). *Agricultural history of the abakuria of kenya from the end of the nineteenth century to the mid 1970's. History. 1999*. Aylesford: DFID Livestock Production.doi: dumas-01277386

Chiejina, S. N., Behnke, J. M., and Fakae, B. B. (2015). Haemonchotolerance in West African Dwarf goats: contribution to sustainable, anthelmintics-free helminth control in traditionally managed Nigerian dwarf goats. *Parasite* 22:7. doi: 10.1051/parasite/2015006

Das, S. M., and Sendalo, D. S. C. (1991). *Small Ruminant Research Highlights In Tanzania*. Dar es Salaam: Ministry of Agriculture and Livestock Development, 40.

De Simoni Gouveia, J. J., Paiva, S. R., McManus, C. M., Caetano, A. R., Kijas, J. W., Facó, O., et al. (2017). Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livestock Sci.* 197, 36–45. doi: 10.1016/j.livsci.2017.01.006

Diop, M., Sarr, J., and Libeau, G. (2005). Evaluation of novel diagnostic tools for peste des petits ruminants virus in naturally infected goat herds. *Epidemiol. Infect.* 133, 711–717. doi: 10.1017/s0950268805003729

Donkin, E., Stewart, C. G., MacGregor, R. G., Els, H. C., Boyazoglu, P. A., and Ramsay, K. (2015). Resistance of Indigenous and Crossbred goats to heartwater (Cowdria ruminantium). *Epidemiol. Infect.* 33:9844. doi: 10.13140/RG.2.1.4323.9844

Edea, Z., Dadi, H., Dessie, T., and Kim, K.-S. (2019). Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes Genomics* 41, 973–981. doi: 10.1007/s13258-019-00820-y

Edea, Z., Jeoung, Y. H., Shin, S.-S., Ku, J., Seo, S., Kim, I.-H., et al. (2018). Genome-wide association study of carcass weight in commercial Hanwoo cattle. *Asian Austr. J. Anim. Sci.* 31, 327–334. doi: 10.5713/ajas.17.0276

Esemu, S. N., Ndip, L. M., and Ndip, R. N. (2011). Ehrlichia species, probable emerging human pathogens in sub-Saharan Africa: environmental exacerbation. *Rev. Environ. Health* 26, 269–279. doi: 10.1515/reveh.2011.034

Esper, L., Roman-Campos, D., Lara, A., Brant, F., Castro, L. L., Barroso, A., et al. (2012). Role of SOCS2 in modulating heart damage and function in a murine model of acute Chagas disease. *Am. J. Pathol.* 181, 130–140. doi: 10.1016/j.ajpath.2012.03.042

ESRI (2011). *ArcGIS Desktop: Release 10*. Redlands: Environmental Systems Research Institute.

FAO (2015). *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*. Rome: FAO.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics* 193, 929–941. doi: 10.1534/genetics.112.147231

Faye, D., Osaer, S., Goossens, B., Van Winghem, J., Dorny, P., Lejon, V., et al. (2002). Susceptibility of trypanotolerant West African Dwarf goats and F1 crosses with the susceptible Sahelian breed to experimental Trypanosoma congolense infection and interactions with helminth infections and different levels of diet. *Vet. Parasitol.* 108, 117–136. doi: 10.1016/s0304-4017(02)00184-x

Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221

Foll, M., Fischer, M. C., Heckel, G., and Excoffier, L. (2010). Estimating population structure from AFLP amplification intensity. *Mole. Ecol.* 19, 4638–4647. doi: 10.1111/j.1365-294X.2010.04820.x

Frichot, E., and François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evolut.* 6, 925–929. doi: 10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and Francois, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973–983. doi: 10.1534/genetics.113.160572

Frichot, E., Schoville, S. D., Bouchard, G., and François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mole. Biol. Evolut.* 30, 1687–1699. doi: 10.1093/molbev/mst063

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genom.* 10:550. doi: 10.1186/1471-2164-10-550

Geerts, S., Osaer, S., Goossens, B., and Faye, D. (2009). Trypanotolerance in small ruminants of sub-Saharan Africa. *Trends Parasitol.* 25, 132–138. doi: 10.1016/j.pt.2008.12.004

Gichohi, C. M. (1998). "Overview of the goat industry in Kenya: - strategies and the way forward," in *Goat Developments In East Africa:- Practical Experiences And The Way Forward*, eds C. O. Ahuya and H. Van Houton (Rome: FAO).

Gifford-Gonzalez, D. (2017). "Animal disease challenges" fifteen years later: The hypothesis in light of new data. *Q. Int.* 436, 283–293. doi: 10.1016/j.quaint.2015.10.054

Gonçalves, R. M., Lima, N. F., and Ferreira, M. U. (2014). Parasite virulence, co-infections and cytokine balance in malaria. *Pathog. Global Health* 108, 173–178. doi: 10.1179/2047773214Y.0000000139

Goossens, B., Osaer, S., Ndao, M., Van Winghem, J., and Geerts, S. (1999). The susceptibility of Djallonké and Djallonké-Sahelian crossbred sheep to Trypanosoma congolense and helminth infection under different diet levels. *Vet. Parasitol.* 85, 25–41. doi: 10.1016/s0304-4017(99)00087-4

Greenlee, J. J. (2019). Review: Update on Classical and Atypical Scrapie in Sheep and Goats. *Vet. Pathol.* 56, 6–16. doi: 10.1177/0300985818794247

Griffin, L., and Allonby, E. W. (1979b). Trypanotolerance in breeds of sheep and goats with an experimental infection of Trypanosoma congolense. *Vet. Parasitol.* 5, 97–105. doi: 10.1016/0304-4017(79)90001-3

Griffin, L., and Allonby, E. W. (1979a). Studies on the epidemiology of trypanosomiasis in sheep and goats in Kenya. *Trop. Anim. Health Prod.* 11, 133–142. doi: 10.1007/bf02237789

Grillo, K. M., Prendergast, M. E., Contreras, D. A., Fitton, T., Gidna, A. O., Goldstein, S. T., et al. (2018). Pastoral Neolithic Settlement at Luxmanda. *Tanzania. J. Field Archaeol.* 43, 102–120. doi: 10.1080/00934690.2018.1431476

Gutierrez, C., Corbera, J. A., Morales, M., and Büscher, P. (2006). Trypanosomosis in Goats. *Anna. N. Y. Acad. Sci.* 1081, 300–310. doi: 10.1196/annals.1373.040

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7443–7448. doi: 10.1073/pnas.1232392100

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hoeven, E., Fidalis, M. N., Leak, S. G. A., Geerts, S., Hanotte, O., and Jianlin, H. (2006). Introgression of the sahelian breed into West African dwarf goats. *Int. Scient. Council* 2005, 622–626.

Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. *Genetics* 198, 497–508. doi: 10.1534/genetics.114.167908

Howard, J. T., Kachman, S. D., Snelling, W. M., Pollak, E. J., Ciobanu, D. C., Kuehn, L. A., et al. (2013). Beef cattle body temperature during climatic stress: a genome-wide association study. *Int. J. Biometeorol.* 58, 1665–1672. doi: 10.1007/s00484-013-0773-5

Hu, H., and Sun, S.-C. (2016). Ubiquitin signaling in immune responses. *Cell Res.* 26, 457–483. doi: 10.1038/cr.2016.40

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68–73. doi: 10.1093/bioinformatics/14.1.68

Informal Expert Group on Gambiense HAT Reservoirs, Büscher, P., Bart, J.-M., Boelaert, M., Bucheton, B., Cecchi, G., et al. (2018). Do Cryptic Reservoirs Threaten Gambiense-Sleeping Sickness Elimination? *Trends. Parasitol.* 34, 197–207. doi: 10.1016/j.pt.2017.11.008

Ishimoto, K., Hayase, A., Kumagai, F., Kawai, M., Okuno, H., Hino, N., et al. (2017). Degradation of human Lipin-1 by BTRC E3 ubiquitin ligase. *Biochem. Biophys. Res. Commun.* 488, 159–164. doi: 10.1016/j.bbrc.2017.04.159

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. (2008). *Hole-Filled Seamless Srtm Datav4*. Colombia: International Centre for Tropical Agriculture.

Kanyari, P. W. N., Munyua, W. K., and Wilson, A. (1986). Goat trypanosomiasis: trypanotolerance and epidemiology among goat breeds in Kenya. *Bull. Anim. Health Prod. Afr.* 34, 93–97.

Katunguka-Rwakishaya, E., Murray, M., and Holmes, P. H. (1997). Susceptibility of three breeds of ugandan goats to experimental infection withTrypanosoma congolense. *Trop Anim. Health Prod.* 29:7. doi: 10.1007/BF02632337

Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., et al. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255–264. doi: 10.1038/hdy.2015.94

Kim, E.-S., Sonstegard, T. S., Silva, M. V. G. B., da, Gasbarre, L. C., and Tassell, C. P. V. (2015). Genome-Wide Scan of Gastrointestinal Nematode Resistance in Closed Angus Population Selected for Minimized Influence of MHC. *PLoS One* 10:e0119380. doi: 10.1371/journal.pone.0119380

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The genome landscape of indigenous African cattle. *Genome Biol.* 18:34. doi: 10.1186/s13059-017-1153-y

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387

Kristjanson, P., Swallow, B., Rowlands, G., Kruska, R., and De Leeuw, P. (1999). Measuring the costs of African animal trypanosomosis, the potentialbenefits of control and returns to research. *Agric. Syst.* 59, 79–98. doi: 10.1016/s0308-521x(98)00086-9

Kusiluka, L., and Kambarage, D. (1996). *Diseases of small ruminants in sub-saharan africa. A handbook*. Scotland: VETAID, 110.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *J. Statist. Softw.* 25, 1–18. doi: 10.18637/jss.v025.i01

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565

Lemos, M. V. A., Chiaia, H. L. J., Berton, M. P., Feitosa, F. L. B., Aboujaoud, C., Camargo, G. M. F., et al. (2016). Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. *BMC Genomics* 17:213. doi: 10.1186/s12864-016-2511-y

LUKNOW (1950). Les races ovines et caprines de l'Afrique Occidentale Française-(Conférence de LUKNOW-13 et 22 février 1950). *Rev. Elev. Méd. Vét. Pays Trop.* 4, 193–201. doi: 10.19182/remvt.6875

Luu, K., Bazin, E., and Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mole. Ecol. Resour.* 17, 67–77. doi: 10.1111/1755-0998.12592

Lwanga-Lunyiigo, S. (1987). *The Colonial Roots Of Internal Conflict In Uganda*, Kampala. Makerere Institute of Social Research.

Ma, T., Qu, D., Yan, B., Zhang, Q., Ren, J., and Hu, Y. (2018). Effect of SLC34A2 gene mutation on extracellular phosphorus transport in PAM alveolar epithelial cells. *Exp. Therap. Med.* 15, 310–314. doi: 10.3892/etm.2017.5380

Machado, F. S., and Aliberti, J. (2006). Impact of lipoxin-mediated regulation on immune response to infectious disease. *Immunol. Res.* 35, 209–218. doi: 10.1385/IR:35:3:209

Machado, F. S., Dutra, W. O., Esper, L., Gollob, K. J., Teixeira, M. M., Factor, S. M., et al. (2012). Current understanding of immunity to Trypanosoma cruzi infection and pathogenesis of Chagas disease. *Semin. Immunopathol.* 34, 753–770. doi: 10.1007/s00281-012-0351-7

Marshall, F., Grillo, K. M., and Arco, L. (2011). in *Sustainable Lifeways: Cultural Persistence in an Ever-Changing Environment*, eds N. F. Miller, K. M. Moore, and K. Ryan (Philadelphia: University of Pennsylvania Museum of Archaeology and Anthropology), 38–73.

Martin, P., Szymanowska, M., Zwierzchowski, L., and Leroux, C. (2002). The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reproduct .Nutr. Devel.* 42, 433–459. doi: 10.1051/rnd:2002036

Masefield, G. B. (1962). *Agricultural Change in Uganda: 1945-1960*, Vol. 3. Rome: Food Research Institute, 1–38.

Mason, I. L. (1969). *A World Dictionary of Livestock Breeds, Types, and Varieties*. England: Commonwealth Agricultural Bureaux.

Mawuena, K. (1986). Trypanosomose des moutons et des chèvres de race Naine Djallonké des régions Sud-guinéennes au Togo. *Rev. Elev. Méd. Vét. Pays Trop.* 39, 307–315. doi: 10.19182/remvt.8557

Mdladla, K. (2016). *Landscape genomic approach to investigate genetic adaptation in South African indigenous goat populations*. Ph.D thesis. South Africa: University of KwaZulu-NatalPietermaritzburg.

Meyer, C. (ed.) (2020). *Dictionnaire des Sciences Animales. [On line]*. France: Cirad.

Minematsu, H., Shin, M. J., Celil Aydemir, A. B., Kim, K.-O., Nizami, S. A., Chung, G.-J., et al. (2011). Nuclear presence of nuclear factor of activated T cells (NFAT) c3 and c4 is required for Toll-like receptor-activated innate inflammatory response of monocytes/macrophages. *Cell. Signal.* 23, 1785–1793. doi: 10.1016/j.cellsig.2011.06.013

MINEPIA (2000). *CAMEROUN. Ministère de l'Elevage, des pêches et des industries animales 2000. Rapport annuel*. Yaoundé: Direction des productions animales.

Monirei, J. M., Murray, M., and Whitelaw, D. D. (1982). *Susceptibility of East African livestock to African trypanosomiasis [Report]. Deutsche Gesellschaft fur Technische Zusammenarbeit*. Availale online at: https://cgspace.cgiar.org/handle/10568/51111.

Mruttu, H., Ndomba, C., Nandonde, S., and Nigussie Brook, K. (2016). *Animal Genetics Strategy And Vision For Tanzania*. Nairobi: Tanzania Ministry of Agriculture.

Murray, M., Morrison, W. I., and Whitelaw, D. D. (1982). Host susceptibility to African trypanosomiasis: trypanotolerance. *Adv. Parasitol.* 21, 1–68. doi: 10.1016/s0065-308x(08)60274-2

Murray, M., Trail, J. C. M., Davis, C. E., and Black, S. J. (1984). Genetic Resistance to African Trypanosomiasis. *J. Infect. Dis.* 149, 311–319. doi: 10.1093/infdis/149.3.311

Mutayoba, B. M., Gombe, S., Waindi, E. N., and Kaaya, G. P. (1989). Comparative trypanotolerance of the small east African breed of goats from different localities to Trypanosoma congolense infection. *Vet. Parasitol. May* 31, 95–105. doi: 10.1016/0304-4017(89)90024-1

Mwacharo, J., Elbeltagy, A. R., Kim, E.-S., Aboul-Naga, A. M., Rischkowsky, B., and Rothschild, M. F. (2018). Intra-population SNP analysis identifies signatures for grazing stress tolerance in Egyptian local goats. *Proc.the World Congr. Genet. Appl. Livestock Product.* 459, 1–6.

Ndayishimiye, J.-P. (1986). Les conditions d'existence des paysans montagnards du Bututsi au Burundi. *Les Cahiers d'Outre-Mer.* 39, 351–370. doi: 10.3406/caoum.1986.3198-8

Newman, J. L. (1995). *The peopling of Africa: a geographicinterpretation*. London: Yale University.

Nguluma, A. S., Msalya, G., and Chenyambuga, S. W. (2016). Phenotypic variation among four populations of small East African goats of Tanzania. *Liv. Res. Rural Devel.* 28:8.

Nsubuga, H. S. K. (1994). "Small ruminants: Goats and sheep in Uganda," in *Small Ruminant Research and Development un Africa Proceedings of the Third Biennial Conference of the African Small Ruminant Research Network UICC* (London: Yale University).

Onzima, R. (2014). "Economic Analysis of Cross Breeding Programs for Indigenous Goat Breeds in Uganda," in *Conference: 10th World Congress on Genetics Applied to Livestock Production*. (London: Yale University).

Onzima, R. B., Upadhyay, M. R., Mukiibi, R., Kanis, E., Groenen, M. A. M., and Crooijmans, R. P. M. A. (2018). Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds. *Anim. Genet.* 49, 59–70. doi: 10.1111/age.12631

Osaer, S., Goossens, B., Clifford, D. J., Kora, S., and Kassama, M. (1994). A comparison of the susceptibility of Djallonké sheep and West African Dwarf goats to experimental infection with two different strains of Trypanosoma congolense. *Vet. Parasitol.* 51, 191–204. doi: 10.1016/0304-4017(94)90156-2

Ouchene-Khelifi, N.-A., Lafri, M., Pompanon, F., Ouhrouch, A., Ouchene, N., Blanquet, V., et al. (2018). Genetic homogeneity of North-African goats. *PLoS One* 13:e0202196. doi: 10.1371/journal.pone.0202196

Pereira, F., and Amorim, A. (2010). *Origin and Spread of Goat Pastoralism. In eLS*. American: American Cancer Society, doi: 10.1002/9780470015902.a0022864

Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Goats.In:Mason's World Encyclopedia of Livestock Breeds and Breeding*. Oxfordshire, UK: CABI, 350–352.

Puig-Oliveras, A., Ballester, M., Corominas, J., Revilla, M., Estellé, J., Fernández, A. I., et al. (2014). A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. *PLoS One* 9:e114862. doi: 10.1371/journal.pone.0114862

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qi, X., Zhang, Q., He, Y., Yang, L., Zhang, X., Shi, P., et al. (2019). The Transcriptomic Landscape of Yaks Reveals Molecular Pathways for High Altitude Adaptation. *Genom. Biol. Evolut.* 11, 72–85. doi: 10.1093/gbe/evy264

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ready, S. (2013). Abnormal Growth of the Parasite Toxoplasma Gondii in Human Cells. *Stud. Res. Creat. Activ. Fair. Reproduc Nutr. Devel.* 42, 433–459.

Rewe, T. O., Ogore, P. B., and Kahi, A. K. (2002). *Integrated goat projects in kenya: impact on genetic improvement*. Oxfordshire, UK: CABI.

Ruvuna, F. (1986). *SR-CRSP research highlights 1980 –1985*. Oxfordshire, UK: CABI.

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Serranito, B., Cavalazzi, M., Vidal, P., Taurisson-Mouret, D., Ciani, E., Bal, M., et al. (2021). Local adaptations of Mediterranean sheep and goats through an integrative approach. *BioRxiv*[Preprint] doi: 10.1101/2021.01.22.427461

Shaw, A., Wint, W., Cecchi, G., Torr, S., Waiswa, C., Temesgen, A., et al. (2017). *Intervening Against Bovine Trypanosomosis in Eastern Africa: Mapping the Costs and Benefits*. Rome: Food & Agriculture Organization of the United Nations.

Shirima, E. (2005). *Benefits from dual purpose goats for crop and livestock production under small-scale peasant systems in Kondoa eroded areas*. Rome: Food & Agriculture Organization of the United Nations.

Silatsa, B. A., Simo, G., Githaka, N., Mwaura, S., Kamga, R. M., Oumarou, F., et al. (2019). A comprehensive survey of the prevalence and spatial distribution of ticks infesting cattle in different agro-ecological zones of Cameroon. *Parasit. Vect.* 12:489. doi: 10.1186/s13071-019-3738-7

Silva, F. F., Bambou, J. C., Oliveira, J. A., Barbier, C., Fleury, J., Machado, T., et al. (2018). Genome wide association study reveals new candidate genes for resistance to nematodes in Creole goat. *Small Rum. Res.* 166, 109–114. doi: 10.1016/j.smallrumres.2018.06.004

Slingenbergh, J. (1992). *Tsetse control and agricultural development in Ethiopia. Revue Mondiale de Zootechnie (FAO)*. Rome: Revista Mundial de Zootecnia (FAO).

Smith, A. B. (1992). Origins and Spread of Pastoralism in Africa. *Annu. Rev. Anthropol.* 21, 125–141. doi: 10.1146/annurev.an.21.100192.001013

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A* 100, 9440–9445. doi: 10.1073/pnas.153050910

Swallow, B. M. (2000). *). Impacts of Trypanosomiasis on African agriculture. PAAT Technical and scientific series 2*. Rome: FAO Food and Agricultural Organization of the United Nations, 46.

Trail, J. C. M., Hoste, C. H., Wissocq, Y. J., Lhoste, P., and Mason, I. L. (1980). *Trypanotolerant livestock in West and Central Africa*, Vol. 2. Rome: FAO Food and Agricultural Organization of the United Nations.

Traoré, A., Álvarez, I., Tamboura, H. H., Fernández, I., Kaboré, A., Royo, L. J., et al. (2009). Genetic characterisation of Burkina Faso goats using microsatellite polymorphism. *Liv. Sci.* 123, 322–328. doi: 10.1016/j.livsci.2008.11.005

Trujillo, A. E., and Bergey, C. M. (2020). *Immunogenetic response to a malaria-like parasite in a wild primate* [Preprint]. *Evolut. Biol.* [Preprint] doi: 10.1101/2020.11.24.396317

Twomey, A. J., Berry, D. P., Evans, R. D., Doherty, M. L., Graham, D. A., and Purfield, D. C. (2019). Genome-wide association study of endo-parasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. *Genet. Select. Evolut.* 51:15. doi: 10.1186/s12711-019-0457-7

Verma, P., Sharma, A., Sodhi, M., Thakur, K., Kataria, R. S., Niranjan, S. K., et al. (2018). Transcriptome Analysis of Circulating PBMCs to Understand Mechanism of High Altitude Adaptation in Native Cattle of Ladakh Region. *Scient. Rep.* 8:7 doi: 10.1038/s41598-018-25736-7

Walker, A. R., Bouattour, A., Camicas, J.-L., Pena, A. E., Latif, A., Pegram, R., et al. (2003). *Ticks of Domestic Animals in Africa: A Guide to Identification of Species*. Edinburgh: Bioscience Reports.

Wang, M., Zhou, Z., Khan, M. J., Gao, J., and Loor, J. J. (2015). Clock circadian regulator (CLOCK) gene network expression patterns in bovine adipose, liver,and mammary gland at 3 time points during the transition from pregnancy into lactation. *J. Dairy Sci.* 98, 4601–4612. doi: 10.3168/jds.2015-9430

Wang, X., Liu, J., Zhou, G., Guo, J., Yan, H., Niu, Y., et al. (2016). Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Scient. Rep.* 6:38932. doi: 10.1038/srep38932

Whitelaw, D. D., Kaaya, G. P., Moulton, J. E., Moloo, S. K., and Murray, M. (1985). Susceptibility of different breeds of goats in Kenya to experimental infection withTrypanosoma congolense. *Trop. Anim. Health. Product.* 17, 155–165. doi: 10.1007/BF02355876

Wilson, R. (2018). Crossbreeding of Cattle in Africa. *J. Agricult. Environ. Sci.* 6:640. doi: 10.15640/jaes.v7n1a3

Wilson, R. T., Araya, A., and Dominique, S. (1990). *Répertoire des projets de recherche et de développement caprins et ovins en Afrique. Draft*. Edinburgh: CIPEA.

Wint, W., and Rodgers, D. (2000). *Consultants report: predicted distributions of tsetse in Africa*. Rome: Food and Agriculture Organization of the United Nations, 1–62.

Wu, J., Xia, L., Yao, X., Yu, X., Tumas, K. C., Sun, W., et al. (2020). The E3 ubiquitin ligase MARCH1 regulates antimalaria immunity through interferon signaling and T cell activation. *Proc. Natl. Acad. Sci.* 117, 16567–16578. doi: 10.1073/pnas.2004332117

Wu, X., Fang, M., Liu, L., Wang, S., Liu, J., Ding, X., et al. (2019). Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. *BMC Genomics* 14:897. doi: 10.1186/1471-2164-14-897

Yang, J., Li, W.-R., Lv, F.-H., He, S.-G., Tian, S.-L., Peng, W.-F., et al. (2016). Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol. Biol. Evolut.* 33, 2576–2592. doi: 10.1093/molbev/msw129

Zheng, N., and Shabek, N. (2017). Ubiquitin Ligases: Structure, Function, and Regulation. *Annu. Rev. Biochem.* 86, 129–157. doi: 10.1146/annurev-biochem-060815-014922

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 August 2021
doi: 10.3389/fgene.2021.723980

Genome-Wide Association Study of Growth Performance and Immune Response to Newcastle Disease Virus of Indigenous Chicken in Rwanda

Richard Habimana[1,2], Kiplangat Ngeno[2], Tobias Otieno Okeno[2], Claire D' andre Hirwa[3], Christian Keambou Tiambo[4] and Nasser Kouadio Yao[5]*

[1]College of Agriculture, Animal Science and Veterinary Medicine, University of Rwanda, Kigali, Rwanda, [2]Animal Breeding and Genomics Group, Department of Animal Science, Egerton University, Egerton, Kenya, [3]Rwanda Agricultural and Animal Resources Development Board, Kigali, Rwanda, [4]Centre for Tropical Livestock Genetics and Health, International Livestock Research Institute, Nairobi, Kenya, [5]Biosciences Eastern and Central Africa – International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

OPEN ACCESS

Edited by:
Younes Miar, Dalhousie University, Canada

Reviewed by:
Sayed Haidar Abbas, Northwest A & F University, China
Suxu Tan, Michigan State University, United States

***Correspondence:**
Richard Habimana
hrichard86@yahoo.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 June 2021*
Accepted: *15 July 2021*
Published: *16 August 2021*

Citation:
Habimana R, Ngeno K, Okeno TO, Hirwa CD, Keambou Tiambo C and Yao NK (2021) Genome-Wide Association Study of Growth Performance and Immune Response to Newcastle Disease Virus of Indigenous Chicken in Rwanda. Front. Genet. 12:723980. doi: 10.3389/fgene.2021.723980

A chicken genome has several regions with quantitative trait loci (QTLs). However, replication and confirmation of QTL effects are required particularly in African chicken populations. This study identified single nucleotide polymorphisms (SNPs) and putative genes responsible for body weight (BW) and antibody response (AbR) to Newcastle disease (ND) in Rwanda indigenous chicken (IC) using genome-wide association studies (GWAS). Multiple testing was corrected using chromosomal false detection rates of 5 and 10% for significant and suggestive thresholds, respectively. BioMart data mining and variant effect predictor tools were used to annotate SNPs and candidate genes, respectively. A total of four significant SNPs (rs74098018, rs13792572, rs314702374, and rs14123335) significantly ($p \leq 7.6E{-}5$) associated with BW were identified on chromosomes (CHRs) 8, 11, and 19. In the vicinity of these SNPs, four genes such as pre-B-cell leukaemia homeobox 1 (PBX1), GPATCH1, MPHOSPH6, and MRM1 were identified. Four other significant SNPs (rs314787954, rs13623466, rs13910430, and rs737507850) all located on chromosome 1 were strongly ($p \leq 7.6E{-}5$) associated with chicken antibody response to ND. The closest genes to these four SNPs were cell division cycle 16 (CDC16), zinc finger, BED-type containing 1 (ZBED1), myxovirus (influenza virus) resistance 1 (MX1), and growth factor receptor bound protein 2 (GRB2) related adaptor protein 2 (GRAP2). Besides, other SNPs and genes suggestively ($p \leq 1.5E{-}5$) associated with BW and antibody response to ND were reported. This work offers a useful entry point for the discovery of causative genes accountable for essential QTLs regulating BW and antibody response to ND traits. Results provide auspicious genes and SNP-based markers that can be used in the improvement of growth performance and ND resistance in IC populations based on gene-based and/or marker-assisted breeding selection.

Keywords: genome-wide association studies, quantitative trait loci, putative gene, single nucleotide polymorphism, growth performance, indigenous chicken

INTRODUCTION

The genome-wide identification of genes linked to complex traits started in the 1990s (Zhang et al., 2012). The most suitable way to detect genetic variation for economically significant traits at the genome-wide level was to map quantitative trait loci (QTL; Soller et al., 2006). Earlier genomic studies have mostly used low-density marker assays like microsatellites (Goddard and Hayes, 2009). However, this approach does not provide novel information any longer (Zhang et al., 2012). Genome-wide association studies (GWAS) may now be undertaken on a larger scale thanks to the development of high-throughput genotyping tools and relevant statistical techniques (Sharmaa et al., 2015). Currently, GWAS is the most commonly used approach for searching a single variant and identifying functional complex traits genes (Berghof et al., 2018; Ji et al., 2019). Compared with previous QTL mapping approaches, GWAS is the most important in the detection of causal variants having simple effects. It is also powerful in delineating narrow genomic regions that have causal variants (MacKay et al., 2009; Sharmaa et al., 2015). GWAS does not assume that certain QTLs or genes are related to specific traits (Sharmaa et al., 2015), rather it gives the relationship between given traits and genetic markers (Berghof et al., 2018; Ji et al., 2019).

The use of GWAS in chickens has been to ascertain major loci related to economic traits (Liu et al., 2019; Saelao et al., 2019; Zhang et al., 2020). Most economic traits in animals display quantitative variation, which is regulated by many QTLs with comparatively small effects and altered by the environment (Tsudzuki et al., 2007). Growth performance traits are among the furthermost important economic traits in a poultry venture (Jin et al., 2015). Significant improvements in the study of growth performance traits in chicken have been realized, and many related genes and QTLs have been reported (Zhang et al., 2020). Over 1,500 QTLs, covering the majority of the chicken genome has been linked with growth performance traits (Hu et al., 2010; Sharmaa et al., 2015; Noorai et al., 2019). The majority of these reported QTLs are from crossbred chicken populations (Mebratie et al., 2019) and were discovered using microsatellites, markers with low map resolution, resulting in the detection of a few causative genes (Sharmaa et al., 2015).

Chicken has been threatened by various diseases (Jie and Liu, 2011) including Newcastle disease (ND; Kapczynski et al., 2013), a highly contagious viral disease of birds (El Sayed et al., 2016). The development of molecular and quantitative genetics assist in breeding for resistance in poultry (Walugembe et al., 2019). Currently, numerous efforts have been made globally to genetically improve disease resistance (Saelao et al., 2019). The immune capacity for specific diseases is a beneficial indicator of a good immunological response. This trait can be assessed and measured in live animals for indirect selection when breeding for resistance (Goddard and Hayes, 2009). Immunological characteristics such as antibody titres are heritable in poultry (Lamont et al., 2009; Lwelamira et al., 2009; Biscarini et al., 2010). This implies that there is a possibility of discovering loci and/or genes related to the immune and/or disease resistance traits. Selection for antibody response (AbR) is a suitable process for effectively improving resistance to diseases such as ND (Lillie et al., 2017). The development and distribution of disease-resistant chicken flocks represent a proactive approach for diseases control as compared to the current methods, which use drugs and vaccination (Dar et al., 2018). More than 10 QTLs associated with the antibody response to ND on chicken have been reported (Saelao et al., 2019; Walugembe et al., 2019).

Despite the existing QTL reports for BW and antibody response to ND in chicken, replication and confirmation of these QTL effects in IC in Rwanda have not been done. Growth performance and antibody response to ND of IC in Rwanda have been evaluated (Habimana et al., 2020b). This study examined the genetic architecture of BW and immune response to ND traits of IC in Rwanda. In this study, GWAS was performed to detect single nucleotide polymorphisms (SNPs) and candidate genes significantly related to growth and immune response to ND traits.

METHODOLOGY

Ethical Statement

All chicken manipulations were carried out as per the revised Animals Act 1986 with the approval of the ethical clearance committee of the College of Agriculture, Animal Sciences and Veterinary Medicine, University of Rwanda (Ref: 031/19/DRI September 2, 2019). The birds were humanely handled, and during the research, none of them was sacrificed.

Experimental Birds

Indigenous chicken (IC) used in this study were kept under the same conditions on- station at the University of Rwanda in the Eastern region, Nyagatare district as explained in Habimana et al. (2020b).

Phenotyping

The live body weights (BWs) of 185 ICs were weighed at the 20th week, and immune response to ND [mainly immunoglobulin Y (IgY)] titres recorded after 7 days from the second immunization (**Table 1**) as outlined in Habimana et al. (2020b).

Genotyping

Blood samples from 185 ICs were collected using an EDTA tube. Promega genomic DNA extraction kit was used to extract genomic DNA from blood. The concentration and quality of Genomic DNA were assessed using NanoDrop™ 2000 spectrophotometer (Thermo Scientific™ Nanodrop 2000)

TABLE 1 | Means and SDs for 20-week body weight (BW) and antibody titres to Newcastle disease (ND) traits.

Traits	Samples	Mean	SD	Minimum	Maximum
Body weight (g)	185	1,435.15	549.62	706.35	3,000.50
Antibody (titre)	185	6,064.12	2,178.67	1,615	12,000

and gel electrophoresis (1% agarose) ensuring they met genotyping requirements. Genotyping-by-sequencing (GBS) approach (at the BecA-ILRI –Integrated Genotyping Service and support, Nairobi – Kenya) was used to obtain the raw reads (Elshire et al., 2011).

Reads Alignment and SNP Calling

Raw reads were trimmed through sickle (Joshi and Fass, 2011) and aligned to Galgal6 (chicken reference genome) by the Burrows-Wheeler Alignment tool (BWA v0.7.17; Li et al., 2008). Picard package was used to remove duplicated reads. SNPs calling was done using SAMtools v1.3.1 (Li et al., 2009).

Resultant SNPs were subjected to the following filtering criteria in Plink v1.07 software (Purcell et al., 2007); minimum SNP quality of 20, 5% missing SNP genotypes, Hardy–Weinberg equilibrium ($p < 10^{-6}$), call rate >95%, heterozygosity >0.4, and minor allele frequency >0.05. Genotype imputation was performed (Marchini and Howie, 2010) to increase the power of genome-wide analysis. Missing genotypes were imputed using the LD KNNi imputation method in Tassel 5.2.60 (Bradbury et al., 2007). For each chromosome, Tassel 5.2.60 estimated pairwise linkage disequilibrium (LD; Bradbury et al., 2007). Autosomal SNPs were pruned by using indep-pairwise parameters described by Wang et al. (2009), resulting in 65,945 independent SNP markers.

Data Analysis

Multidimensional scaling (MDS) based on the centred identical-by-state (IBS) approach was used to test the population structure. Pairwise IBS distances were computed using independent SNPs. MDS components were got using the MDS-plot option based on the IBS matrix. Due to the internal population structure, the first MDS component was included as a covariate in the statistical model for assessing SNP effects on growth and antibody response attributes to account for sample stratification. Tassel 5.2.60 was used to construct a relative kinship matrix (K) using 65,945 independent SNP markers (Bradbury et al., 2007).

The analysis for GWAS was performed using Tassel 5.2.60 (Bradbury et al., 2007). SNPs significantly associated with body weight and antibody response to ND were identified using the following mixed linear model (MLM) as implemented in Tassel (Zhang et al., 2010).

$$\mathbf{y} = \mathbf{X} + \mathbf{Zu} + \mathbf{e}$$

where y is the vector of quantitative traits (20-week bodyweight and antibody response to Newcastle disease); β, vectors containing fixed effect (sex, gene pools, and SNPs) and covariates (population structure); u, a vector of random effect of the relative kinship matrix constructed by matrix simple matching coefficients based on the independent SNPs; e, a vector of random residuals; and X and Z are design matrices.

The family-wise error rate was controlled by using a Bonferroni correction. Based on the estimated number of independent SNP markers, the threshold value of p of the 5% Bonferroni genome-wide significance was computed. Therefore, the threshold value of p of the 5% Bonferroni genome-wide significance was 7.6E−5 (0.05/65,945). Similarly, the threshold value of p for the significance of suggestive linkage allowing one false-positive effect in a genome-wide test (Lander and Kruglyak, 1995) was calculated using the same approach as above and was 1.5E−5 (1/65,945). Relationship of normal theoretical quantiles of the probability distributions between expected (x-axis) and observed (y-axis) values of p from each, the, respectively, associated trait was shown by the Quantile-quantile (QQ) plots (**Supplementary Datasheet 2**). A comprehensive view of all values of p for each trait's SNP markers was observed *via* the Manhattan plot. Manhattan and QQ plots were produced by the qqman package of the R software (Turner, 2018).

BioMart data-mining (Kinsella et al., 2011) and variant effect predictor (Dashti and Gamieldien, 2017) tools were used to annotate all significant SNPs found in GWAS, as well as genes situated 100 kb upstream and downstream of these SNPs, respectively. This was done in order to catalogue all of the genes found around the discovered SNPs, resulting in the compilation of gene lists for BW and antibody response to ND traits.

Database for Annotation, Visualization, and Integrated Discovery (DAVID; Dennis et al., 2003) was used to analyze putative genes for each trait. Gene ontology and functional annotation clustering analysis were performed to figure out what the mapped genes mean biologically. When enrichment score (ES) of the DAVID, an adjusted Fisher exact value of p is higher, it reflects more enriched clusters. When an ES is more than unity, the functional category is overrepresented.

RESULTS

Genome-wide association study using genotyping by sequencing identified SNPs and putative genes linked to growth performance and immune response to Newcastle disease virus in indigenous chicken in Rwanda.

Population Structure of Indigenous Chicken in Rwanda

Multidimensional scaling analysis of 65,945 SNPs with the first two components revealed an internal population structure as explained by the variance among individuals' population stratification (**Figure 1**).

Genome-Wide Association Studies

Possible Causal Variant for Body Weight in Indigenous Chicken in Rwanda

Location and annotation of all significant and suggestive SNPs identified by GWAS are displayed in **Figure 2; Tables 2** and **3**. A total of four SNPs (rs74098018, rs13792572, rs314702374, and rs14123335) significantly ($p \leq$ 7.6E−5) associated with 20-week BW were detected on chromosomes 8, 11, and 19. Furthermore, rs318161016, a suggestive SNP ($p \leq$ 1.5E−5) associated also with BW was detected on chromosome 4 (**Figure 2**).

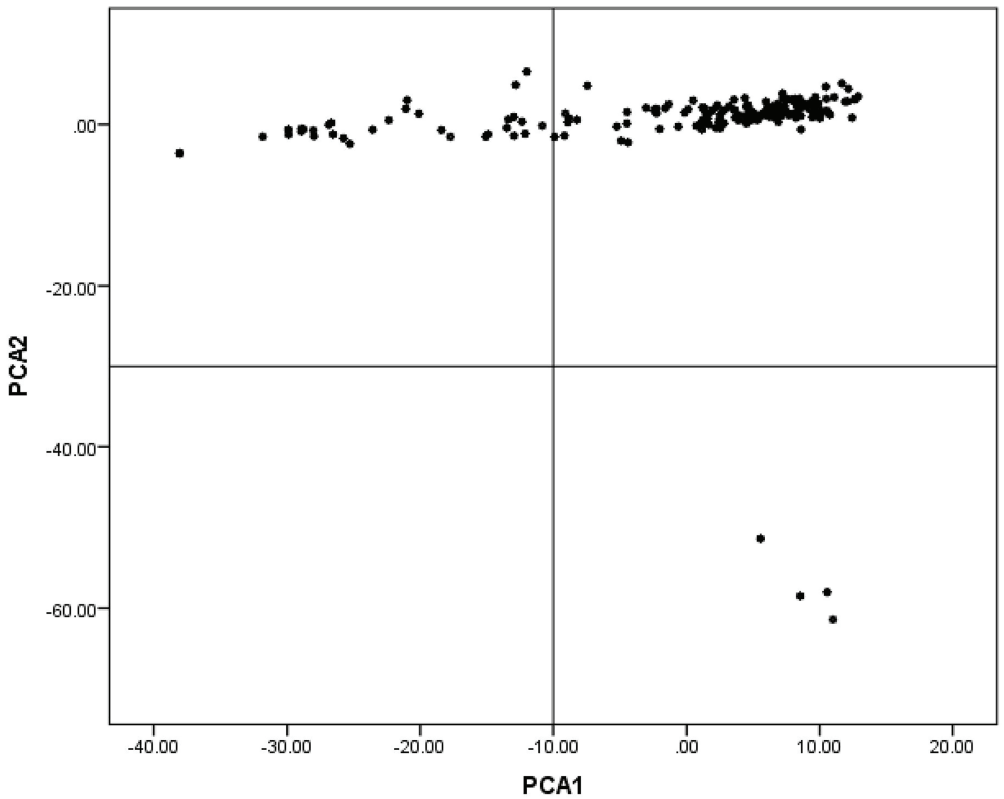

FIGURE 1 | Population structure revealed by multidimensional scaling (MDS) analysis in indigenous chicken (IC) in Rwanda.

Lists of candidate genes situated 100 kb downstream and upstream of the significant SNPs for BW at week 20 are summarized in **Table 2** and **Supplementary Table S1**. Among all the mapped genes in the candidate genomic regions, only four known genes namely *pre-B-cell leukaemia homeobox 1* (*PBX1*), *G-patch domain containing 1* (*GPATCH1*), *M-phase phosphoprotein 6* (*MPHOSPH6*), and *mitochondrial rRNA methyltransferase 1* (*MRM1*) were the closest to the significant SNPs (**Table 2**). One gene called *Fibroblast growth factor 2* (*FGF2*) was reported in the vicinity of a suggestive SNP (**Table 2**).

Functional annotation clustering analysis of all mapped genes for BW exposed the presence of enriched gene groups related to transcription factor activity, sequence-specific DNA binding, arylamine *N*-acetyltransferase activity, and growth factor activity (**Supplementary Table S2**). The biological functions of all genes neighboring the significant and suggestive SNPs for BW were identified and are presented in the **Supplementary Table S3**.

Possible Causal Variant for Antibody Response to Newcastle Disease in Indigenous Chicken in Rwanda

Overall view of values of p for all the SNPs contributing to the antibody response to ND showed that a genomic region on chromosome 1 with four SNPs (rs314787954, rs13623466, rs13910430, and rs737507850) was strongly ($p \leq 7.6E{-}5$) associated with the IC antibody response to ND after 7 days from the second immunization. Additionally, 10 suggestive SNPs ($p \leq 1.5E{-}5$) detected on chromosomes 2, 4, 8, 13, 17, 19, and 26 influenced also antibody response to ND in IC in Rwanda (**Figure 3**).

All genes located at 100 kb upstream and downstream of the four significant and suggestive SNPs were identified for antibody response to ND (**Supplementary Table S1**). Among the known genes, four including *cell division cycle 16* (*CDC16*), *zinc finger, BED-type containing 1* (*ZBED1*), *myxovirus (influenza virus) resistance 1* (*MX1*), and *growth factor receptor bound protein 2 (GRB2) related adaptor protein 2* (*GRAP2*) were the closest to the significant SNPs (**Table 2**). In addition, a total of 10 known genes namely *ubiquitin-associated domain containing 1* (*UBAC1*), *tubulin epsilon and delta complex 1* (*TEDC1*), *Ras association domain family member 5* (*RASSF5*), *intraflagellar transport 22* (*IFT22*), *jumonji and AT-rich interaction domain containing 2* (*JARDI2*), *gamma-aminobutyric acid A receptor, beta 2* (*GABRB2*), *zinc finger homeobox 4* (*ZFHX4*), *adenylate cyclase-activating polypeptide 1* (*ADCYAP1*), *pterin-4 alpha-carbinolamine dehydratase 2* (*PCBD2*), and *albumin* (*ALB*) were mapped in the neighborhood of the identified suggestive SNPs (**Table 3**). Biological functions of these genes mapped nearby significant and suggestive SNPs for antibody response (Ab) response to ND are presented in **Supplementary Table S3**.

Functional annotation clustering analysis exposed the presence of enriched gene clusters associated with regulation of transcription,

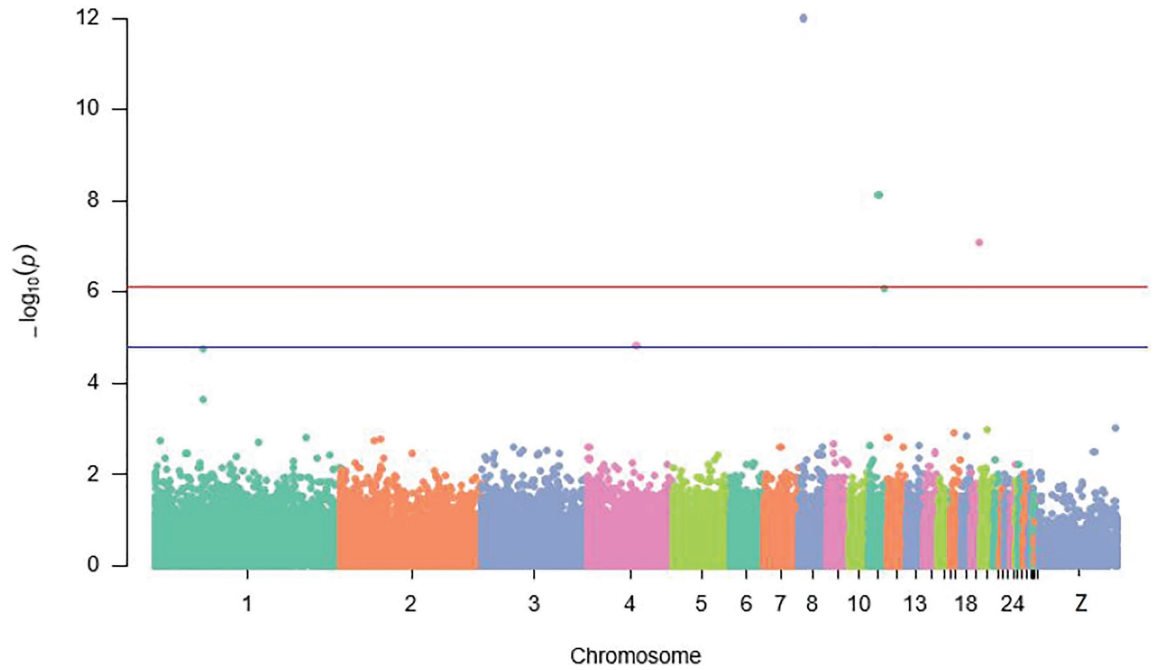

FIGURE 2 | Manhattan plot displaying significant and suggestive single nucleotide polymorphisms (SNPs) associated with IC body weight {the red line designates a Bonferroni-adjusted genome-wide threshold [–log10 (p) ≥ 6.1] and the blue line indicates a suggestive genome-wide threshold [–log10 (p) ≥ 4.8]}.

DNA-templated, metal-ions binding and transport, ATP binding, and cytokine activity (**Supplementary Table S2**).

DISCUSSION

Genome-wide association study is a prevailing tool for genetic analysis of essential traits in domestic animals. In this study, a GWAS experiment was performed to investigate the genetic basis of the BW and Ab response to ND among IC populations in Rwanda. A supposition made in GWAS is that it is possible to identify significant associations since SNPs are in LD with causal mutations for the traits of interest (Zhang et al., 2012).

Population Structure of Indigenous Chicken in Rwanda

The structured population witnessed in this study is consistent with findings in the previous report on the IC population in Rwanda (Habimana et al., 2020a). This population stratification could, however, confound GWAS. Nevertheless, the MLM method used in this study corrected for the stratification and removed confounding effects (Price et al., 2010) and markedly reduced the number of false-positive associations (Sharmaa et al., 2015).

Genome-Wide Association Studies

Possible Causal Variant for Body Weight in Indigenous Chicken in Rwanda

Bodyweight is under complex genetic control (Zhao et al., 2012). Detection of the molecular growth mechanism would contribute to a more efficient selection for growth in chicken (Deeb and Lamont, 2002). In this study, the GWAS resulted in the identification of four significant and one suggestive SNPs associated with BW. Of these four leading SNPs, one was detected on chromosome 8, two on chromosome 11, and one on chromosome 19. However, the strongest association signal was revealed on chromosome 8. This study confirmed the existing candidate genomic regions for BW on chromosomes 8 (Pértille et al., 2017) and 11 (Jennen et al., 2004; Jacobsson et al., 2005; McElroy et al., 2006). However, the region on chromosome 19, which is associated with BW can be taken as a novel candidate genomic region controlling BW in IC, since QTL or SNP was never reported in this genomic region.

In the present study, a suggestive QTL for 20-week BW was discovered on chromosome 4. This suggestive QTL corroborated previous studies, which revealed QTLs for BW on the same chromosome (Tsudzuki et al., 2007; Sharmaa et al., 2015; Pértille et al., 2017). Our study also confirmed a study by Johansson et al. (2010), which reported that a region on chromosome 4 was recounted to be subjected to intense selection in several chicken lines with different selection on BW for up to 50 generations. This phenomenon could be attributed to the fact that one QTL on chromosome 4 harbors genes with vast effects on developmental increases in BW (Ankra-Badu et al., 2010). Chromosome 4 could harbor an important genomic region for further analysis as the different selection responses in chicken lines are triggered by the genomic region covering one or more genes with genetic background-dependent effects. In previous studies, apart from chromosomes

TABLE 2 | Summary of significant SNPs and nearest genes detected at false discovery rate of 5% for body weight and antibody response to Newcastle disease in indigenous chicken in Rwanda.

Traits	SNP ID	CHR	SNP position (bp)	GWAS value of *p*	Allele	Candidate gene	Distance (bp)	Biological function
BW	rs740980181	8	5,672,840	1E−10	A	*PBX1*	Within	Positive regulation of cell proliferation and negative regulation of neuron differentiation
	rs13792572	11	10,145,725	7.6E−09	G	*GPATCH1*	7,341^{D}	mRNA processing
	rs314702374	19	8,356,215	8E−08	G	*MPHOSPH6*	22,019^{U}	Maturation of 5.8S rRNA
	rs14123335	11	15,831,341	8.4E−07	C	*MRM1*	54^{U}	Enzyme-directed rRNA 2'-O-methylation
IR	rs314787954	1	137,692,275	4E−10	C	*CDC16*	Within	Protein K11-linked ubiquitination
	rs13623466	1	129,958,897	6.42E−08	T	*ZBED1*	7,631^{U}	Regulation of transcription from RNA polymerase II promoter
	rs13910430	1	101,256,943	1E−07	T	*MX1*	29,169^{U}	Innate immune response, organelle fission, and defense response to the virus
	rs737507850	1	50,383,558	1.55E−07	C	*GRAP2*	44,310^{U}	Leukocyte-specific protein-tyrosine kinase signaling, innate and adaptive immune response

CHR, chromosome; BW, body weight; IR, immune response to Newcastle disease; SNP, Single nucleotide polymorphism; PBX1, pre-B-cell leukaemia homeobox 1; GPATCH1, G-patch domain containing 1; MPHOSPH6, M-phase phosphoprotein 6; MRM1, mitochondrial rRNA methyltransferase 1; CDC16, cell division cycle 16; ZBED1, zinc finger, BED-type containing 1; MX1, myxovirus (influenza virus) resistance 1; GRAP2, GRB2 related adaptor protein 2; superscript letters U and D indicate that the SNP is located upstream and downstream of the nearest gene, respectively.

4, 8, 11,16, 19, 20, 22, and 25, the remaining chromosomes including Z were reported to harbor QTLs affecting BW in chicken at different ages (Gu et al., 2011; Rikimaru et al., 2011; Xie et al., 2012). However, major QTLs are commonly mapped on chromosome 1 (Liu et al., 2007; Tsudzuki et al., 2007; Uemoto et al., 2009).

The differences in QTL affecting BW could be attributed to several factors such as breed types and generations (Gu et al., 2011; Xie et al., 2012) and experimental population used (McElroy et al., 2006). The magnitude of QTL affecting BW may be population-dependent and the frequency of alleles in any population depends on that population's adaptive environment. For instance, the growth dynamics of local chicken are different from layers, broilers, and crossbreed populations (Ankra-Badu et al., 2010), and grandparental breeds used for the construction of the inter-crossed lines (Tsudzuki et al., 2007). Age differences also account for the differences in QTLs (Uemoto et al., 2009; Pértille et al., 2017). This study indicates the existence of several genes involved in the growth and development of chicken at various stages of life. In addition, age-specific QTLs have been found to control BW on chromosomes 1 and 4 in earlier studies; QTL controlling BW at initial stages was identified on chromosome 1 (Liu et al., 2007; Uemoto et al., 2009). This confirms the presence of major QTLs on chromosome 1 modulating early development (Rikimaru et al., 2011). Contrarily, QTLs for BW at later ages like in this study were uncovered on chromosome 4 (Rikimaru et al., 2011; Pértille et al., 2017) with few exceptional cases reported by Wardęcka et al. (2002) and Kerje et al. (2003). Consideration should, therefore, be given to the age-specificity of the QTLs. Consequently, QTL studies on BW, each week from hatching to adult would be needed for the identification of age-specific QTLs. The discrepancy in QTLs in previous studies could be also a result of the mapping approach used. Earlier genomic studies have mostly used low-density microsatellites markers (Tsudzuki et al., 2007). Currently, GWAS employs, however, SNPs as potential markers that are dispersed throughout the whole genome at a higher density (Zhang et al., 2012). Finally, the choice of the statistical model could explain the inconsistent results (McElroy et al., 2006). Some studies used a general linear model (GLM; Wang et al., 2015; Zhang et al., 2015a), and others including this study, used a MLM (Zhang et al., 2013; Mebratie et al., 2019). The study by Mebratie et al. (2019) emphasized that using GLM and MLM in GWAS, does not essentially give the same results even without the presence of a robust population structure in the data. Kennedy et al. (1992) reported that in the presence of relations between animals, the use of GLM analysis results in an inflated *F*-test. This leads to an excess of spurious significant effects of QTLs, even if there is no effect or bias in effect estimation. The MLM approaches have, however, demonstrated worthwhile in controlling population structure and relatedness within GWAS. This analysis method offers the utmost power to discover QTLs while controlling at the preferred level for false-positive rate or false discovery rate and providing the greatest accurate estimates of QTL effect (Zhang et al., 2010; Ekine et al., 2014).

Regions on chromosomes 8, 11, and 19 seemed to be promising genomic regions for putative genes controlling indigenous chicken BW. Within these genomic regions, five candidate genes namely *PBX1*, *GPATCH1*, *MPHOSPH6*, *MRM1*, and *FGF2* were discovered. *PBX1* gene, localized in the nucleus (Berthelsen et al., 1999; Mercader et al., 1999), has been known as a homeodomain transcription factor that can form heterodimers with homeodomain proteins. These proteins are encoded by the *homeotic selector* (*Hox*) gene complexes and rise their DNA-binding affinity and specificity. McWhirter et al. (1997) suggested that interactions between *PBX1* gene and homeodomain proteins were required for Hox proteins to control downstream target genes that in turn regulate growth, differentiation, and morphogenesis during development. The expression pattern of

TABLE 3 | Summary of suggestive SNPs and nearest genes detected at false discovery rate of 10% for body weight and antibody response to Newcastle disease in indigenous chicken in Rwanda.

Traits	SNP ID	CHR	SNP position (bp)	GWAS value of *p*	Allele	Candidate genes	Distance (bp)	Biological function
BW	rs318161016	4	51,805,394	0.00001535	A	*FGF2*	8,033[U]	Growth factor-dependent regulation of skeletal muscle satellite cell proliferation
IR	rs735333650	17	8,458,226	0.000000939	T	*UBAC1*	Within	Regulation of cell survival and proteasomal degradation
	rs15900019	8	4,304,485	0.00000138	A	*TEDC1*	Within	Duplication and assembly of centrioles and basal bodies
	rs1060144701	26	2,497,963	0.00000145	A	*RASSF5*	Within	Negative regulation of cell proliferation and positive regulation of protein ubiquitination
	rs14118744	19	3,336,281	0.00000184	G	*IFT22*	Within	Small GTPase mediated signal transduction
	rs736576816	2	61,140,781	0.00000191	G	*JARID2*	Within	Liver, spleen, and thymus development, stem cell differentiation, and positive regulation of histone H3-K9 methylation
	rs741342879	13	8,272,097	6.2874E−06	T	*GABRB2*	67,439[D]	Ion transport and negative regulation of neuron apoptotic process
	rs736427856	2	119,443,614	0.00000729	T	*ZFHX4*	377[D]	Zinc ion binding and sequence-specific DNA binding
	rs1060031521	2	93,149,588	0.0000079	T	*ADCYAP1*	89,139[U]	Positive regulation of the cAMP biosynthetic process, drinking behavior, and positive regulation of hormone secretion
	rs739117494	13	16,352,162	0.00000832	T	*PCBD2*	Within	Tetrahydrobiopterin biosynthetic process, positive regulation of transcription, DNA-templated, protein homotetramerisation, and protein heterooligomerisation
	rs740392770	4	51,805,394	0.00001201	T	*ALB*	7,988[U]	Negative regulation of apoptotic process and cellular response to starvation, vitamin A, and virus

CHR, chromosome; BW, body weight; IR, immune response; SNP, single nucleotide polymorphism; FGF2, Fibroblast growth factor 2; UBAC1, ubiquitin-associated domain containing 1; TEDC1, tubulin epsilon and delta complex 1; RASSF5, Ras association domain family member 5; IFT22, intraflagellar transport 22; JARDI2, jumonji and AT-rich interaction domain containing 2; GABRB2, gamma-aminobutyric acid (GABA) A receptor, beta 2; ZFHX4, zinc finger homeobox 4; ADCYAP1, adenylate cyclase-activating polypeptide 1; PCBD2, pterin-4 alpha-carbinolamine dehydratase 2; ALB, albumin; superscript letters U and D indicate that the SNP is located upstream and downstream of the nearest gene, respectively.

this gene suggests its importance in early nervous system development. This gene was initially identified for its role in the translocation of the chromosome which happens in pre-B acute lymphoblastic leukaemia (Kamps et al., 1990; Nourse et al., 1990). Besides, Charboneau et al. (2006) revealed the involvement of *PBX1* in pro-angiogenic Hox DNA binding and transcriptional activity in endothelial cells. A study in the rodent fibroblast revealed that *PBX1* encodes a new different member of the fibroblast growth factor (FGF) family (McWhirter et al., 1997). The same study unveiled that *PBX1* homeodomain was required for the induction of FGF. This confirms the presence of *FGF2* among the identified candidate gene in this study. Therefore, we may hypothesize that *PBX1* has a significant effect on chicken growth. *FGF2* is known for its role in regulating skeletal muscle satellite cell proliferation. Also, *FGF2* is famous to be related to chicken growth (Xue et al., 2017). Another study confirms this by revealing two genes of the family of FGF namely growth factor binding protein 1 (*FGFBP1*) and 2 (*FGFBP2*) in modulating chicken growth (Ankra-Badu et al., 2010). *GPATCH1* gene is involved in mRNA processing. Besides its role in RNA transport inside the cells, the *G-patch domain* has been linked with other RNA processing functions (Yiu et al., 2004). This gene was identified and was believed to be associated with biogenesis and the developmental process in Goat (Brito et al., 2017). Moreover, it has been characterized as a novel cholesterol metabolism regulator, reducing cholesterol synthesis, and increasing the concentration of low-density lipoprotein uptake in chicken (Zhang et al., 2019). The role of the *GPATCH1* gene in modulating BW in chicken has not yet been elucidated. *MPHOSPH6* gene is involved in maturation of 5.8S rRNA and a novel target in Cytokine-mediated modulation of the hepatic miRNome (Kirchmeyer et al., 2018). Besides, *MPHOSPH6* regulates the shrimp cell cycle and development of ovary in black tiger (Zhou et al., 2013). However, its role in BW regulation in chicken has not been yet discussed. *MRM1* gene is an enzyme-directed rRNA 2'-O-methylation (Lee et al., 2013; Lee and Bogenhagen, 2014), but its function in controlling BW has not been established.

In the present study, some QTLs for 20-week BW are overlapping those already found in previous studies and others have never been reported. This study, therefore, postulates that these genomic regions may play a key role in the fundamental molecular mechanisms that are responsible for BW. These genomic regions are hence separate in many various chicken populations, confirming the difference in BW revealed in IC gene pools in Rwanda (Habimana et al., 2020a). Besides, this study offers a preliminary assumption that genes uncovered in these genomic regions could be promising candidate genes for 20-week BW in IC populations in Rwanda. Extra confirmation experimentations, however, suggest *PBX1*, *GPATCH1*, *MPHOSPH6*, and *MRM1* be novel targets for 20-week BW. For additional validation of the importance of these genes, studies including their silencing and overexpression need to be conducted both *in-vivo* and *in-vitro*. Since growth performance is the utmost crucial trait of selection, getting profound insights into the growth molecular mechanism would result in a more efficient selection for growth performance in chickens.

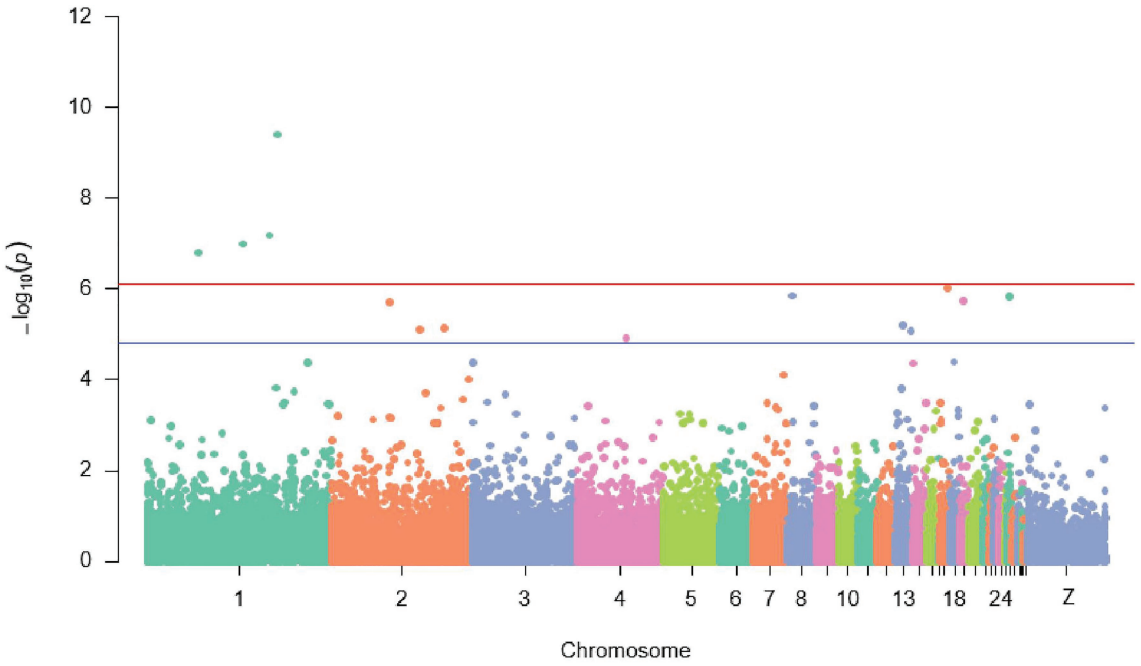

FIGURE 3 | Manhattan plot displaying significant and suggestive SNPs associated with antibody response to Newcastle disease {the red line designates a Bonferroni-adjusted genome-wide threshold [–log10 (p) ≥ 6.1] and the blue line indicates a suggestive genome}.

Possible Causal Variant for Antibody Response to Newcastle Disease in Indigenous Chicken in Rwanda

Host antibody response to viruses is a composite process (Fischer et al., 2013). Immune response to ND could be considered as a quantitative trait under polygenic control, but having few QTLs (Biscarini et al., 2010; Luo et al., 2013; Saelao et al., 2019). In this present study, all four significant SNPs strongly ($p \leq 7.6E{-}5$) associated with antibody (Ab) response to ND were detected on chromosome 1 (50,383,558–137,692,275 bp). This study confirmed the findings by Luo et al. (2013) and Saelao et al. (2019) who recounted chromosome 1 to harbor genomic region associated with Ab response to ND, hence strongly suggests the presence of an important regulatory region for ND control in this region. Also, this genomic region on chromosome 1 overlays the QTLs for the Ab response to sheep red blood cells (Dorshorst et al., 2011). Therefore, this region could have great importance for chicken immune response and, probably, disease resistance in general.

Conversely, results from this study differed from those previously reported (Yonash et al., 2001; Biscarini et al., 2010; Wang et al., 2015). Yonash et al. (2001) reported that QTLs for the Ab response to ND in broiler chicken were located on chromosomes 2 and 18. Biscarini et al. (2010) reported 13 QTLs related to Ab response to ND on chromosomes 3, 4, 5, 9, 13, 16, and 22, and Z. Wang et al. (2015) revealed six QTLs affecting Ab response to ND on chromosomes 2, 4, and Z, whereas Jacob et al. (2000) discovered on chromosome 16, major histocompatibility complex (MHC) which is associated with chicken immunity (Zhou and Lamont, 2003; Zhang et al., 2015b). These inconsistent results could have happened for numerous reasons; comprising a dose of ND vaccine applied, time post-vaccination, markers used, choice of statistical models, the genetic composition of the experimental populations, and limited power of most QTLs mapping studies (McElroy et al., 2006; Ioannidis et al., 2009; Saelao et al., 2019).

The current study used IC resource populations, while other previous studies were based on broiler and layer populations (Yonash et al., 2001; Biscarini et al., 2010). The present work used an MLM while the previous studies performed GWAS through a GLM (Wang et al., 2015). The majority of the previous studies used microsatellite markers whereas this study performed GWAS using SNPs. In addition, the nature of the Ab response targeted could result in different QTLs (Saelao et al., 2019). Previous studies focused on the primary Ab response to ND virus, while the present study scrutinized the secondary Ab response to ND virus. In the primary Ab response, the main class of Ab produced is immunoglobulin M (IgM) whereas in the secondary Ab response is IgY (Sarker et al., 1999); thus, the ranking of chicken on Ab response to the primary vaccination could be different to that after the second one leading to the detection of different QTLs (Biscarini et al., 2010). The QTLs for the Ab response to ND in this study could reflect the aptitude of the memory cell pool to respond to ND.

In addition to the significant SNPs, 10 suggestive SNPs ($p \leq 1.5E{-}5$) were detected on chromosomes 2, 4, 8, 13, 17, 19, and 26. However, apart from chromosomes 2, 4, and 13 which were reported earlier (Rowland et al., 2018; Saelao et al., 2019; Walugembe et al., 2019), there were no other earlier studies

on QTLs modulating ND antibody response on chromosomes 8, 17, 19, and 26. Contrarily, Luo et al. (2013) uncovered a candidate QTL for the ND antibody response on chromosome 12. This could be attributed to the different techniques used in the current study to identify SNPs, which might have permitted the detection of novel SNPs. Surprisingly and consistently, one QTL currently identified on chromosome 4 was also associated with BW at 20 weeks of age. This could be explained by the pleiotropic mechanisms of the QTL region on that chromosome.

Some QTLs for ND antibody response reported in the current study confirmed those found in earlier studies, and other ones have never been reported. This study, therefore, hypothesizes that the reported genomic regions could contribute to the important molecular mechanisms responsible for the total effective host AbR. These genomic regions segregate in several diverse chicken populations, confirming the difference in Ab response to ND existing in IC gene pools in Rwanda (Habimana et al., 2020a). However, further investigations preferably with independent and large populations (Spelman and Bovenhuis, 1998; Marklund et al., 1999), are needed to validate the findings from this study.

Disease resistant genes encode antibodies, microRNA, and other materials helping the host to repel the harm caused by pathogens (Dar et al., 2018). So far, lots of anti-disease genes (*MH*, *NRAMP1*, *IFN*, *MX*, *ANTI-ALV*, *ZYXIN*, *TVB*, *CD1B*, *CD1CB*, *ROBO1*, *ROBO2*, *CHMP2B*, *MHC*, *SEMA5A*, and *TGFBR2*) have been revealed with the development of several molecular technologies and assays in chickens (Zhang et al., 2015b; Deist et al., 2017; Lillie et al., 2017). Other genes related to poultry immunity (*CAMK1d*, *CCDC3 TIRAP*, *ETS1*, and *KIRREL3*) are still under a validation process (Saelao et al., 2019). However, the accurate mechanism of disease resistance is not entirely clear (Jie and Liu, 2011), and only a few causal genes have been identified due to low map resolution (Luo et al., 2013). Many more genes associated with disease resistance (Jie and Liu, 2011) could be revealed with a high map resolution.

This study revealed, however, 14 putative genes (*CDC16*, *ZBED1*, *MX1*, *GRAP2*, *UBAC1*, *TEDC1*, *RASSF5*, *IFT22*, *JARDI2*, *GABRB2*, *ADCYAP1*, *PCBD2*, *ZFHX4*, and *ALB*) associated with Ab response to ND in IC populations in Rwanda. The gene ontology annotation advises that all these genes contribute to the regulation of transcription, binding, transport, cytokine activity, and immune responses. Among these genes, four were near the significant SNPs (*CDC16*, *ZBED1*, *MX1*, and *GRAP2*). These genes might be used as putative genes to be further explored to determine associations with Ab response to ND in IC populations in Rwanda.

The *CDC16* gene identified in this stud is yet to be characterized in chicken immunity. In mammals, the *CDC16* gene is, however, necessary for the normal coupling of DNA replication to mitosis and might act downstream of *CDC28* to negatively control DNA replication (Heichman and Roberts, 1998). This gene plays a significant role in cellular functions in mammals, and defects of which are closely associated with various disease processes (Shi et al., 2018). Besides, Paglialunga et al. (2017) established a vital role for *CDC16* in maintaining *in vivo* β-cell mass. This gene may have a similar function in chickens and could be a promising gene for Ab response for ND in chickens. *ZBED1* gene also known as DNA replication-related element-binding factor uncovered in this study, was initially identified as a transcription factor in *Drosophila* (Hirose et al., 1996). It binds to box A and positively regulates genes involved in DNA replication and cell proliferation, such as the proliferating cell nuclear antigen and DNA polymerase (Matsukage et al., 2008). Later on, a study found that *ZBED1* played a crucial role in promoting proliferation and decreases the chemosensitivity of gastro cancer cells (Jiang et al., 2018). Still, no information concerning its function and key mechanisms in chicken immunity is reported. *GRAP2* gene also known as *GADS*, *GRAP-2*, *GRB2L*, *GRBLG*, *GRID*, *GRPL*, *GrbX*, *Grf40*, *Mona*, and *P38* has been identified in the current study. The function of this gene has not yet been established in chicken but it has been shown in humans that the *GRAP2* gene is involved in leukocyte-specific protein-tyrosine kinase signaling and immune response by stimulating T cells (Dufner et al., 2015; Breuning and Brown, 2017). This role makes this gene a promising gene for Ab response for ND in Chicken. This study confirmed the previous results which revealed that *MX1* gene was associated with Ab response to ND in chicken (Mpenda et al., 2020). Apart from that, the *MX1* gene is associated with resistance to avian influenza (AI) and infectious bursal disease virus (IBDV; Yin et al., 2010; Ewald et al., 2011; Jie and Liu, 2011). Thus, this is an indication that the *MX* gene is involved in chicken immunity.

Ten genes (*UBAC1*, *TEDC1*, *RASSF5*, *IFT22*, *JARDI2*, *GABRB2*, *ADCYAP1*, *PCBD2*, *ZFHX4*, and *ALB*) were found to be associated with suggestively correlated SNPs for Ab response to ND in this study. All these candidate genes for Ab response to ND have not been heretofore directly associated with immune response in poultry (Adhikari and Davie, 2018; Elbeltagy et al., 2019). Some of these genes are new genes in the chicken and have mostly been discussed in other species (Zhou et al., 2014; Fleming et al., 2016; Zhuang et al., 2020). Further examination is therefore required to prove these novel genes as putative genes for Ab response to ND in chicken. Kompetitive Allele Specific Polymerase Chain Reaction (KASP)-based markers (Semagn et al., 2014; Islam and Blair, 2018) could be developed using the SNPs significantly associated with Ab response to ND and validated before being used to screen targeted chicken breeds to establish the involvement of these genes in chicken immunity to ND. Another approach will be to develop an RNA-guided cas9 nuclease from the microbial clustered regularly interspaced short palindromic repeats (CRISPR-CAS9) protocol (Ran et al., 2013; Doudna and Charpentier, 2014) for chicken with the ultimate aim of silencing these genes to ascertain their role in the chicken immunity to ND.

CONCLUSION

Genomic regions that putatively control body weight and antibody response to Newcastle disease in an indigenous chicken were identified in the current study. This information provides

insights on the genetic control of these traits and makes available genetic markers that could be useful for selective breeding programs to improve growth performance and ND resistance in IC. Few of the genomic regions overlapped with hitherto reported QTL regions provide evidence for confirmation of these QTLs and their corresponding effects. In addition, the variants and genes uncovered for the first time in this study, merit further scrutiny to understand the fundamental molecular mechanisms before practical application.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by The Research Screening and Ethical Clearance Committee of the College of Agriculture, Animal Sciences, and Veterinary Medicine, University of Rwanda. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

RH and KN: conception of the work and data analysis. CH: contributed to the data acquisition. RH, KN, and NY: results interpretation. RH: drafting the article. KN, CK, NY, and TO: critical revision of the article and final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

The ABCF programmes are funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership, the Syngenta Foundation for Sustainable Agriculture (SFSA), the Bill and Melinda Gates Foundation (BMGF), the UK Department for International Development (DFID), and the Swedish International Development Cooperation Agency (SIDA). The study was carried out with the aid of grant number 108392-001 from UNESCO and the International Development Research Centre, Ottawa, Canada. This material is also based upon work supported by the United States Agency for International Development, as part of the Feed the Future initiative, under the CGIAR Fund, award number BFS-G-11-00002, and the predecessor fund the Food Security and Crisis Mitigation II grant, award number EEM-G-00-04-00013.

ACKNOWLEDGMENTS

The authors acknowledge the financial and technical support from BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) programmes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.723980/full#supplementary-material

REFERENCES

Adhikari, A., and Davie, J. (2018). JARID2 and the PRC2 complex regulate skeletal muscle differentiation through regulation of canonical Wnt signalling. *Epigenetics Chromatin* 11:46. doi: 10.1186/s13072-018-0217-x

Ankra-Badu, G. A., Bihan-Duval, E., Le Mignon-Grasteau, S., Pitel, F., Beaumont, C., Duclos, M. J., et al. (2010). Mapping QTL for growth and shank traits in chickens divergently selected for high or low body weight. *Anim. Genet.* 41, 400–405. doi: 10.1111/j.1365-2052.2009.02017.x

Berghof, T. V. L., Visker, M. H. P. W., Arts, J. A. J., Parmentier, H. K., van der Poel, J. J., Vereijken, A. L. J., et al. (2018). Genomic region containing toll-like receptor genes has a major impact on total IGM antibodies including KLH-binding IgM natural antibodies in chickens. *Front. Immunol.* 8:1879. doi: 10.3389/fimmu.2017.01879

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes Dev.* 13, 946–953. doi: 10.1101/gad.13.8.946

Biscarini, F., Bovenhuis, H., Van Arendonk, J. A. M., Parmentier, H. K., Jungerius, A. P., and Van Der Poel, J. J. (2010). Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet.* 41, 26–38. doi: 10.1111/j.1365-2052.2009.01960.x

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Breuning, J., and Brown, M. H. (2017). T cell costimulation by CD6 is dependent on bivalent binding of a GADS/SLP-76 complex. *Mol. Cell. Biol.* 37:e00071-17. doi: 10.1128/MCB.00071-17

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., et al. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* 18:229. doi: 10.1186/s12864-017-3610-0

Charboneau, A., East, L., Mulholland, N., Rohde, M., and Boudreau, N. (2006). Pbx1 is required for Hox D3-mediated angiogenesis. *Angiogenesis* 8, 289–296. doi: 10.1007/s10456-005-9016-7

Dar, M. A., Mumtaz, P. T., Bhat, S. A., Nabi, M., Taban, Q., Shah, R. A., et al. (2018). "Genetics of disease resistance in chicken," in *Application of Genetics and Genomics in Poultry Science*. ed. H. Mikkola (United Kingdom: InTech).

Dashti, M. J. S., and Gamieldien, J. (2017). A practical guide to filtering and prioritizing genetic variants. *Biotechniques* 62, 18–30. doi: 10.2144/000114492

Deeb, N., and Lamont, S. J. (2002). Genetic architecture of growth and body composition in unique chicken populations. *J. Hered* 93, 107–118. doi: 10.1093/jhered/93.2.107

Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2017). Novel mechanisms revealed in the trachea transcriptome

of resistant and susceptible chicken lines following infection with Newcastle disease virus. *Clin. Vaccine Immunol.* 24:e00027-17. doi: 10.1128/CVI.00027-17

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:R60. doi: 10.1186/gb-2003-4-9-r60

Dorshorst, B. J., Siegel, P. B., and Ashwell, C. M. (2011). Genomic regions associated with antibody response to sheep red blood cells in the chicken. *Anim. Genet.* 42, 300–308. doi: 10.1111/j.1365-2052.2010.02146.x

Doudna, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

Dufner, A., Kisser, A., Niendorf, S., Basters, A., Reissig, S., Schonle, A., et al. (2015). The ubiquitin-specific protease USP8 is critical for the development and homeostasis of T cells. *Nat. Immunol.* 16, 950–960. doi: 10.1038/ni.3230

Ekine, C. C., Rowe, S. J., Bishop, S. C., and de Koning, D. J. (2014). Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3* 4, 341–347. doi: 10.1534/g3.113.008706

El Sayed, M. S., Ahmed, A. S., and Alyousef, Y. M. (2016). Characterization of Newcastle disease antibody response and some related performance indicators of two local Saudi chicken lines and two cross lines during the rearing period. *J. Anim. Plant Sci.* 26, 1236–1241.

Elbeltagy, A. R., Bertolini, F., Fleming, D. S., Van Goor, A., Ashwell, C. M., Schmidt, C. J., et al. (2019). Natural selection footprints among African chicken breeds and village ecotypes. *Front. Genet.* 10:376. doi: 10.3389/fgene.2019.00376

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Ewald, S. J., Kapczynski, D. R., Livant, D. R.., Suarez, D. L., Ralph, J., Mcleod, S., et al. (2011). Association of Mx1 Asn631 variant alleles with reductions in morbidity, early mortality, viral shedding, and cytokine response in chickens infected with a highly pathogenic avian influenza virus. *Immunogenetics.* 63, 363–375. doi: 10.1007/s00251-010-0509-1

Fischer, U., Koppang, E. O., and Nakanishi, T. (2013). Teleost T and NK cell immunity. *Fish Shellfish Immunol.* 35, 197–206. doi: 10.1016/j.fsi.2013.04.018

Fleming, D. S., Koltes, J. E., Markey, A. D., Schmidt, C. J., Ashwell, C. M., Rothschild, M. F., et al. (2016). Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600 k genotyping array. *BMC Genomics* 17:407. doi: 10.1186/s12864-016-2711-5

Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575

Gu, X., Feng, C., Ma, L., Song, C., Wang, Y., Da, Y., et al. (2011). Genome-wide association study of body weight in chicken F2 resource population. *PLoS One* 6:e21872. doi: 10.1371/journal.pone.0021872

Habimana, R., Ngeno, K., Shyaka, A., and Ntawubizi, M. (2020b). Growth performance and immune response to Newcastle disease in four gene pools of indigenous chicken in Rwanda. *Genet. Resour.* 1, 42–50. doi: 10.46265/genresj.LPJS9396

Habimana, R., Okeno, T. O., Ngeno, K., Mboumba, S., Assami, P., Gbotto, A., et al. (2020a). Genetic diversity and population structure of indigenous chicken in Rwanda using microsatellite markers. *PLoS One* 15:e0225084. doi: 10.1371/journal.pone.0225084

Heichman, K. A., and Roberts, J. M. (1998). CDC16 controls initiation at chromosome replication origins. *Mol. Cell* 1, 457–463. doi: 10.1016/S1097-2765(00)80046-5

Hirose, F., Yamaguchi, M., Kuroda, K., Omori, A., Hachiya, T., Ikeda, M., et al. (1996). Isolation and characterization of cDNA for DREF, a promoter-activating factor for *Drosophila* DNA replication-related genes. *J. Biol. Chem.* 271, 3930–3937. doi: 10.1074/jbc.271.7.3930

Hu, Z. L., Park, A. C., Fritz, E. R., and Reecy, J. M. (2010). "QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes." *Invited Lecture With a Full Paper Published Electronically on The 9th World Congress on Genetics Applied to Livestock Production.* August 1–6, 2010; Leipzig, Germany.

Ioannidis, J. P. A., Thomas, G., and Daly, M. J. (2009). Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* 10, 318–329. doi: 10.1038/nrg2544

Islam, A. S. M. F., and Blair, M. W. (2018). Molecular characterization of mung bean germplasm from the USDA core collection using newly developed KASP-based SNP markers. *Crop Sci.* 58, 1659–1670. doi: 10.2135/cropsci2018.01.0044

Jacob, J. P., Milne, S., Beck, S., and Kaufman, J. (2000). The major and a minor class II β chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex. *Immunogenetics* 51, 138–147. doi: 10.1007/s002510050022

Jacobsson, L., Park, H. B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P. B., et al. (2005). Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genet. Res.* 86, 115–125. doi: 10.1017/S0016672305007767

Jennen, D. G. J., Vereijken, A. L. J., Bovenhuis, H., Crooijmans, R. P. M. A., Veenendaal, A., van der Poel, J. J., et al. (2004). Detection and localization of quantitative trait loci affecting fatness in broilers. *Poult. Sci.* 83, 295–301. doi: 10.1093/ps/83.3.295

Ji, J., Luo, C. L., Zou, X., Lv, X. H., Xu, Y. B., Shu, D. M., et al. (2019). Association of host genetics with intestinal microbial relevant to body weight in a chicken F2 resource population. *Poult. Sci.* 98, 4084–4093. doi: 10.3382/ps/pez199

Jiang, S., Wang, Y., Xiong, Y., Feng, Y., Tang, J., and Song, R. (2018). High expression of ZBED1 affects proliferation and apoptosis in gastric cancer. *Int. J. Clin. Exp. Pathol.* 11, 4019–4025.

Jie, H., and Liu, Y. P. (2011). Breeding for disease resistance in poultry: opportunities with challenges. *Worlds Poult. Sci. J.* 67, 687–696. doi: 10.1017/S0043933911000766

Jin, C. F., Chen, Y. J., Yang, Z. Q., Shi, K., and Chen, C. K. (2015). A genome-wide association study of growth trait-related single nucleotide polymorphisms in Chinese Yancheng chickens. *Genet. Mol. Res.* 14, 15783–15792. doi: 10.4238/2015.December.1.30

Johansson, A. M., Pettersson, M. E., Siegel, P. B., and Carlborg, Ö. (2010). Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6:e1001188. doi: 10.1371/journal.pgen.1001188 (Accessed November 20, 2020).

Joshi, N., and Fass, J. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at: https://Github.Com/Najoshi/Sickle (Accessed November 20, 2020).

Kamps, M. P., Murre, C., Sun, X. H., and Baltimore, D. (1990). A new homeobox gene contributes to the DNA binding domain of the t(1;19) translocation protein in pre-B all. *Cell* 60, 547–555. doi: 10.1016/0092-8674(90)90658-2

Kapczynski, D. R., Afonso, C. L., and Miller, P. J. (2013). Immune responses of poultry to Newcastle disease virus. *Dev. Comp. Immunol.* 41, 447–453. doi: 10.1016/j.dci.2013.04.012

Kennedy, B. W., Quinton, M., and van Arendonk, J. A. (1992). Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70, 2000–2012. doi: 10.2527/1992.7072000x

Kerje, S., Carlborg, Ö., Jacobsson, L., Schütz, K., Hartmann, C., Jensen, P., et al. (2003). The twofold difference in adult size between the red junglefowl and white Leghorn chickens is largely explained by a limited number of QTLs. *Anim. Genet.* 34, 264–274. doi: 10.1046/j.1365-2052.2003.01000.x

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011:bar030. doi: 10.1093/database/bar030

Kirchmeyer, M., Servais, F. A., Hamdorf, M., Nazarov, P. V., Ginolhac, A., Halder, R., et al. (2018). Cytokine-mediated modulation of the hepatic miRNome: miR-146b-5p is an IL-6-inducible miRNA with multiple targets. *J. Leukoc. Biol.* 104, 987–1002. doi: 10.1002/JLB.MA1217-499RR

Lamont, S. J., Pinard-van der Laan, M. H., Cahaner, A., van der Poel, J. J., and Parmentier, H. K. (2009). "Selection for disease resistance: direct selection on the immune response," in *Poultry Genetics, Breeding and Biotechnology*. eds. W. M. Muir and S. E. Aggrey (UK: CABI Publishing), 399–418.

Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11, 241–247. doi: 10.1038/ng1195-241

Lee, K.-W., and Bogenhagen, D. F. (2014). Assignment of 2'-O-methyltransferases to modification sites on the mammalian mitochondrial large subunit 16S ribosomal RNA (rRNA). *J. Biol. Chem.* 289, 24936–24942. doi: 10.1074/jbc.C114.581868

Lee, K.-W., Okot-Kotber, C., LaComb, J. F., and Bogenhagen, D. F. (2013). Mitochondrial ribosomal RNA (rRNA) methyltransferase family members are positioned to modify nascent rRNA in foci near the mitochondrial DNA nucleoid. *J. Biol. Chem.* 288, 31386–31399. doi: 10.1074/jbc.M113.515692

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851–1858. doi: 10.1101/gr.078212.108

Lillie, M., Sheng, Z., Honaker, C. F., Dorshorst, B. J., Ashwell, C. M., Siegel, P. B., et al. (2017). Genome-wide standing variation facilitates long-term response to bidirectional selection for antibody response in chickens. *BMC Genomics* 18:99. doi: 10.1186/s12864-016-3414-7

Liu, X., Li, H., Wang, S., Hu, X., Gao, Y., Wang, Q., et al. (2007). Mapping quantitative trait loci affecting body weight and abdominal fat weight on chicken chromosome one. *Poult. Sci.* 86, 1084–1089. doi: 10.1093/ps/86.6.1084

Liu, Z., Yang, N., Yan, Y., Li, G., Liu, A., Wu, G., et al. (2019). Genome-wide association analysis of egg production performance in chickens across the whole laying period. *BMC Genet.* 20:67. doi: 10.1186/s12863-019-0771-7

Luo, C., Qu, H., Ma, J., Wang, J., Li, C., Yang, C., et al. (2013). Genome-wide association study of antibody response to Newcastle disease virus in chicken. *BMC Genet.* 14:42. doi: 10.1186/1471-2156-14-42

Lwelamira, J., Kifaro, G. C., and Gwakisa, P. S. (2009). Genetic parameters for body weights, egg traits and antibody response against Newcastle disease virus (NDV) vaccine among two Tanzania chicken ecotypes. *Trop. Anim. Health Prod.* 41, 51–59. doi: 10.1007/s11250-008-9153-2

MacKay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612

Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11, 499–511. doi: 10.1038/nrg2796

Marklund, L., Nyström, P. E., Stern, S., Andersson-Eklund, L., and Andersson, L. (1999). Confirmed quantitative trait loci for fatness and growth on pig chromosome 4. *Heredity* 82, 134–141. doi: 10.1038/sj.hdy.6884630

Matsukage, A., Hirose, F., Yoo, M. A., and Yamaguchi, M. (2008). The DRE/DREF transcriptional regulatory system: a master key for cell proliferation. *Biochim. Biophys. Acta Gene Regul. Mech.* 1779, 81–89. doi: 10.1016/j.bbagrm.2007.11.011

McElroy, J. P., Kim, J. J., Harry, D. E., Brown, S. R., Dekkers, J. C. M., and Lamont, S. J. (2006). Identification of trait loci affecting white meat percentage and other growth and carcass traits in commercial broiler chickens. *Poult. Sci.* 85, 593–605. doi: 10.1093/ps/85.4.593

McWhirter, J. R., Goulding, M., Weiner, J. A., Chun, J., and Murre, C. (1997). A novel fibroblast growth factor gene expressed in the developing nervous system is a downstream target of the chimeric homeodomain oncoprotein E2A-Pbx1. *Development* 124, 3221–3232. doi: 10.1242/dev.124.17.3221

Mebratie, W., Reyer, H., Wimmers, K., Bovenhuis, H., and Jensen, J. (2019). Genome-wide association study of body weight and feed efficiency traits in a commercial broiler chicken population, a re-visitation. *Sci. Rep.* 9:922. doi: 10.1038/s41598-018-37216-z

Mercader, N., Leonardo, E., Azplazu, N., Serrano, A., Morata, G., Martínez-A, C., et al. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* 402, 425–429. doi: 10.1038/46580

Mpenda, F. N., Lyantagaye, S. L., and Buza, J. (2020). Association of chicken Mx1 polymorphisms with susceptibility in chicken embryos challenged with virulent Newcastle disease virus. *Asian J. Anim. Sci.* 14, 9–15. doi: 10.3923/ajas.2020.9.15

Noorai, R. E., Shankar, V., Freese, N. H., Gregorski, C. M., and Chapman, S. C. (2019). Discovery of genomic variations by whole-genome resequencing of the north American Araucana chicken. *PLoS One* 14:e0225834. doi: 10.1371/journal.pone.0225834

Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D., et al. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell* 60, 535–545. doi: 10.1016/0092-8674(90)90657-Z

Paglialunga, S., Simnett, G., Robson, H., Hoang, M., Pillai, R., Arkell, A. M., et al. (2017). The RAB-GTPase activating protein, TBC1D1, is critical for maintaining normal glucose homeostasis and β-cell mass. *Appl. Physiol. Nutr. Metab.* 42, 647–655. doi: 10.1139/apnm-2016-0585

Pértille, F., Moreira, G. C. M., Zanella, R., Nunes, J. D. R. D. S., Boschiero, C., Rovadoscki, G. A., et al. (2017). Genome-wide association study for performance traits in chickens using genotype by sequencing approach. *Sci. Rep.* 7:41748. doi: 10.1038/srep41748

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11, 459–463. doi: 10.1038/nrg2813

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Rikimaru, K., Sasaki, O., Koizumi, N., Komatsu, M., Suzuki, K., and Takahashi, H. (2011). Mapping of quantitative trait loci affecting growth traits in a Japanese native chicken cross. *Asian-Australas. J. Anim. Sci.* 24, 1329–1334. doi: 10.5713/ajas.2011.11004

Rowland, K., Wolc, A., Gallardo, R. A., Kelly, T., Zhou, H., Dekkers, J. C. M., et al. (2018). Genetic analysis of a commercial egg laying line challenged with Newcastle disease virus. *Front. Genet.* 9:326. doi: 10.3389/fgene.2018.00326

Saelao, P., Wang, Y., Chanthavixay, G., Gallardo, R. A., Wolc, A., Dekkers, J. C. M., et al. (2019). Genetics and genomic regions affecting response to Newcastle disease virus infection under heat stress in layer chickens. *Genes* 10:61. doi: 10.3390/genes10010061

Sarker, N., Tsudzuki, M., Nishibori, M., and Yamamoto, Y. (1999). Direct and correlated response to divergent selection for serum immunoglobulin M and G levels in chickens. *Poult. Sci.* 78, 1–7. doi: 10.1093/ps/78.1.1

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele-Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Sharmaa, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. H., et al. (2015). Stories and challenges of genome-wide association studies in livestock – a review. *Asian-Australas. J. Anim. Sci.* 28, 1371–1379. doi: 10.5713/ajas.14.0715

Shi, M.-T., Zhang, Y., and Zhou, G.-Q. (2018). The critical roles of TBC proteins in human diseases. *Yi Chuan* 40, 12–21. doi: 10.16288/j.yczz.17-343

Soller, M., Weigend, S., Romanov, M. N., Dekkers, J. C. M., and Lamont, S. J. (2006). Strategies to assess structural variation in the chicken genome and its associations with biodiversity and biological performance. *Poult. Sci.* 85, 2061–2078. doi: 10.1093/ps/85.12.2061

Spelman, R. J., and Bovenhuis, H. (1998). Moving from QTL experimental results to the utilization of QTL in breeding programmes. *Anim. Genet.* 29, 77–84. doi: 10.1046/j.1365-2052.1998.00238.x

Tsudzuki, M., Onitsuka, S., Akiyama, R., Iwamizu, M., Goto, N., Nishibori, M., et al. (2007). Identification of quantitative trait loci affecting shank length, body weight and carcass weight from the Japanese cockfighting chicken breed, Oh-Shamo (Japanese Large Game). *Cytogenet. Genome Res.* 117, 288–295. doi: 10.1159/000103190

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *J. Open Source Softw.* 3:731. doi: 10.21105/joss.00731

Uemoto, Y., Sato, S., Odawara, S., Nokata, H., Oyamada, Y., Taguchi, Y., et al. (2009). Genetic mapping of quantitative trait loci affecting growth and carcass traits in F2 intercross chickens. *Poult. Sci.* 88, 477–482. doi: 10.3382/ps.2008-00296

Walugembe, M., Mushi, J. R., Amuzu-Aweh, E. N., Chiwanga, G. H., Msoffe, P. L., Wang, Y., et al. (2019). Genetic analyses of Tanzanian local chicken ecotypes challenged with Newcastle disease virus. *Genes* 10:546. doi: 10.3390/genes10070546

Wang, D., Sun, Y., Stang, P., Berlin, J. A., Wilcox, M. A., and Li, Q. (2009). Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proc.* 3(Suppl. 7):S109. doi: 10.1186/1753-6561-3-s7-s109

Wang, W., Zhang, T., Zhang, G., Wang, J., Han, K., Wang, Y., et al. (2015). Genome-wide association study of antibody level response to NDV and IBV in Jinghai yellow chicken based on SLAF-seq technology. *J. Appl. Genet.* 56, 365–373. doi: 10.1007/s13353-014-0269-y

Wardęcka, B., Olszewski, R., Jaszczak, K., Zięba, G., Pierzchała, M., and Wicińska, K. (2002). Relationship between microsatellite marker alleles on chromosomes 1-5 originating from the Rhode Island red and green-legged

Partrigenous breeds and egg production and quality traits in the F2 mapping population. *J. Appl. Genet.* 43, 319–329.

Xie, L., Luo, C., Zhang, C., Zhang, R., Tang, J., Nie, Q., et al. (2012). A genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS One* 7:e30910. doi: 10.1371/journal.pone.0030910

Xue, Q., Zhang, G., Li, T., Ling, J., Zhang, X., and Wang, J. (2017). Transcriptomic profile of leg muscle during early growth in chicken. *PLoS One* 12:e0173824. doi: 10.1371/journal.pone.0173824

Yin, C. G, Zhang, C. S., Zhang, A. M., Qin, H. W., Wang, X. Q, Du, L. X., et al. (2010). Expression analyses and antiviral properties of the Beijing-You and White Leghorn myxovirus resistance gene with different amino acids at position 631. *Poult. Sci* 89, 2259–2264. doi: 10.3382/ps.2010-00826

Yiu, W. H., Poon, J. W. M., Tsui, S. K. W., Fung, K. P., and Waye, M. M. Y. (2004). Cloning and characterization of a novel endoplasmic reticulum localized G-patch domain protein, IER3IP1. *Gene* 337, 37–44. doi: 10.1016/j.gene.2004.04.033

Yonash, N., Cheng, H. H., Hillel, J., Heller, D. E., and Cahaner, A. (2001). DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chickens. *Poult. Sci.* 80, 22–28. doi: 10.1093/ps/80.1.22

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies NIH public access author manuscript. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhang, G. X., Fan, Q. C., Zhang, T., Wang, J. Y., Wang, W. H., Xue, Q., et al. (2015a). Genome-wide association study of growth traits in the Jinghai yellow chicken. *Genet. Mol. Res.* 14, 15331–15338. doi: 10.4238/2015.November.30.10

Zhang, L., Li, P., Liu, R., Zheng, M., Sun, Y., Wu, D., et al. (2015b). The identification of loci for immune traits in chickens using a genome-wide association study. *PLoS One* 10:e0117269. doi: 10.1371/journal.pone.0117269

Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., et al. (2013). Genome-wide association studies for growth and meat production traits in sheep. *PLoS One* 8:e66569. doi: 10.1371/journal.pone.0066569

Zhang, H., Shen, L., Li, Y., Xu, Z., Zhang, X., Yu, J., et al. (2019). Genome-wide association study for plasma very low-density lipoprotein concentration in chicken. *J. Anim. Breed. Genet.* 136, 351–361. doi: 10.1111/jbg.12397

Zhang, H., Shen, L. Y., Xu, Z. C., Kramer, L. M., Yu, J. Q., Zhang, X. Y., et al. (2020). Haplotype-based genome-wide association studies for carcass and growth traits in chicken. *Poult. Sci.* 99, 2349–2361. doi: 10.1016/j.psj.2020.01.009

Zhang, H., Wang, Z., Wang, S., and Li, H. (2012). Progress of genome-wide association study in domestic animals. *J. Anim. Sci. Biotechnol.* 3:26. doi: 10.1186/2049-1891-3-26

Zhao, X. H., Wang, J. Y., Zhang, G. X., Wei, Y., Gu, Y. P., and Yu, Y. B. (2012). Single nucleotide polymorphism in the STAT5b gene is associated with the body weight and reproductive traits of the Jinghai yellow chicken. *Mol. Biol. Rep.* 39, 4177–4183. doi: 10.1007/s11033-011-1202-7

Zhou, H., and Lamont, S. J. (2003). Chicken MHC class I and II gene effects on antibody response kinetics in adult chickens. *Immunogenetics* 55, 133–140. doi: 10.1007/s00251-003-0566-9

Zhou, J., Qiu, L., Jiang, S., Zhou, F., Huang, J., Yang, L., et al. (2013). Molecular cloning and mRNA expression of M-phase phosphoprotein 6 gene in black tiger shrimp (*Penaeus monodon*). *Mol. Biol. Rep.* 40, 1301–1306. doi: 10.1007/s11033-012-2173-z

Zhou, X. H., Yang, C. Q., Zhang, C. L., Gao, Y., Yuan, H. B., and Wang, C. (2014). RASSF5 inhibits growth and invasion and induces apoptosis in osteosarcoma cells through activation of MST1/LATS1 signalling. *Oncol. Rep.* 32, 1505–1512. doi: 10.3892/or.2014.3387

Zhuang, Z., Xu, L., Yang, J., Gao, H., Zhang, L., Gao, X., et al. (2020). Weighted single-step genome-wide association study for growth traits in Chinese Simmental beef cattle. *Genes* 11:189. doi: 10.3390/genes11020189

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 August 2021
doi: 10.3389/fgene.2021.723964

Mitochondrial Sequence Variation, Haplotype Diversity, and Relationships Among Dromedary Camel-Types

*Randa Alaqeely[1], Bader H. Alhajeri[1], Faisal Almathen[2,3] and Hasan Alhaddad[1]**

[1] Department of Biological Sciences, Kuwait University, Kuwait City, Kuwait, [2] Department of Veterinary Public Health, College of Veterinary Medicine, King Faisal University, Al-Ahsa, Saudi Arabia, [3] The Camel Research Center, King Faisal University, Al-Ahsa, Saudi Arabia

Dromedary camels are outstanding livestock that developed efficient abilities to tolerate desert conditions. Many dromedary camel-types (i.e., named populations) exist but lack defined specific breed standards, registries, and breeders' governing organizations. The breed status of dromedary camel-types can partly be assessed by exploring mitochondrial DNA (mtDNA) variation. Accordingly, this study aimed to examine the breed status and the inter-population relationships of dromedary camel-types by analyzing sequence variation in the mtDNA control region and in three coding genes [*cytochrome b*, *threonine, and proline tRNA*, and part of the displacement loop (D-loop)] (867 bp region). Tail hair samples (n = 119) that represent six camel-types from Kuwait were collected, extracted, sequenced, and compared to other publicly available sequences (n = 853). Within the sequenced mitochondrial region, 48 polymorphic sites were identified that contributed to 82 unique haplotypes across 37 camel-types. Haplotype names and identities were updated to avoid previous discrepancies. When all sequences were combined (n = 972), a nucleotide diversity of 0.0026 and a haplotype diversity of 0.725 was observed across the dromedary-types. Two major haplogroups (A and B) were identified and the B1 haplotype was predominant and found in almost all dromedary-types whereas the A haplotypes were more abundant in African regions. Non-metric multidimensional scaling revealed an increased similarity among Arabian Peninsula "Mezayen" camel-types, despite their defining coat colors. The relationships among dromedary camel-types can partly be explained by mtDNA. Future work aimed at a deeper understanding of camel-type breed status should focus on a high number of nuclear markers.

Keywords: camel, mtDNA, haplogroup, polymorphism, population

OPEN ACCESS

Edited by:
Younes Miar, Dalhousie University, Canada

Reviewed by:
Martha Nchang Bemji, Federal University of Agriculture, Abeokuta, Nigeria
Guillermo Giovambattista, CONICET Institute of Veterinary Genetics (IGEVET), Argentina

****Correspondence:***
Hasan Alhaddad
hassan.alhaddad@ku.edu.kw;
hhalhaddad@gmail.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 June 2021*
Accepted: *12 August 2021*
Published: *30 August 2021*

Citation:
Alaqeely R, Alhajeri BH, Almathen F and Alhaddad H (2021) Mitochondrial Sequence Variation, Haplotype Diversity, and Relationships Among Dromedary Camel-Types. Front. Genet. 12:723964. doi: 10.3389/fgene.2021.723964

INTRODUCTION

The dromedary, *Camelus dromedarius*, is well-known for its adaptations to harsh desert conditions. The adaptations include structural (Alsafy et al., 2013; Achaaban et al., 2016), physiological (Adamsons et al., 1956; Schmidt-Nielsen et al., 1956), and behavioral traits (Mitchell et al., 2002; Djazouli Alim et al., 2012). The natural adaptations of the dromedaries were anthropologically

exploited via (1) its domestication around 2000–3000 years B.C., (2) the expansion of their uses, and (3) the development of unique populations (i.e., camel-types) (Uerpmann and Uerpmann, 2002; Almathen et al., 2016; Orlando, 2016). However, unlike other domesticated animals (e.g., cattle, sheep, horses, dogs, and cats), dromedaries do not currently have breed definitions, standards, registries, or breeders' organizations (Arman, 2007; Lynghaug, 2009; Alhaddad and Alhajeri, 2019). Named dromedary populations are locally known and occasionally documented, but little is known about their breed status. As a result, named camel populations are referred to here as "camel-types" instead of breeds (Alaskar et al., 2021). Porter et al. (2016) has reported about 200 different camel-types, yet many displayed overlapping characteristics and thus may include types with synonymous names. Using a few STR markers, AlAskar et al. (2020) partially explored the breed status of dromedary camel-types; the inconclusive conclusions suggested exploring the populations using mitochondrial DNA (mtDNA) sequence variation or a higher number of nuclear markers.

Mitochondrial DNA variation can be used to gain a better understanding of dromedary populations, types, evolution, and domestication history. Analyses of mtDNA variation have been used to identify (1) maternal lineages (Jansen et al., 2002; Yang et al., 2018), (2) wild ancestry (Kadwell et al., 2001; Naderi et al., 2008; Kimura et al., 2010), and (3) geographic origins of different species (Cieslak et al., 2010; Di Lorenzo et al., 2015; Almathen et al., 2016). With a focus on domesticated animals, mtDNA variation have been used to study the breed relationships of Bactrian camels (Ming et al., 2017), horses (Hristov et al., 2017), donkeys (Cozzi et al., 2018), goats (Kibegwa et al., 2016), and cattle (Di Lorenzo et al., 2018). The mitochondrial genome of dromedary camels is ∼16.6 kb in length and consists of genes encoding *tRNAs* (22 genes), *rRNAs* (2 genes), *sequence tagged sites* (*STS*) (3 sites), *NADH dehydrogenase* (7 genes), *cytochrome c oxidase* (3 genes), *ATP synthase* (2 genes), and *cytochrome b* (a single gene) (GenBank accession number: NC_009849, Huang et al., unpublished). The control region is the longest within the mitochondrial genome (1,124 bp) and the most variable non-coding region (Stoneking et al., 1991). The displacement loop (D-loop), which is located within the control region, exhibits the highest levels of polymorphism and accordingly is used for evolutionary studies (McMillan and Palumbi, 1997). mtDNA variation has been investigated in dromedary camels with a focus on the D-loop in addition to *tRNA* (mostly proline and threonine) and *cytochrome b* sequences. Using this localized sequence variation, Almathen et al. (2016) found that dromedary populations, combined based on country of origin, exhibited no clear phylogeographic clustering. Also, the authors reported two major haplogroups (A and B), which consisted of 76 haplotypes of different frequencies (Almathen et al., 2016). Nonetheless, the naming and assignment of haplotypes into haplogroups lacked a clear methodology, and the haplotype sequences and positions of mutation(s) were not reported.

The objectives of this study were to: (1) re-examine the molecular variation within an mtDNA region using 972 dromedary samples, (2) evaluate the molecular variation within and among dromedary camel-types, (3) classify and report haplogroups and haplotypes, (4) investigate the relationship between the identified haplotypes and the camel-type's naming system and geography, (5) test the hypothesis that having an apparent selection criteria in specific camel-types may affect haplotype variability, and (6) evaluate the relationships within "Mezayen" camel-types also known as "beauty pageant camels."

MATERIALS AND METHODS

Samples and DNA Extraction

Tail-hair specimens (n = 119) of unrelated dromedary camels were selected from the Cdrom Archive (Alhaddad and Alhajeri, 2018, 2019) for the current study. Relatedness was avoided not only by looking at the information associated with each dromedary camel and its pedigree, but also by avoiding the inclusion of more than one sample per breeder when possible. Selected samples belonged to six dromedary camel-types: Majaheem, Sofor, Shaele, Shageh, Homor, and Waddah (Porter et al., 2016; **Supplementary Table 1**). DNA was extracted from approximately 30 tail-hair follicles using a DNA extraction kit (PureLink Genomic DNA Mini Kit, Thermo Fisher Scientific) following an established protocol (Alhaddad et al., 2019). The quality of the extracted DNA was evaluated using a 1.5% agarose gel and the quantity and purity of the extracted DNA was assessed using eight channel nanodrop spectrophotometry (NanoDrop^{TM} 8000 Spectrophotometer, Thermo Fisher Scientific^{TM}) at the Biotechnology Center at Kuwait University.

Amplifying and Sequencing of the Target Region Using PCR

A mtDNA fragment of 867 bp length was amplified using a primer-pair previously designed and published Almathen et al. (2016): CB_F 5′ CCTAGCATTTATCCCCGCACTA3′ and tPRO_R 5′ GGTTGTATGATGCGGGTAAATG 3′. This fragment included the end of *cytochrome b* (184 bp), transfer RNA *threonine and proline* (134 bp), and the beginning of a control region spanning STRs (549 bp). PCR reaction was carried out in a total volume of 20 μl containing: 4–84 ng genomic DNA, 0.6 μM of each primer, 10 μl of *Taq* PCR Master Mix Kit (Qiagen^{TM}) and completed to the final volume with nuclease free water. The PCR cycle was as follows: (1) an initial denaturation step at 94°C for 3 min, (2) 40 cycles each of denaturation at 94°C for 30 s, annealing at 60°C for 45 s, extension at 72°C for 90 s, and (3) a final extension step at 72°C for 5 min. The amplified DNA product was visualized in a 1.5% agarose gel then purified using ExoSAP-IT^{TM} PCR Product Cleanup Reagent (Applied Biosystems^{TM}) as recommended by the manufacturer.

Using both forward and reverse PCR primers, independently, the PCR product was sequenced using Sanger sequencing (BigDye Terminator v3.1 Cycle Sequencing Kit, Thermo Fisher Scientific) following the manufacturer's protocol. The sequencing reaction was carried out in a final volume of 20 μl containing: (1) 8 μl of BigDye Terminator Master Mix, (2) 2 μl of each primer (0.6 μM) in two separate reactions, (3) 8 μl of deionized water, and (4) 2 μl of the amplified PCR product. Sequencing products were purified using the BigDye

XTerminator™ Purification Kit (Applied Biosystems™) and its protocol. Sequences were analyzed using ABI 3130XL Genetic Analyzer at the Biotechnology Center in Kuwait University and submitted to GenBank (accession numbers MT164347 – MT164465).

Sequence Quality and Alignment

Each sequence was visually inspected for quality, and only sequences with a clear chromatogram were included in the downstream analyses. Sequences of each individual (one using the forward primer and one using the reverse primer) were subjected to manual editing and cleaning using FinchTV (FinchTV® 1.5.0, Geospiza, Inc., Seattle, WA, United States). Cleaned forward and reverse sequences were aligned to obtain the consensus sequences using BioEdit v.7.2.5 (Hall, 2013). A multiple sequence alignment was created for the generated consensus sequences of all samples using the CLUSTALW method as implemented in MEGAX (Kumar et al., 2018). Although the flanking regions were obtained using designed primers, sequences were cropped to match/align them with publicly available sequences of previously established studies (Almathen et al., 2016).

Beside the generated sequences in this study, publicly available sequences (n = 759) were retrieved and used in this study (Accession numbers JX946206-JX946273 and KF719283-KF719290) (Almathen et al., 2016) in addition to 95 unpublished sequences obtained from Saudi Arabia (Almathen, unpublished). The combined sequences belonged to 37 dromedary camel-types from 21 countries (**Supplementary Figure 1** and **Supplementary Table 2**).

Nucleotide and Haplotype Diversities

Sequences were assigned into mitochondrial haplotypes (mitotypes) using DnaSP software v 6.0, (Rozas et al., 2017). Using *Pegas* package (Paradis, 2010) in R software (R Development Core Team, 2018), the nucleotide diversity (Nei, 1987) was calculated using the function (*nuc.div*) and the haplotype diversity using (*hap.div*). To compare the effect of using D-loop independently to the use of the D-loop in addition to coding genes in terms of mitochondrial haplotype assignment, two sequence sets were created. The first included the whole 867 bp mtDNA (including *cytochrome b*, *threonine, and proline tRNA*, and D-loop) for all sequences. The second set of sequences included only the control (D-loop) region (552 bp in length).

Haplotype frequencies were analyzed in relation to geography and the dromedary camel-type naming system. Mitochondrial haplotypes with a frequency less than 0.019 were considered as "low frequency." The aforementioned cut-off was calculated using the sample size formula ($ss = \frac{Z^2 \times p \times (1-p)}{c^2}$), where: $Z = Z$ value (e.g., 1.96 for 95% confidence level), p, population proportion expressed as a decimal (0.5 was used), c, confidence interval, expressed as a decimal (e.g., 0.04 = ± 4). So, when applying a confidence level of 95%, confidence interval of 20, and population size equals the total obtained haplotypes of 82, it was found that the minimum required sample size for a haplotype to be 19 individuals.

Amendment of the Existing Haplotypes Nomenclature

The initial haplogroup naming system (A and B) was retained as previously reported (Almathen et al., 2016) to avoid confusion. The major haplotypes extending from haplogroups A and B were renamed based on sequence similarity to the major groups and their frequencies. For example, five major haplotypes were identified in haplogroup B, with frequencies (in percentage) equal to 50.5, 7.7, 7.1, 6.8, and 2.9, which were named as B1, B2, B3, B4, and B5, respectively. Haplotypes with frequencies less than 1.9% (this cut-off is based on the sample size formula – see above) were named sequentially following the names of the major haplotypes. For instance, haplotypes that directly originated from B1 were given names sequentially from B6 to B29, and haplotypes that originated from B2 were named starting with B30. A total of 82 haplotypes were assigned to the two haplogroups, where haplogroup A had 24, and haplogroup B had 58 haplotypes.

Mitochondrial DNA Relationships

Phylogenetic relationships between identified haplotypes were inferred using the Bayesian method, as implemented in MRBAYES software v 3.2.7a (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012). HKY + I + G (Hasegawa et al., 1985) nucleotide substitution model with gamma correction (α = 0.0221) was used as the best fitting model for the 82 identified haplotypes based on lowest Akaike Information Criterion with correction for small sample size (AICc) (Akaike, 1974) value using jModelTest v 2.1.10 (Posada, 2008). Two independent Markov Chain Monte Carlo runs of two million generations each were used, with trees sampled every 1000 generations from the posterior distribution; the first 25% of the trees were discarded as burn-in generations (Almathen et al., 2016). Clade support was determined using Bayesian posterior probabilities and the same starting tree was used for each chain. Nonetheless, using different random starting trees showed similar results. The convergence and stationary of the post-burn-in tress was confirmed as well as the post-burn-in effective samples size (ESS) using TRACER v 1.7.1 (Rambaut et al., 2018). The relationships between the identified haplotypes were summarized via Median Joining Networks (MJ) using NETWORK software v 5.0.1.1 (Bandelt et al., 1999). The generated networks were modified using Network Publisher v 2.1.2.5.

The level of genetic differentiation between the dromedary camel-types was deduced by performing an Analysis of Molecular Variance (AMOVA) and by calculating mtDNA pairwise genetic differences F_{st} using ARLEQUIN v 3.5.2.2 (Excoffier and Lischer, 2010). Pairwise genetic distances between samples were visually inspected in a non-metric multidimensional scaling (NMDS) plot using the R package *vegan* (Oksanen et al., 2013).

Dromedary camel haplotypes were compared to five ancient dromedary camel samples (KT334309-KT334313), four Bactrian camel samples (KF640731, FJ792680, FJ792683, and KF640727), four wild Bactrian camel samples (FJ792685, FJ792684, EF212038, and NC_009629), seven guanaco samples (JQ754689-JQ754692, JQ754705, AY535173, and AY535174),

sixteen vicuña samples (JQ754672-JQ754688), and six horse samples (MH032886-MH032891). The relationship between dromedary camel haplotypes and other species was investigated via a neighbor joining tree with a bootstrap value based on 1000 iterations using MEGAX software (Kumar et al., 2018).

RESULTS

Mitochondrial DNA Polymorphism and Nucleotide Diversity

Over the 867 bp sequence alignment, 48 polymorphic sites (substitutions) were observed, 38 of which were parsimony informative sites (i.e., contain at least two different nucleotides and at least two of them occur with a minimum frequency of two) while the rest were singleton variable sites (i.e., contain at least two different nucleotides with one of polymorphisms in overall high frequency). The majority of the detected polymorphisms were in the D-loop region (**Figure 1** and **Supplementary Tables 3**, **4**). The identified polymorphic sites resulted in 82 unique haplotypes. Analysis of the polymorphic sites at *cytochrome b* and *threonine tRNA* genes with MAF (minor allele frequency) greater than 0.1 showed no effect of geographic separation among samples (i.e., alleles are equally present in different dromedary camel-types of different countries as well as different continents).

The average nucleotide and haplotype diversities across all studied dromedary samples were 0.0026 and 0.725, respectively. The level of nucleotide diversity ranged from 0.0005 in the Targui camel-type to 0.0054 in the Hawari camel-type, and the level of haplotype diversity ranged from 0.378 in the Shageh dromedary camel-type to 0.964 in the Omani camel-type. Kuwaiti Majaheem samples displayed signs of homogeneity, having the smallest nucleotide diversity value within the type, and a low haplotype diversity. In fact, all camel-types from Kuwait (Majaheem, Sofor, and Shaele) showed low nucleotide and haplotype diversities compared to their counterparts from Saudi Arabia (**Supplementary Figure 2**).

Haplotype Frequencies and Relationships

The already established haplotype nomenclature was slightly modified to correct discrepancies in a published study (Almathen et al., 2016). The first discrepancy was that identical sequences were assigned different names [e.g., haplotype B59 (JX946241) is identical to B73 (KF719287), B60 (JX946242) is identical to B74 (KF719288), and A65 (JX946240) is identical to A75 (KF719289)]. The second discrepancy was that two haplotypes were named after haplogroup B while belonging to haplogroup A (B69 and B76). Furthermore, nine new haplotypes were discovered in the current study (**Supplementary Figure 3**).

The investigated dromedary camel samples (n = 972 sequences) were represented by two haplogroups (A and B). Most haplotypes were classified under haplogroup B, which together contained 58 unique haplotypes (B1–B58), while haplogroup A contained 24 haplotypes (A1–A24) (**Figure 2**). Haplotype sequences (only parsimony informative sites) along with their accession numbers, and old and new names are listed in **Supplementary Table 5**. Among the 82 identified haplotypes, seven exhibited high frequency and were found in 806 dromedary camels (82.9%) and thus referred to as major haplotypes (B–B5 and A1–A2) (**Supplementary Figure 4**). A haplotypes exhibited the A allele at nucleotide position 375 (Mt reference position nt 15495) while B haplotypes showed the G allele. A Median Joining Network of the dromedary camel sequence haplotypes displayed the two haplogroups (A and B) connected to each other through a median vector (i.e., unsampled haplotype) (**Supplementary Figure 4**). Both haplogroups included a mixture of haplotypes of different frequencies in the analyzed populations. The interrelationships between mitochondrial haplotypes are illustrated in **Figure 3**.

When only D-loop (control region) was analyzed, 27 polymorphic sites were identified, which assigned the sequences to 58 haplotypes. Despite the decreased number of identified haplotypes compared to the analysis of the entire sequenced region, the proportions of haplotypes remained generally the same across dromedary camel-types and countries except for the low-frequency haplotypes (**Supplementary Figure 5**). Dromedary populations of some countries displayed a slight reduction in the proportions of A haplotypes (e.g., Egypt and Iran) (**Supplementary Figure 5**). Since the difference was minor, the entire region was used for all subsequent analyses, both to capture the maximum variation possible and to be consistent with previous studies (Almathen et al., 2016).

Haplotypes in Relation to Population Names and Geography

The analysis of the haplotype variation of dromedary camel-types considering their names revealed that those named based on phenotype (including Mezayen types) were almost homogenous with reduced observed diversity compared to other naming systems (e.g., named after a geographic region). The A haplotypes were mainly represented with low frequency haplotypes, while the predominant B haplotype was B1 (**Figure 4**). The highest haplotype variation was observed within dromedary camel-types named after regions (different countries and continents). Camel-types named after tribal affiliation showed variation mainly in haplogroup A, but these types were represented by small sample sizes (**Figure 4**).

All haplotypes were observed across all geographic regions, but A haplotypes were slightly overrepresented in the region of the Horn of Africa (e.g., Somalia, Ethiopia, and Kenya). Haplogroup B was predominant in Asian countries and dromedaries from India and Yemen showed no A haplotypes and Pakistani camels had only ∼8% A haplotypes (**Supplementary Figure 6**). The haplotype frequency pattern in individual dromedary camel-types from each country was further investigated (**Supplementary Figure 7**). Raidi and Kohi, which are both Pakistani camel-types displayed differences in haplotype pattern especially in B haplotypes where the Kohi type has a comparatively higher number of low frequency haplotypes. Similarly, the Algerian camel-types, Targui and Sahraoui, showed

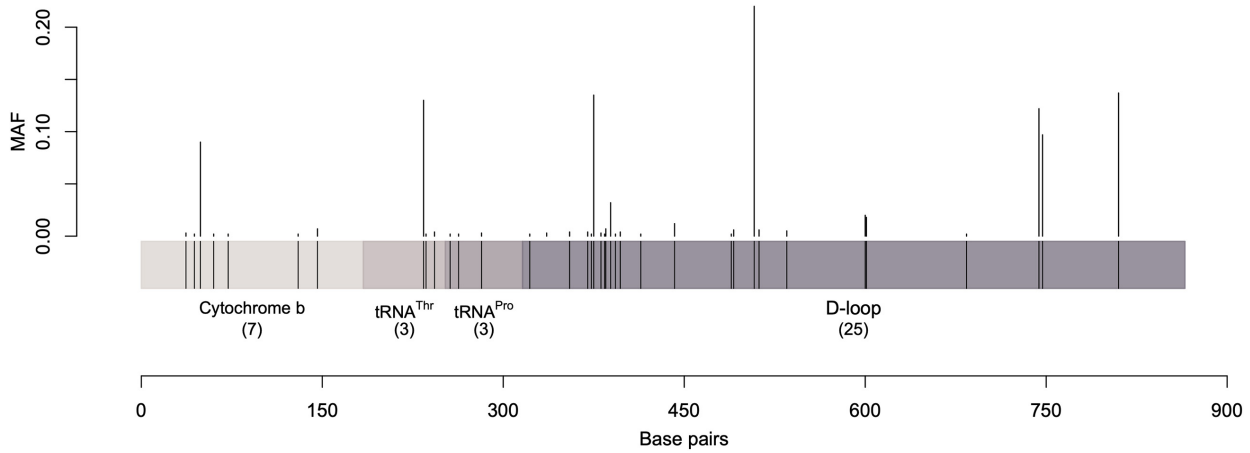

FIGURE 1 | A graphical representation of the sequenced mtDNA fragment along with the parsimony informative sites and their minor allele frequency. Vertical lines represent the minor allele frequencies (MAF) of parsimony informative sites. Singletons were excluded.

similar patterns; the Targui type had no A haplotypes and no low frequency B haplotypes (**Supplementary Figure 7**). Dromedary camel-types located in the southern regions of Africa and in the Arabian Peninsula displayed increased proportions of A haplotypes. For instance, Omani, Batinah, and Dhofar types had ≤40% A haplotypes as well as the Awarik of Saudi Arabia. Among the dromedary camel-types that lacked haplogroup A were the Baladia (Sudan), Ja (Niger), Targui (Algeria), and the Sahlia (from Saudi Arabia) (**Supplementary Figure 7**).

Pairwise Relationships Among Dromedary Camel-Types

Using the genetic difference index (F_{st}), several dromedary camel-types exhibited significant genetic differences (**Supplementary Table 6**). The Raka (Syrian camel-type) and the Baladia (Sudanese camel-type) types displayed significant genetic differences from the majority of the other camel-types, whereas the Borena (Ethiopian camel-type) type exhibited no genetic differences from all other camel-types. Among camel-types from Kuwait and Saudi Arabia, Sahlia, Hadana, Awadi, and Majaheem appeared genetically different from all other local and distant dromedary camel-types (**Supplementary Table 6**). The Sahlia type was significantly difference from most dromedary camel-types, even with other camel-types located within the same country and was the only type of Saudi Arabia that was statistically different from the Kenyan and the Nigerian camel-types (**Supplementary Table 6**). Similarly, the Pakistani camel-types, Raidi and Kohi, were also significantly difference from one another despite residing in the same geographic locality. Camel-types within Africa such as Kababish and Baladia were genetically distant from one another, while the Kababish, Kala and the Ja types did not show significant genetic differences (**Supplementary Table 6**). Dromedary camel-type differentiation was inspected visually using an NMDS plot (**Figure 5**). A distinct cluster of African dromedary camel-types was observed with only the Kurri and the Turkana camel-types separated as distinct groups. Omani, Batinah, and Dhofar types were distant from each other, despite being in the same country (Oman) (**Figure 5A**). Awarik and Hadana were both genetically distant from other dromedary camel-types of Saudi Arabia. Mezayen camel-types clustered together, with the Maghateer being most distant from the centroid of this group (**Figure 5B**).

The relationships between dromedary camel-types of the same country were visualized using separate NMDS plots. This analysis revealed that the Kuwaiti dromedary camel-types were homogenous as well as the shared camel-types between Kuwait and Saudi Arabia (i.e., Majaheem, Sofor, and Shaele) (**Supplementary Figure 8**). However, collectively dromedary camel-types from Saudi Arabia were largely heterogenous and varied greatly from each other. The Omani, Pakistani, and Algerian camel-types were distant from each other within their respective countries. Although camel-types from Sudan clustered together, they showed more variation between one another than camel-types from Kuwait for example. Camel-types from Kenya were very distant from each other on the NMDS plot compared to Sudanese camel-types (**Supplementary Figure 8**). When AMOVA test was run on the whole data set, it was found that 91.93% of genetic variation was distributed within dromedary camel-types and only 8.07% was distributed between types with F_{st} value of 0.081 and p-value < 0.001 (percentages were for the obtained covariates).

A preliminary result showed that dromedary camel haplotypes were distinct from other camelids including guanacos, vicuñas, and Bactrian camels, and formed a monophyletic group including archeological samples (**Supplementary Figure 9**). Deep clades were supported by bootstrap values greater than 50. However, most of the clade showed low bootstrap values that might be a result of relatively short sequences used. Additionally, the dromedary haplotypes were the most derived sequences. Four out of the five used archeological samples clustered with the B haplotypes. The A haplotypes were ancestral

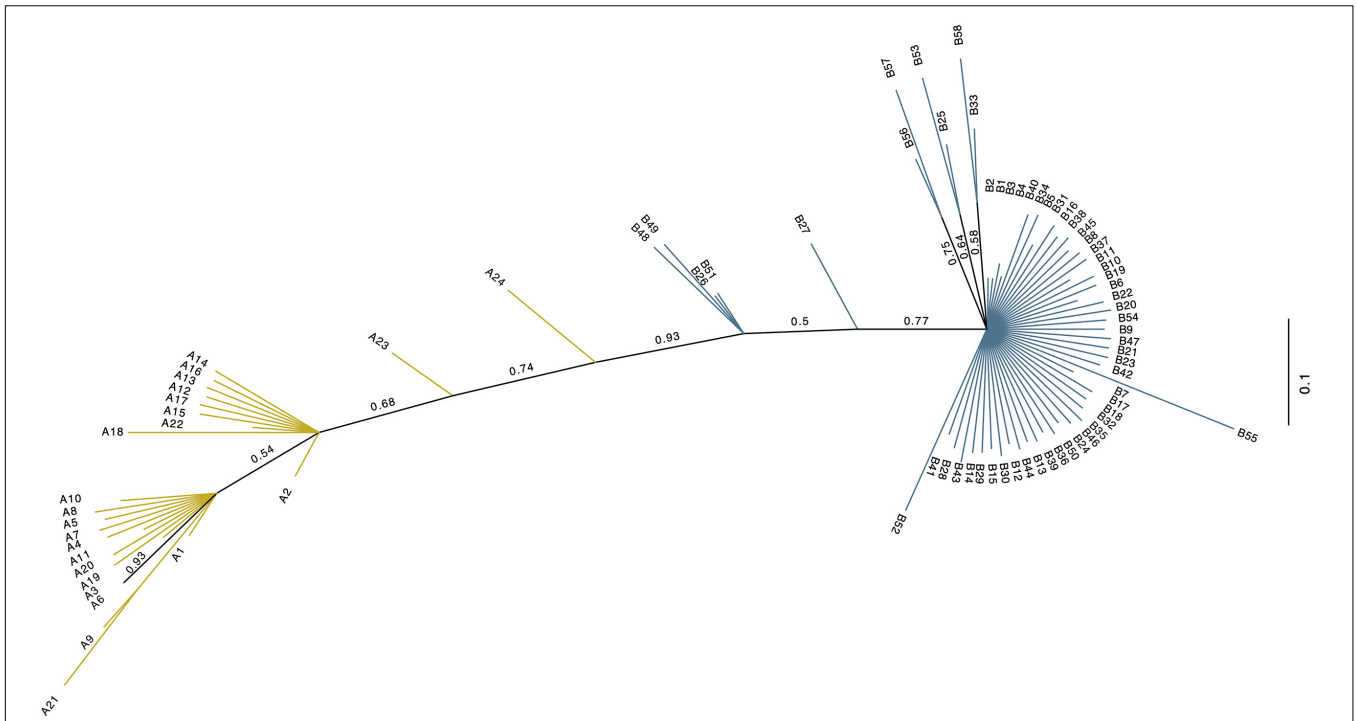

FIGURE 2 | Unrooted phylogenetic tree of the 82 dromedary camel haplotypes. The new haplotype naming system based on this study. Values above branches are posterior probabilities (PP). Blue branch lines indicate haplogroup B sequences with posterior probability values = 1, while the yellow line designate haplogroup A sequences with a posterior probability = 1. Branches colored in black have *PP* values < 1.

to the B haplotypes, despite their low frequency among the sampled dromedary camel-types. Wild Bactrian camels shared the closest ancestry to the dromedary camel haplotypes (**Supplementary Figure 9**).

Mezayen Camel-Types

The Mezayen camel-types were predominantly represented by the B haplogroup and more specifically B1 haplotypes (**Supplementary Figure 10A**). No clear distinction in haplotype frequencies was observed between the two subgroups of Mezayen camel-types [i.e., Majaheem (black colored) and Malaween (multicolored)]. The same general pattern was found in the proportions of the A haplotypes (**Supplementary Figures 10B,C**). Most Mezayen samples exhibited B haplotypes and only ∼4% were of A haplotypes (a total of eight camels had A haplotypes, of which only two samples had A1 and only one had the A2 haplotype) (**Figure 6** and **Supplementary Figure 11**). Few samples belonged to A haplotypes, and those included Shaele, Shageh, Sofor, Majaheem, and Maghateer camel-types. Among these haplotype A dromedary camels, only two samples were from Kuwait, and the rest were from Saudi Arabia (**Figure 6**). An examination of the genetic difference (F_{st}) among the Mezayen camel-types indicated a significant difference between the Majaheem camels to each of the Shaele, Homor, Shageh, and Maghateer camels (**Supplementary Table 7**). Sofor camels, on the other hand, was like all other Malaween camels except for Homor camels.

DISCUSSION

Different dromedary camel-types have been named, yet little or no documentation can be found about their breed status (Arman, 2007; Lynghaug, 2009; Porter et al., 2016; Alhaddad and Alhajeri, 2019). Several camel-type naming systems were previously described such as those based on ecotype (e.g., hill and riverine), country (e.g., Omani and Sudani), region of origin (e.g., Raka and Turkana), tribal affiliation (e.g., Kenani and Borena), and phenotype (e.g., Waddah and Shaele) (Leese, 1927; Mburu et al., 2003; Mehta et al., 2006; Ishag et al., 2010; Mahrous et al., 2011; Porter et al., 2016; Saad et al., 2017). Dromedary camel-type names currently in use in the literature exceed 200; all of which lacked registries and standard breeding criteria (Porter et al., 2016). Therefore, it is a necessary to investigate the relationships between dromedary camel-types and assess their population structure (Alaskar et al., 2021). The population structure and breed status examination of dromedary camel-types have recently been investigated using several STR markers (AlAskar et al., 2020). Analyzing mtDNA polymorphism is another way to evaluate the genetic relationships between dromedary camel-types (Hutchison Iii et al., 1974; Atig et al., 2009; Ahmed et al., 2016; Almathen et al., 2016; Saad et al., 2017). A general overview of mtDNA sequences revealed that across the sequenced region, 48 recognized variable sites were mostly transitions (Jukes, 1987; Almathen et al., 2016). As expected, most of the variable sites were found in the D-loop region, as it has lower selection pressure compared to the coding regions (McMillan and Palumbi, 1997).

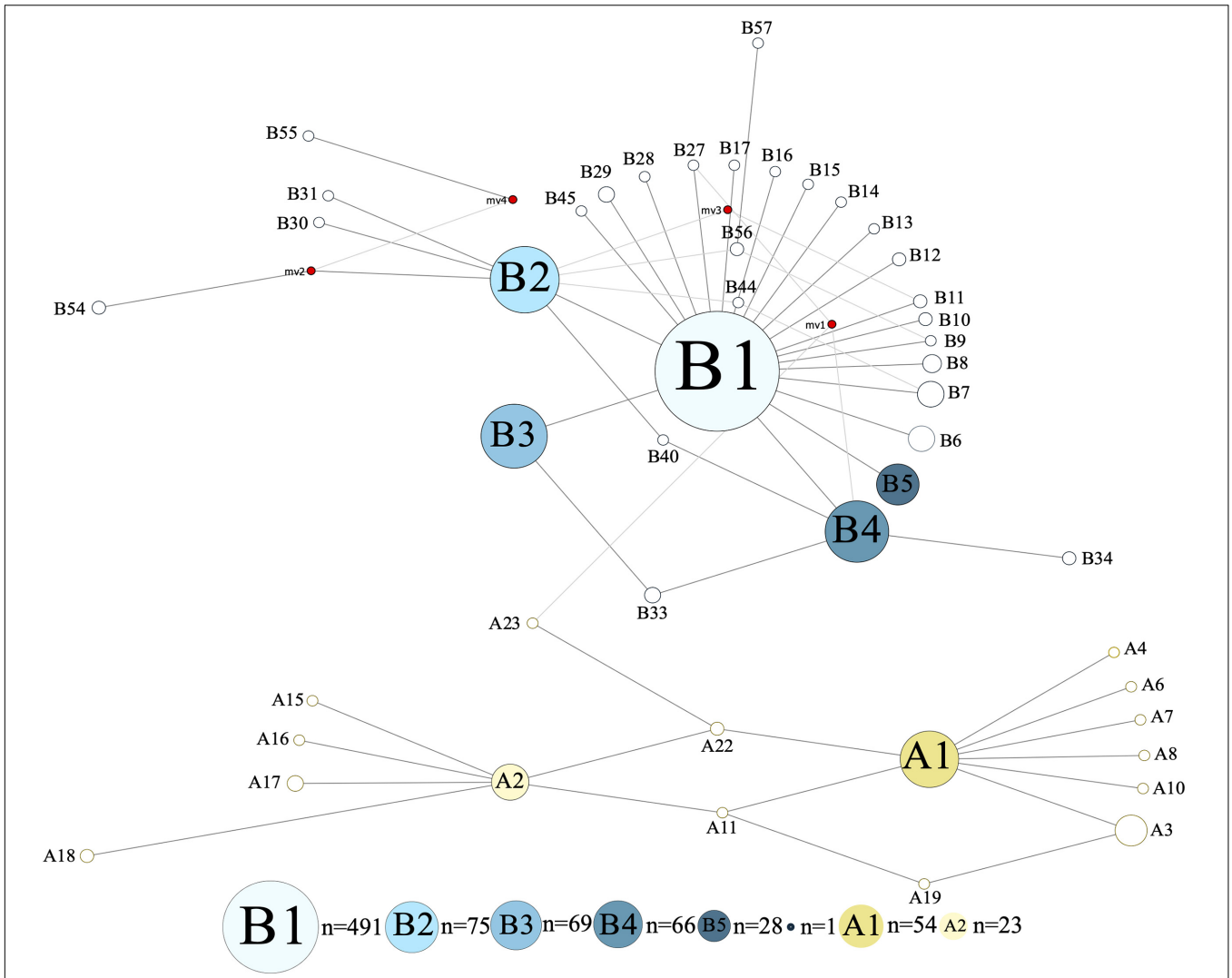

FIGURE 3 | Median Joining Network of 867 bp camel mtDNA sequences. Singletons were excluded resulting in 47 haplotypes. Red circles represent median vectors. Colors are based on haplotype identity and frequency.

Overall Genetic Diversity

Both the nucleotide and the haplotype diversity indices suggested low differences between the dromedary camel-types, as indicated by the very low nucleotide diversity (0.00026) and the moderate to high haplotype diversity (0.725). These diversity indices might indicate a much lower selection pressure on the camels when compared to other livestock that are more intensively bred for a specified function.

This study updated the existing haplotype naming system to correct the inconsistencies previously reported (Almathen et al., 2016). Two haplogroups (A and B) were identified denoting two different maternal lineages. The current study identified a total of 82 haplotypes in dromedary camels. This is relatively high when compared to Bactrians, llamas, and vicuñas, which only had 15, 17, and 57 haplotypes, respectively (Ming et al., 2017; Casey et al., 2018; González et al., 2019). However, these numbers might be correlated with the overall population size of the different camelid species, since dromedaries have a much higher population size than these other species (FAO, 2013, 2017). Most previous studies that investigated maternal lines of species or breeds used the D-loop region for its high molecular variability (Kavar et al., 1999; Doosti and Dehkordi, 2011; Kawabe et al., 2014). This study used additional mitochondrial regions (coding genes) since these additional regions did not alter the resulting haplotype frequencies (**Supplementary Figure 5**).

Relationship Between mtDNA Haplotypes and Dromedary Camel-Type Naming Criteria

Dromedary camel-types are named based on: names of individual camels, country names, ecotypes, phenotype, regions, tribal affiliation, and other criteria (Porter et al., 2016; Alaskar et al., 2021). This study aimed to test the hypothesis

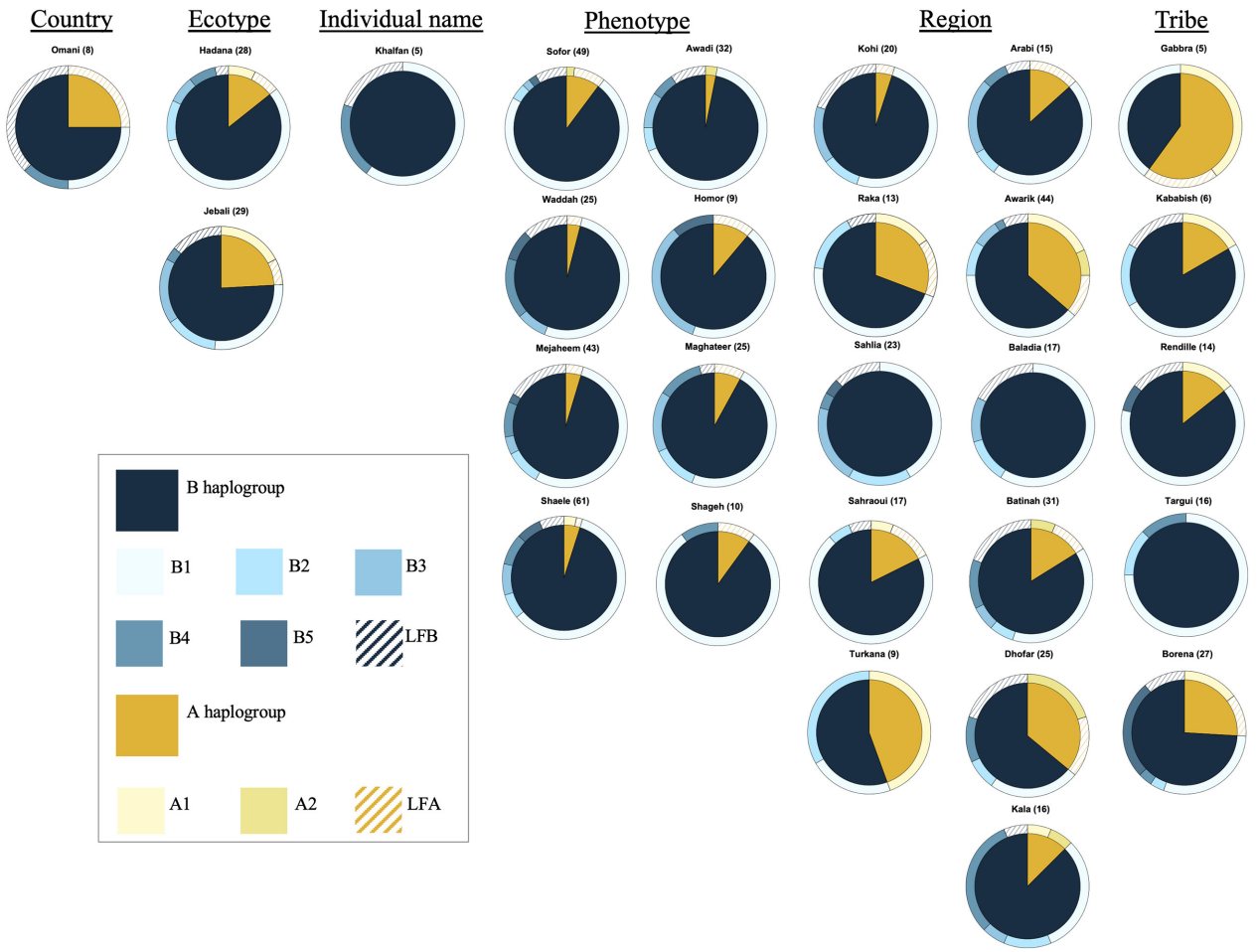

FIGURE 4 | Frequencies of haplotypes of different camel-types grouped according to their naming criteria. Camel-types represented by <5 individuals were not plotted. Camel-types with unknown naming system are not shown.

that dromedary camel-types would display haplotype variability depending on their names (AlAskar et al., 2020, Alaskar et al., 2021). Dromedary camel-types named after the name of an individual camel most likely represent lineages. An example of this camel-type is the Khalfan, which represents a racing camel-type in the UAE (Porter et al., 2016). A single haplotype was observed for the five analyzed Khalfan camels, which supports the hypothesis that types named after individuals are generally a homogenous group and represent a lineage. Dromedary camel-types named after ecotypes (e.g., Sahlia means beach camels, Hadana meaning hill camels, and Jebali meaning mountain camels) exhibited variation both in the A and the B haplotypes, indicating that these camel-types are mixtures of dromedaries of multiple origins or camel-types that experience recurrent introduction of genetically distant dromedaries. Also, this variability in haplotypes within ecotypic camel-types shows that there is no genetic uniqueness among individuals (i.e., signs of a breed). The Sahlia camel-type (meaning beach camels) differed in the haplotype frequency pattern from the other dromedary camel-types especially of the same country, mostly due to the lack of A haplotypes. This camel-type is located near the coast of the Red Sea, and is characterized by being short, and having round humps, narrow feet, and a short, thick neck (Porter et al., 2016). Therefore, it is likely that Sahlia camel-types possess distinct adaptive traits, which are suited for the high humidity of the coastal habitat. These adaptive traits and the inability of other types to flourish in the same habitat results in Sahlia's reduced gene flow with neighboring dromedary camel-types.

Although the Omani camel-type is broadly named after a country, it was genetically different from almost all other camel-types, which suggests a genetically homogeneous group that approaches the status of a true breed. Dromedary camel-types named after tribes (e.g., Rendille and Targui) and geographic regions (e.g., Awarik, Baladia, and Batinah) generally display high variability in haplotype composition, which is indicative of admixture amongst its individuals with other populations. This is in accordance with previously reported measures of genetic variability reports (Mburu et al., 2003; Legesse et al., 2018). Dromedary camel-types named based on phenotype (e.g., Majaheem, Waddah, and Awadi) are usually selectively bred for

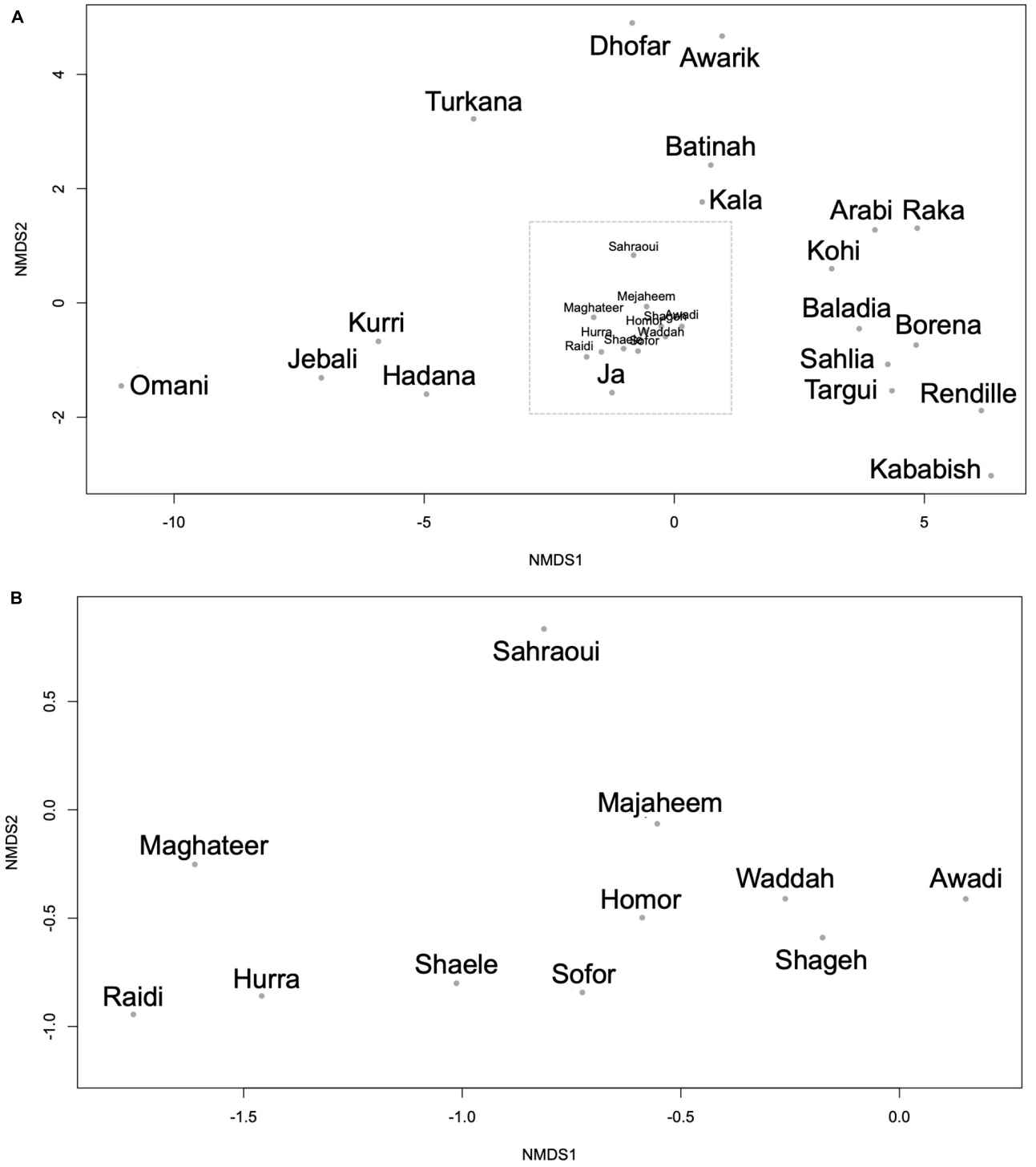

FIGURE 5 | Non-metric multidimensional scaling plot constructed using the F_{st} values between each pair of the 30 camel-types. **(A)** All camel-types. **(B)** Expanded view of the camel-types enclosed in the dashed box in panel **(A)**. Each dot represents a camel-type.

distinctive phenotypes (e.g., coat color) (Almathen et al., 2016; Porter et al., 2016; Saad et al., 2017; Alhaddad and Alhajeri, 2019). These camels formed the most homogenous group based on haplotype; with B haplotypes being overrepresented in this group. This is in accordance with previously reported findings using microsatellite data (Mahmoud et al., 2019). The handful of individuals with A haplotypes within the Mezayen camel-types signify possible crossbreeding with distant camel-types.

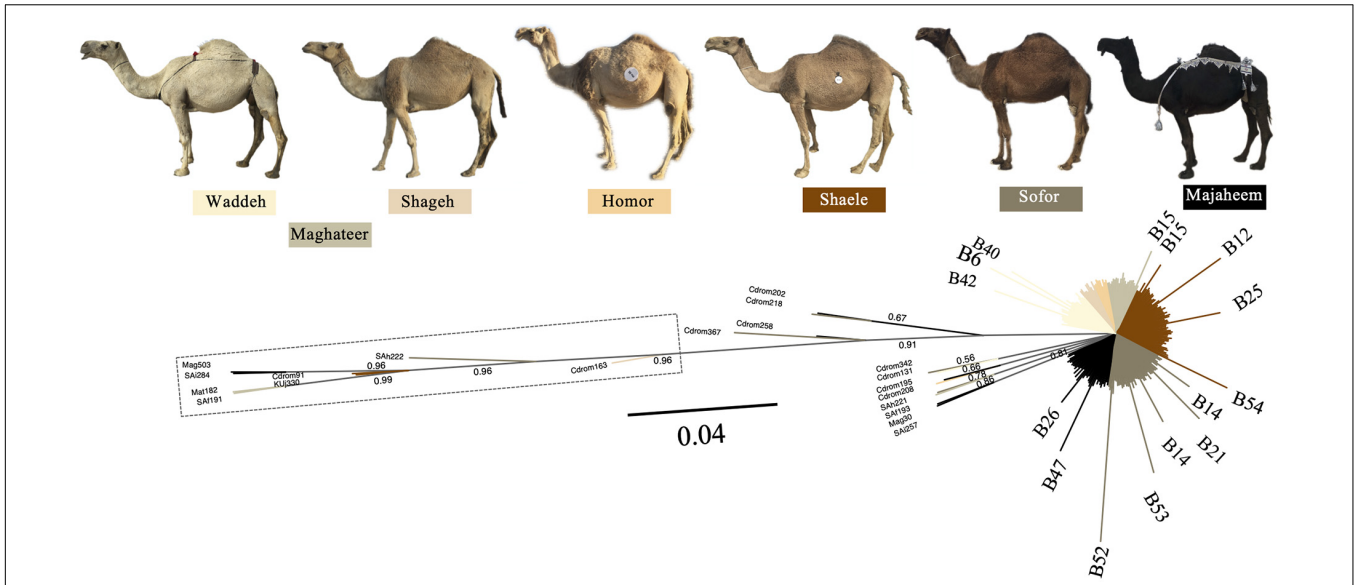

FIGURE 6 | Unrooted phylogenetic tree of Mezayen camel-type haplotypes. Values above branches correspond to the posterior probability values. Posterior probability values = 1 were omitted for clarity. The dashed box contains haplotype A sequences. Non-major (low frequency) B haplotypes were illustrated at the tip of the branches. Photos are taken by HA.

Dromedary Camel-Type Population Structure

Although no distinct phylogeographic structure was observed as previously reported (Almathen et al., 2016), A haplotypes were relatively of high frequency in African countries compared to Asian countries, except for Syria (**Supplementary Figure 7**). The low frequency of A haplotypes in Asian camels might be a result of periodic crossbreeding with African camels, since most African camel-types are of moderate to large body size (i.e., heavy pack) (Porter et al., 2016). Targui and Sahraoui are two camel-types of the Sahara (Oulad Belkhir et al., 2013) and despite occupying the same geographical locations, exhibited noticeable differences both in haplotype identity and frequency, which might be attributable to their phenotypic differences that limits intentional interbreeding (Oulad Belkhir et al., 2013). The Ja camel-type of Niger had no A haplotypes, which is unique among the dromedary camel-types of this country (Kala and Kurri). However, no genetic distinction was found when dromedary camel-types of Niger were analyzed using STR (Abdussamad et al., 2015; AlAskar et al., 2020).

Analysis of Molecular Variance results imply little or no clear sub-structuring both in relation to the dromedary camel-types and their geographical distributions. Most of the mtDNA variation was observed within the studied camel-type samples while 8% of the observed variation was among the 37 studied camel-types, which might be due to a shared origin of the two haplogroups and the continuous gene flow amongst the various dromedary camel-types. Pairwise F_{st} values revealed that some dromedary camel-types of the same geographical location (i.e., country) are genetically distant from one another (see Saudi Arabia, and Oman, etc.). This can be attributed to differences between camel owners, in selected camel qualities, and/or breeding systems (Al-Hazmi et al., 1994; Abdallah and Faye, 2012, 2013).

Mezayen Dromedary Camel-Types

In general, a low frequency of A haplotypes was found in the six Mezayen camel-types. This low frequency suggests little crossbreeding with non-Mezayen dromedary camels. The mitochondrial relationships within the six Mezayen camel-types indicate that: (1) Majaheem camels are different from the rest of the Mezayen camel-types (i.e., Malaween), (2) Maghateer camels, which is a name given to different camel groups depending on their location and tribal ownership, appears to be more diverse than all other Mezayen types, and (3) Malaween camel-types are generally similar. These findings were supported by a study on dromedary camel torso using geometric morphometric methods (Alhajeri et al., 2019). The uniqueness of Majaheem camels, and its separation from the other Mezayen types agrees with the phenotypic differences that may prevent interbreeding and gene flow (Alhaddad and Alhajeri, 2019). The observed diversity of Maghateer camels may be related to the fact that certain camel breeders define this dromedary camel-type as a mixture of Waddah and Shageh camels or even all colored Mezayen camel-types (Alhaddad and Alhajeri, 2019).

CONCLUSION

Dromedary camel mitochondrial haplotypes were more distinct than mitochondrial haplotypes in other camelids, and haplogroup A may represent the ancestral form of the more abundant B haplogroup. Little genetic difference can be observed between dromedary camel-types. The observed geographic distribution of the mitochondrial haplotypes could be due to

the physical separation of the dromedary camel-types. Also, an introgression event could have helped to introduce the A haplotypes into the Asian camel-types. Mezayen camel-types most probably represent the true breeds of the Arabian Peninsula as they exhibited homogenic haplotype mixture as well as having a set of well-identified phenotypic traits as selection criteria. The investigation of camel mtDNA is probably not sufficient to fully explain the relationships between dromedary camel-types and identify true breeds. Nuclear genome markers such as SNPs, STR, or even whole genome sequencing should be used for more comprehensive conclusions to be reached.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MT164347 – MT164465.

AUTHOR CONTRIBUTIONS

RA and HA designed the experiments. HA, BA, and FA collected the samples. RA performed the experiments and analyzed the data. RA, HA, and BA wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

RA was funded by the College of Graduate Studies (CGS), Kuwait University.

ACKNOWLEDGMENTS

We thank the Biotechnology Center at Kuwait University for providing us full access to their equipment and their staff for providing us with technical support. We are also grateful to the numerous Kuwaiti camel breeders who supplied us with the needed biological samples as well as their verbally inherited knowledge about camels.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.723964/full#supplementary-material

REFERENCES

Abdallah, H., and Faye, B. (2013). Typology of camel farming system in Saudi Arabia. *Emirates J. Food Agric.* 25, 250–260.

Abdallah, H. R., and Faye, B. (2012). Phenotypic classification of Saudi Arabian camel (*Camelus dromedarius*) by their body measurements. *Emirates J. Food Agric.* 24, 272–280.

Abdussamad, A. M., Charruau, P., Kalla, D. J. U., and Burger, P. A. (2015). Validating local knowledge on camels: colour phenotypes and genetic variation of dromedaries in the Nigeria-Niger corridor. *Livestock Sci.* 181, 131–136. doi: 10.1016/j.livsci.2015.07.008

Achaaban, M. R., Mouloud, M., Tligui, N. S., and El Allali, K. (2016). Main anatomical and histological features of the tonsils in the camel (*Camelus dromedarius*). *Trop. Anim. Health Prod.* 48, 1653–1659. doi: 10.1007/s11250-016-1139-x

Adamsons, K. Jr., Engel, S. L., Van Dyke, H. B., Schmidt-Nielsen, B., and Schmidt-Nielsen, K. (1956). The distribution of oxytocin and vasopressin (antidiuretic hormone) in the neuro-hypophysis of the camel. *Endocrinology* 58, 272–278. doi: 10.1210/endo-58-2-272

Ahmed, S., Grobler, P., Madisha, T., and Kotze, A. (2016). Mitochondrial D-loop sequences reveal a mixture of endemism and immigration in Egyptian goat populations. *Mitochond. DNA Part A* 28, 711–716. doi: 10.3109/24701394.2016.1174225

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19, 716–723.

AlAskar, H., Alhajeri, B. H., Almathen, F., and Alhaddad, H. (2020). Genetic diversity and population structure of dromedary camel-types. *J. Hered.* 111, 405–413. doi: 10.1093/jhered/esaa016

Alaskar, H. M., Alaqeely, R., Alhajeri, B. H., and Alhaddad, H. (2021). The enigma of camel-types: localities, utilities, names, and breed statuses. *J. Camelid Sci.* 14, 22–34.

Alhaddad, H., and Alhajeri, B. H. (2018). SamplEase: a simple application for collection and organization of biological specimen data in the field. *Ecol. Evol.* 8, 10266–10271. doi: 10.1002/ece3.4503

Alhaddad, H., and Alhajeri, B. H. (2019). Cdrom archive: a gateway to study camel phenotypes. *Front. Genet.* 10:48. doi: 10.3389/fgene.2019.00048

Alhaddad, H., Maraqa, T., Alabdulghafour, S., Alaskar, H., Alaqeely, R., Almathen, F., et al. (2019). Quality and quantity of dromedary camel DNA sampled from whole-blood, saliva, and tail-hair. *PLoS One* 14:e0211743. doi: 10.1371/journal.pone.0211743

Alhajeri, B. H., Alaqeely, R., and Alhaddad, H. (2019). Classifying camel breeds using geometric morphometrics: a case study in Kuwait. *Livestock Sci.* 230:103824. doi: 10.1016/j.livsci.2019.103824

Al-Hazmi, M., Ghandour, A., and ElGohar, M. (1994). A study of the biometry of some breeds of arabian camel (*camelus dromedarius*) in Saudi Arabia. *Science* 6, 87–99.

Almathen, F., Charruau, P., Mohandesan, E., Mwacharo, J. M., Orozco-terWengel, P., Pitt, D., et al. (2016). Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6707–6712. doi: 10.1073/pnas.1519508113

Alsafy, M., El-Gendy, S., and El Sharaby, A. (2013). Anatomic reference for computed tomography of paranasal sinuses and their communication in the Egyptian buffalo (*Bubalus bubalis*). *Anat. Histol. Embryol.* 42, 220–231. doi: 10.1111/ahe.12005

Arman, K. (2007). A new direction for kennel club regulations and breed standards. *Can. Vet. J.* 48, 953–965.

Atig, R. K., Hsouna, S., Beraud-Colomb, E., and Abdelhak, S. (2009). Mitochondrial DNA: properties and applications. *Arch. Inst. Pasteur Tunis* 86, 3–14.

Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.

Casey, C. S., Orozco-terWengel, P., Yaya, K., Kadwell, M., Fernández, M., Marín, J. C., et al. (2018). Comparing genetic diversity and demographic history in co-distributed wild South American camelids. *Heredity* 121, 387–400. doi: 10.1038/s41437-018-0120-z

Cieslak, M., Pruvost, M., Benecke, N., Hofreiter, M., Morales, A., Reissmann, M., et al. (2010). Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5:e15311. doi: 10.1371/journal.pone.0015311

Cozzi, M. C., Valiati, P., Cherchi, R., Gorla, E., Prinsen, R. T. M. M., Longeri, M., et al. (2018). Mitochondrial DNA genetic diversity in six Italian donkey breeds (*Equus asinus*). *Mitochondr. DNA Part A* 29, 409–418.

Di Lorenzo, P., Ceccobelli, S., Panella, F., Attard, G., and Lasagna, E. (2015). The role of mitochondrial DNA to determine the origin of domestic chicken. *World's Poult. Sci. J.* 71, 311–318.

Di Lorenzo, P., Lancioni, H., Ceccobelli, S., Colli, L., Cardinali, I., Karsli, T., et al. (2018). Mitochondrial DNA variants of Podolian cattle breeds testify for a dual maternal origin. *PLoS One* 13:e0192567. doi: 10.1371/journal.pone.0192567

Djazouli Alim, F. Z., Rodriguez, M. J., Andrade, C., Lebaili, N., and Mahy, N. (2012). Adaptation of *Camelus dromedarius* pars nervosa of the hypophysis to winter and summer living conditions. *Folia Histochem. Cytobiol.* 50, 203–212. doi: 10.5603/FHC.2012.002

Doosti, A., and Dehkordi, P. G. (2011). Genetic polymorphisms of mitochondrial genome D-loop region in Bakhtiarian population by PCR-RFLP. *Int. J. Biol.* 3:41.

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

FAO (2013). *In Vivo Conservation of Animal Genetic Resources*. Rome: Food and Agriculture Organization of the United Nations.

FAO (2017). *Food and Agricultural Organization, United Nations.* [Online]. Available online at: http://www.fao.org/faostat/en/#data (accessed on 2018)

González, B. A., Vásquez, J. P., Gómez-Uchida, D., Cortés, J., Rivera, R., Aravena, N., et al. (2019). Phylogeography and population genetics of *Vicugna vicugna*: evolution in the arid andean high plateau. *Front. Genet.* 10:445. doi: 10.3389/fgene.2019.00445

Hall, T. (2013). *BioEdit, Version 7.2. 5*. Carlsbad, CA: Ibis Biosciences.

Hasegawa, M., Kishino, H., and Yano, T.-A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174. doi: 10.1007/bf02101694

Hristov, P., Yordanov, G., Ivanova, A., Mitkov, I., Sirakova, D., Mehandzyiski, I., et al. (2017). Mitochondrial diversity in mountain horse population from the South-Eastern Europe. *Mitochondr. DNA Part A* 28, 787–792. doi: 10.1080/24701394.2016.1186667

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Hutchison Iii, C. A., Newbold, J. E., Potter, S. S., and Edgell, M. H. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251, 536. doi: 10.1038/251536a0

Ishag, I., Reissmann, M., Peters, K., Musa, L., and Ahmed, M. (2010). Phenotypic and molecular characterization of six Sudanese camel breeds. *South Afr. J. Anim. Sci.* 40, 319–326.

Jansen, T., Forster, P., Levine, M. A., Oelke, H., Hurles, M., Renfrew, C., et al. (2002). Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10905. doi: 10.1073/pnas.152330099

Jukes, T. H. (1987). Transitions, transversions, and the molecular evolutionary clock. *J. Mol. Evol.* 26, 87–98.

Kadwell, M., Fernandez, M., Stanley, H. F., Baldi, R., Wheeler, J. C., Rosadio, R., et al. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 268, 2575–2584.

Kavar, T., Habe, F., Brem, G., and Dovč, P. (1999). Mitochondrial D-loop sequence variation among the 16 maternal lines of the Lipizzan horse breed. *Anim. Genet.* 30, 423–430. doi: 10.1046/j.1365-2052.1999.00557.x

Kawabe, K., Worawut, R., Taura, S., Shimogiri, T., Nishida, T., and Okamoto, S. (2014). Genetic diversity of mtDNA D-loop polymorphisms in laotian native fowl populations. *Asian Austr. J. Anim. Sci.* 27, 19–23. doi: 10.5713/ajas.2013.13443

Kibegwa, F., Githui, K., Jung'a, J., Badamana, M., and Nyamu, M. (2016). Mitochondrial DNA variation of indigenous goats in Narok and Isiolo counties of Kenya. *J. Anim. Breed. Genet.* 133, 238–247.

Kimura, B., Marshall, F. B., Chen, S., Rosenbom, S., Moehlman, P. D., Tuross, N., et al. (2010). Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc. R. Soc. B Biol. Sci.* 278, 50–57.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Leese, A. S. (1927). *A Treatise on the One-Humped Camel in Health and in Disease*. Stanford: Haines.

Legesse, Y. W., Dunn, C. D., Mauldin, M. R., Ordonez-Garza, N., Rowden, G. R., Gebre, Y. M., et al. (2018). Morphometric and genetic variation in 8 breeds of Ethiopian camels (*Camelus dromedarius*). *J. Anim. Sci.* 96, 4925–4934. doi: 10.1093/jas/sky351

Lynghaug, F. (2009). *The Official Horse Breeds Standards Guide: The Complete Guide to the Standards of All North American Equine Breed Associatio*. Minneapolis, MN: Voyageur Press.

Mahmoud, A. H., Abu-Tarbush, F. M., Alshaik, M., Aljumaah, R., and Saleh, A. (2019). Genetic diversity and population genetic structure of six dromedary camel (*camelus dromedarius*) populations in Saudi Arabia. *Saudi J. Biol. Sci.* 27, 1384–1389. doi: 10.1016/j.sjbs.2019.11.041

Mahrous, K. F., Ramadan, H., Abdel-Aziem, S. H., Abd-El Mordy, M., and Hemdan, D. M. (2011). Genetic variations between camel breeds using microsatellite markers and RAPD techniques. *J. Appl. Biosci.* 39, 2626–2634.

Mburu, D., Ochieng, J., Kuria, S., Jianlin, H., Kaufmann, B., Rege, J., et al. (2003). Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Anim. Genet.* 34, 26–32. doi: 10.1046/j.1365-2052.2003.00937.x

McMillan, W. O., and Palumbi, S. R. (1997). Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae). *J. Mol. Evol.* 45, 473–484.

Mehta, S., Mishra, B., and Sahani, M. (2006). Genetic differentiation of Indian camel (*Camelus dromedarius*) breeds using random oligonucleotide primers. *Anim. Genet. Resourc.* 39, 77–88. doi: 10.1017/S1014233900002157

Ming, L., Yi, L., Sa, R., Wang, Z. X., Wang, Z., and Ji, R. (2017). Genetic diversity and phylogeographic structure of Bactrian camels shown by mitochondrial sequence variations. *Anim. Genet.* 48, 217–220. doi: 10.1111/age.12511

Mitchell, D., Maloney, S. K., Jessen, C., Laburn, H. P., Kamerman, P. R., Mitchell, G., et al. (2002). Adaptive heterothermy and selective brain cooling in arid-zone mammals. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 131, 571–585. doi: 10.1016/S1096-4959(02)00012-X

Naderi, S., Rezaei, H.-R., Pompanon, F., Blum, M. G., Negrini, R., Naghash, H.-R., et al. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17659–17664.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia university press.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., et al. (2013). Package 'vegan'. *Commun. Ecol. Pack. Vers.* 2, 1–295.

Orlando, L. (2016). Back to the roots and routes of dromedary domestication. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6588–6590. doi: 10.1073/pnas.1606340113

Oulad Belkhir, A., Chehma, A., and Faye, B. (2013). Phenotypic variability of two principal Algerian camel's populations (Targui and Sahraoui). *Emirates J. Food Agric.* 25, 231–237.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26, 419–420.

Porter, V., Alderson, L., Hall, S. J., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. Wallingford: CABI.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Saad, Y. M., El Hanafy, A. A., Alkarim, S. A., Almehdar, H. A., and Redwan, E. M. (2017). Analysis of genetic variations in camel breeds (*Camelus dromedarius*).

World Acad. Sci. Eng. Technol. Int. J. Biol. Biomol. Agric. Food Biotechnol. Eng. 11, 564–568.

Schmidt-Nielsen, B., Schmidt-Nielsen, K., Houpt, T. R., and Jarnum, S. A. (1956). Water balance of the camel. *Am. J. Physiol. Legacy Content* 185, 185–194. doi: 10.1152/ajplegacy.1956.185.1.185

Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., and Erlich, H. A. (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* 48, 370–382.

Uerpmann, H.-P., and Uerpmann, M. (2002). The appearance of the domestic camel in south-east Arabia. *J. Oman Stud.* 12, 235–260.

Yang, L., Kong, X., Yang, S., Dong, X., Yang, J., Gou, X., et al. (2018). Haplotype diversity in mitochondrial DNA reveals the multiple origins of Tibetan horse. *PLoS One* 13:e0201564. doi: 10.1371/journal.pone.0201564

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fgene.2021.723360

Species and Phenotypic Distribution Models Reveal Population Differentiation in Ethiopian Indigenous Chickens

Fasil Getachew Kebede[1,2], Hans Komen[1], Tadelle Dessie[2], Setegn Worku Alemu[2], Olivier Hanotte[2,3] and John W. M. Bastiaansen[1]**

[1] Animal Breeding and Genomics, Wageningen University & Research, Wageningen, Netherlands, [2] International Livestock Research Institute, Addis Ababa, Ethiopia, [3] Cells, Organism and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

OPEN ACCESS

Edited by:
Filippo Biscarini,
National Research Council (CNR), Italy

Reviewed by:
Tania Bobbo,
University of Padua, Italy
Gabriele Senczuk,
University of Molise, Italy
Arianna Manunza,
Institute of Agricultural Biology and Biotechnology, Italian National Research Council, Italy

***Correspondence:**
Fasil Getachew Kebede
fasilgetachew7@gmail.com
John W. M. Bastiaansen
John.Bastiaansen@wur.nl

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 10 June 2021
Accepted: 20 August 2021
Published: 08 September 2021

Citation:
Kebede FG, Komen H, Dessie T, Alemu SW, Hanotte O and Bastiaansen JWM (2021) Species and Phenotypic Distribution Models Reveal Population Differentiation in Ethiopian Indigenous Chickens. Front. Genet. 12:723360. doi: 10.3389/fgene.2021.723360

Smallholder poultry production dominated by indigenous chickens is an important source of livelihoods for most rural households in Ethiopia. The long history of domestication and the presence of diverse agroecologies in Ethiopia create unique opportunities to study the effect of environmental selective pressures. Species distribution models (SDMs) and Phenotypic distribution models (PDMs) can be applied to investigate the relationship between environmental variation and phenotypic differentiation in wild animals and domestic populations. In the present study we used SDMs and PDMs to detect environmental variables related with habitat suitability and phenotypic differentiation among nondescript Ethiopian indigenous chicken populations. 34 environmental variables (climatic, soil, and vegetation) and 19 quantitative traits were analyzed for 513 adult chickens from 26 populations. To have high variation in the dataset for phenotypic and ecological parameters, animals were sampled from four spatial gradients (each represented by six to seven populations), located in different climatic zones and geographies. Three different ecotypes are proposed based on correlation test between habitat suitability maps and phenotypic clustering of sample populations. These specific ecotypes show phenotypic differentiation, likely in response to environmental selective pressures. Nine environmental variables with the highest contribution to habitat suitability are identified. The relationship between quantitative traits and a few of the environmental variables associated with habitat suitability is non-linear. Our results highlight the benefits of integrating species and phenotypic distribution modeling approaches in characterization of livestock populations, delineation of suitable habitats for specific breeds, and understanding of the relationship between ecological variables and quantitative traits, and underlying evolutionary processes.

Keywords: chickens, local adaptation, niche and agroecology, species distribution models (SDMs), phenotypic distribution models (PDMs), phenotypic differentiation, breeds and ecotypes, poultry genetics and breeding

INTRODUCTION

Smallholder farmers in Africa keep scavenging poultry as a source of affordable animal protein and a means of income. The sustainability of this type of poultry production in tropical low-and medium-input systems depends on the availability of adaptive genotypes that can produce and thrive under adverse conditions such as climatic extremes, high prevalence of tropical diseases and parasites, and periodic feed shortage. The presence of selective pressures in these environments has led to adaptation of indigenous chicken populations to production constraints (Bettridge et al., 2018).

Local adaptation refers to local individuals having higher fitness in their environment than individuals from elsewhere (Williams, 1966). Environmental heterogeneity is known to be one of the main drivers of within species diversity and local adaptation (Darwin et al., 1858). Understanding the drivers of local adaptation provides essential information for designing research and development programs aiming at improving productivity while retaining resilience. A starting point in genetic improvement of the existing local chicken populations or in considering the introduction of new genotypes is to understand how the environment is driving local adaptation (Bettridge et al., 2018). This knowledge would allow breeding of indigenous ecotypes that are more productive under village conditions while retaining locally acceptable morphological and adaptive traits (Dana et al., 2010; Muchadeyi and Dzomba, 2017; Birhanu et al., 2021).

Present day African chickens are a result of an intricate interplay between domestication and natural selection. Ethiopia is an ecological microcosm of Africa, with a rich geomorphology, where people closely interacted with the environment and practiced agriculture for millennia. Because of its cultural diversity, geographical position, complex topography, and varying climatic patterns, the country harbors rich domestic animal biodiversity. The earliest osteological evidence of domestic chicken in Africa (921–801 BCE) was recovered from Ethiopia (Woldekiros and D'Andrea, 2017). The geomorphological landscape of the country is characterized by wide range of elevation (from −155 m to 4,620 m.a.s.l.) and diverse climate (Billi, 2015).

Recent technological advances in remote sensing and GIS, increased availability of environmental data, and improved computational power facilitate the understanding of the selective forces associated with local adaptation. Species distribution models (SDMs), implemented in MaxEnt (Phillips et al., 2006) and similar software, predict distribution of a species based on presence-only data, estimate the contribution of environmental variables, and help identify suitable habitats in current and future environments. Gheyas et al. (2021) and Lozano-Jaramillo M. et al. (2019) applied SDMs to produce suitability maps of Ethiopian chickens and identify important environmental variables associated with habitat suitability in chickens, without relating ecological differences with phenotypic variation among study populations. When used alone, SDMs treat a species as an evolutionarily homogenous entity and fail to consider possible population differences pertaining to local adaptation (Hampe, 2004). SDMs also make assumptions in their modeling approach (Wiens et al., 2009) which necessitate their combined use with additional approaches, such as phenotypic distribution models (PDMs).

Phenotypic distribution models use associations between phenotypes and environmental variables to map the phenotypes of populations within that species' distribution (Michel et al., 2017). These phenotype-environment associations, are well documented for natural populations of several wild plant and animal species (Bergmann, 1848; Clausen et al., 1940; Mayr, 1942; Cain and Sheppard, 1954; Langerhans, 2008; Phillimore et al., 2010; Maloney et al., 2012; Michel et al., 2017; Smith et al., 2017) and can be applicable to predict phenotype distribution among domestic animals.

Phenotypic differentiation represents the fraction of phenotypic variance between populations over the total phenotypic variance and helps understand evolutionary processes shaping populations (Storz, 2002; Leinonen et al., 2006; Schmid and Guillaume, 2017). With the exception of Lozano-Jaramillo A. et al. (2019) who applied PDMs to predict performance of improved chicken strains, distribution models have seldom been applied in indigenous livestock to identify environmental factors associated with phenotypic differentiation and to define their ecotypes. In contrast to introduced strains which have been subjected to intense artificial selection in a relatively short period of time, indigenous populations have been exposed to natural selection over multiple generations which permits a better understanding of evolutionary processes. Even with natural populations of animals, correlation between a phenotype and environment could be spurious if PDM are used on their own (Etterson and Shaw, 2001; Michel, 2011; Michel et al., 2017) and this requires their combination with additional analytical approaches, such as SDMs.

To overcome possible limitations in the use of SDMs in domesticated species like livestock, where humans may have interfered in the geographic distribution of the study species, we have taken corrective measures in our study design. Our sampling strategy was elaborate enough to ensure environments potentially habitable by chickens are included in sufficient sample size, while those uninhabitable are excluded in the sampling frame. We targeted random mating, nondescript indigenous chicken populations from separate livestock market-sheds, clustered along environmental gradients, to maximize ecological and phenotypic variation between sample populations. We followed a novel approach integrating SDMs with PDMs through generalized additive models (GAMs) to identify the most important environmental variables contributing to habitat suitability and evaluate their relationships with phenotypic differentiation among Ethiopian indigenous chicken populations.

MATERIALS AND METHODS

Sampling Strategy

A hybrid strategy, maximizing both environmental and geographical representativeness of sampling sites, increases statistical power by reducing false discovery rates caused by

demographic processes and confounding effects (De Mita et al., 2013; Lotterhos and Whitlock, 2015; Selmoni et al., 2020). We used a hybrid sampling strategy that covered the target area, ensuring high environmental variability, wide geographic distributions, and considering the demographic and biotic processes influencing the Ethiopian indigenous chicken populations (**Figure 1**). Chickens were sampled from four spatial gradients with a minimum distance between gradients of 500 km. Each gradient comprised three environmental clusters, primarily delineated based on elevation (400–1,800; 1,800–2,400; and 2,400–3,500 m.a.s.l.).

While we did not consider administrative regions in Ethiopia in our sampling strategy, we would like to describe the four regions covered in the present study (Amhara, Oromia, Benishangul-Gumuz, and Southern regions) to give a brief view of the geographic landscape. Gradient-I stretched from the Rift valley lowlands of northeastern Ethiopia (McConnell, 1967), along the territories of Afar region to the highlands of Wollo province within Amhara region. Gradient-II, starts from the Rift valley lowlands in central Ethiopia, crosses the highlands of Hararghe, including Mount Gara Muleta, and stretches to eastern Ethiopia within Oromia region. Gradient-III stretches from the highlands of northwestern Ethiopia and goes down to the lowlands along the Ethiopian–Sudanese border within Benishangul-Gumuz region. Gradient-IV spans from the highlands of western Ethiopia in Oromia region to the lowlands along the Ethiopian–Kenyan border in Southern region. Areas around the national borders of Ethiopia have low elevation, which gradually culminate to highland plateaus in the center of the country creating a striking contrast in agroecology (Ethiopian Mapping Authority, 1988). Geographic coordinates and phenotypic measurements were not taken from areas which are not habitable by chickens because of their unconducive environments (below 400 and above 3,500 m.a.s.l.).

We made sure that clusters within a gradient were distant by at least 100 km and the target chicken populations were sampled from households which visit isolated, i.e., not connected livestock market-sheds. The concept of market-shed refers to a geographic area, where households therein are in sufficient proximity to exchange their animals in various ways (e.g., sale, gift), most commonly traveling on foot. Each cluster along the spatial gradient constituted two to three populations. A total of 26 populations were sampled (**Figure 1** and **Supplementary Table 1**). The sampling frame is spatially evenly spread to capture high inter- and intrapopulation environmental and quantitative trait variability. The research design integrating SDMs and PDMs is presented in **Figure 2**.

The sample locations in our study covered 13 out of the total of 18 agroecological zones (MoA, 1998; Tadesse et al., 2006) in Ethiopia. Agroecological zonation utilizes biophysical attributes of soil, terrain, and climate to organize land-use types or production systems into relatively homogenous units (Hurni, 1998). The five agroecologies that were not covered, are areas where chickens have either not been reared due to extreme climates, cannot be kept at all (e.g., water bodies, undisturbed forests), or have only been recently introduced.

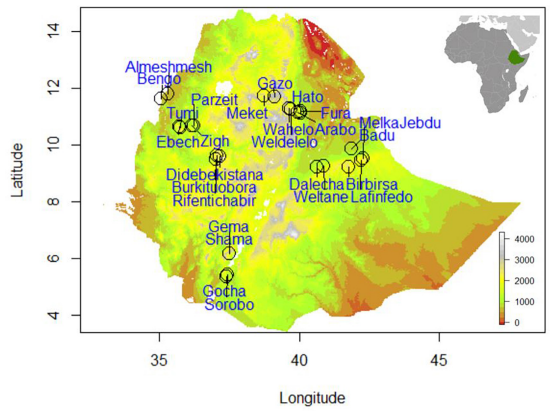

FIGURE 1 | Topographic map of Ethiopia depicting the 26 indigenous chicken sample populations.

Environmental Data

A total of 34 environmental variables were selected for their potential effects on chicken adaptive evolution. Data on these variables was extracted from online databases (**Supplementary Table 2**). The environmental data included bioclimatic (n = 24), vegetation (n = 2), and soil (n = 8) variables. Values for bioclimatic variables (temperature, precipitation, solar radiation, and water vapor pressure) in different seasons were obtained from WorldClim database[1] at a spatial resolution of 30 s (∼1 km^2; Fick and Hijmans, 2017) based on mean values of 30 years (1970–2000). Cropland extent at 30-m resolution was attained from Global Food Security Analysis-Support Data (Xiong et al., 2017). The SoilGrids system at 250 m resolution with standard numeric soil properties (organic carbon, bulk density, cation exchange capacity, pH, soil texture fractions at 15–30 cm depth was accessed from ISRIC database; Hengl et al., 2015, 2017). In addition to the 34 environmental variables elevation data was downloaded from DIVA-GIS[2] (Hijmans et al., 2001; Farr et al., 2007) at a spatial resolution of 30 s (∼1 km^2).

Species Distribution Models

Species distribution models (also called niche, envelope, or bioclimatic models) associate georeferenced observations of a biotic response variable – typically species occurrence or abundance – with multiple environmental predictors using a broad array of statistical learning methods to describe species' niches (Elith and Leathwick, 2009; Franklin, 2010; Elith and Franklin, 2013; MacKenzie et al., 2017). MaxEnt is a general-purpose machine learning algorithm developed to model species distributions from presence-only records (Phillips et al., 2006).

For every population, a single geographic coordinate was taken at the center of the village during sampling of chickens. Coordinates from nine additional grids (1.44 km^2), covering a total of 12.96 km^2, were then drawn around a recorded location and extracted using Google Earth Pro v 7.3.2 to ensure

[1]http://www.worldclim.org/; version 2

[2]http://www.diva-gis.org/gdata

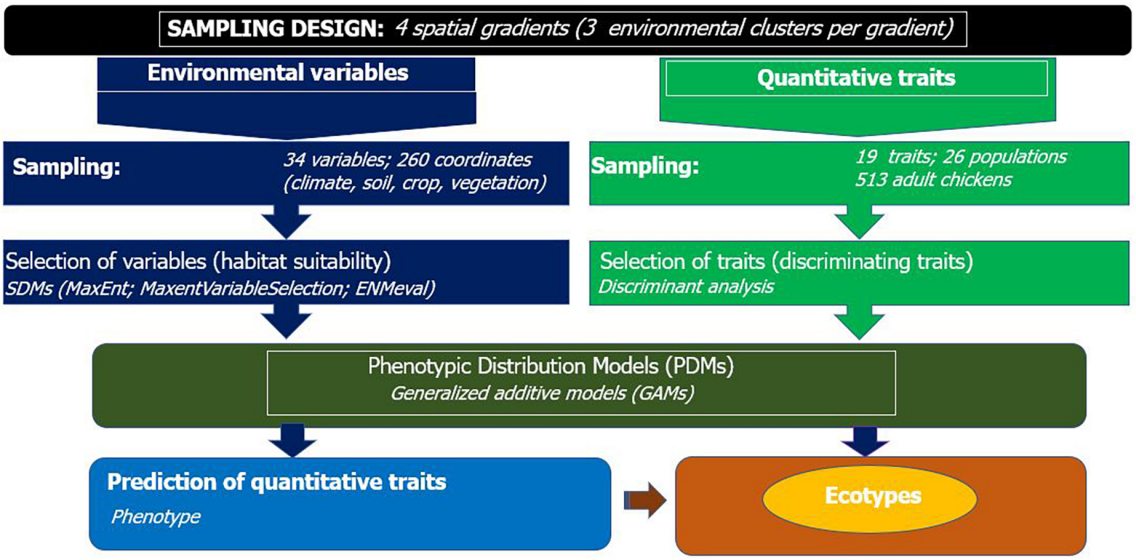

FIGURE 2 | Workflow for integrated species and phenotypic distribution modeling to detect population differentiation and define ecotypes of indigenous chickens.

high representation of environmental variability affecting the population (Gheyas et al., 2021). This way the total number of "presence" or "occurrence" points used in SDMs for the 26 sample populations comprised 260 coordinates. Different R software packages: "*sp*" (Pebesma et al., 2012), "*raster*" (Hijmans et al., 2015), "*rgdal*" (Bivand et al., 2015), "*maptools*" (Bivand et al., 2021), "*rgeos*" (Bivand et al., 2017), and "*dismo*" (Hijmans et al., 2017), were used to extract, read, and visualize geospatial data. Dimension and extent of the grids were corrected and homogenized for 1 km^2 based on the WGS84 geodetic reference system (Decker, 1986).

Selection of Environmental Variables

To constrain model complexity and increase the performance of SDMs, the highest contributing set of uncorrelated environmental variables were identified and Maxent's regularization multiplier was fine-tuned using the R package "*MaxentVariableSelection*" (Jueterbock et al., 2016). The predictive performance of the most important environmental variables was measured using test gain in MaxEnt v.3.4.1 (Phillips et al., 2006; Phillips and Dudík, 2008).

Configuration of Model Parameters

Species-specific tuning of model parameters can improve the performance of MaxEnt model compared to the default settings (Elith et al., 2011; Radosavljevic and Anderson, 2014). The large set of feature types was subsequently reduced to the optimal subset to improve model fit and the optimum regularization multiplier for model training was identified by the R package "*ENMeval*" (Muscarella et al., 2014) by using spatial blocks method (Radosavljevic and Anderson, 2014). Regularization refers to smoothing the model, making it more regular, to avoid fitting too complex a model. It is a common approach in model selection and penalizes coefficients (the betas) to values that allow both accurate prediction and generality (Tibshirani, 1996; Elith et al., 2011).

Species' responses to environmental covariates or independent variables (e.g., temperature, elevation) tends to be complex and usually requires fitting of non-linear functions (Austin, 2002). In machine learning algorithms this is achieved by applying transformations of the original covariates into features. MaxEnt currently has six feature classes: linear, product, quadratic, hinge, threshold, and categorical (Elith et al., 2011). We built models with regularization multiplier values ranging from 0.5 to 4.0 (increments of 0.5) and with six different feature combinations (H, LQH, HQP, HQC, LQHP, LQHPT; where L, linear; Q, quadratic; H, hinge; P, product; and T, threshold); this resulted in 48 individual model runs. The parameter configuration with the lowest delta AICc value was chosen to run the model (**Supplementary Table 3**). To reduce the influence of sampling bias, we included a bias file (Phillips et al., 2009) and preferentially sampled pseudo background points from areas near our presence points based on kernel density function (Venables and Ripley, 2002).

Tests of Niche Similarity

A niche is a description of the conditions in which a species maintains a viable population. Populations in a species that are adapted to a specific local habitat or niche show genetically induced phenotypic differences in response to environmental selective pressures and are regarded as "ecotypes" (Müntzing, 1971; Knüpffer et al., 2003). Niche similarity between one or more pairs of populations was measured according to Warren et al. (2008). Raster files (.ASCII) of predicted habitat suitability produced by MaxEnt in logistic output (no probability and complete probability of presence designated by 0 and 1, respectively) were used as inputs to perform correlation test by ENMTools (Warren et al., 2010). Correlation tests were

used to cluster sampling sites on the selected environmental variables and build dendrogram through hierarchical clustering with R package *cluster* (Maechler et al., 2013). The grouping of sampling locations into environmental niches was based on "Euclidean" distance. Different clustering methods (Ward's minimum variance method, complete linkage, average linkage, and single linkage) were compared. Visualization of the cluster memberships of locations of populations based on niche similarity, measured by correlations tests on the most important environmental variables, was accomplished using the R package *factoextra* (Kassambara and Mundt, 2017).

Quantitative Trait Data

A total of 19 phenotypic traits (**Supplementary Table 4**), selected for their potential role in adaptation in chicken based on available literature, were measured on 513 adult chickens (380 hens and 133 cocks) from the 26 nondescript indigenous chicken sample populations. We had three environmental clusters (lowland: 400–1,800; midaltitude: 1,800–2,400; and highland: 2,400–3,500 m.a.s.l.) stretching across each of the four elevational gradients in this study. A total of 12 environmental clusters from the four elevational gradients were included. Each environmental cluster is represented by two randomly selected chicken populations, except in two instances where we took three populations. A population refers to the total number of nondescript indigenous chickens available in an administrative village. Adult chickens were selected randomly for phenotyping through transect walk across villages. This method entailed walking along a defined path (transect) across a village and sampling one chicken from each farming household until a total of 15 hens and five cocks were measured.

The age of the chickens was estimated by interviewing owners to confirm that females were in their second clutch (7–8 months-of-age) and males were above 12 months-of-age. The researchers also visually appraised cocks (roosters) for presence of well-developed spurs. One chicken was sampled per household. Under rare circumstances ($n = 9$ households), two chickens were sampled per household when farmers proved their animals have no family relationship.

Live bodyweight of individuals was taken in the morning on fasting chickens. Accurate morphological measurements were made by using different tools (digital balance, measuring tapes, and image processing software) **Supplementary Table 4**. The pictures of individual chickens taken in a sheltered environment to achieve appropriate resolution were digitally analyzed using *ImageJ* (Rasband, 1997). To reduce systematic error, the same operator measured all chickens, which were held in the same position by a technician. A steel ruler was placed in the background of every picture as a distance reference.

Selection of Quantitative Traits

A multivariate test of differences between populations with stepwise selection (Klecka et al., 1980) was performed through linear discriminant function analysis (SAS, 2002) to identify the traits which were most useful in classifying populations. Principal component analysis (PCA) was run with R "*stats*" package on quantitative trait data to see how much percent of the variation is explained by the first nine principal components (PCs).

Clustering of Nondescript Chicken Populations Into Ecotypes

The 26 nondescript Ethiopian chicken populations sampled in this study are heterogenous in terms of qualitative traits (e.g., coat color, comb shape, and feather pattern) and quantitative traits. We used the most discriminant quantitative traits, which are most useful because of their variability, to group populations into ecotypes. We expect that populations of chickens within the same niche are affected by similar environmental variables and cluster into the same ecotype. The phenotypic values of these traits were analyzed by the average silhouette method to decide on the optimal number of clusters. The average silhouette method measures how well each experimental unit lies within its cluster and is less ambiguous than the elbow method to decide on the number of clusters (Rousseeuw, 1987; Kaufman and Rousseeuw, 2009).

Different hierarchical clustering methods (Ward's minimum variance method, complete linkage, average linkage, and single linkage) were compared via R packages "*cluster*" (Maechler et al., 2013) and "*factoextra*" (Kassambara and Mundt, 2017) to make a valid comparison of population memberships between dendrograms produced on similarity of phenotypes. We used the same approach for clustering of environmental and phenotypic data to avoid any possible bias associated with the use of different tools.

Phenotypic Distribution Models

While species can vary genetically and phenotypically across their range and populations can be locally adapted, SDMs assume that all populations respond homogenously to the range of environmental conditions experienced by the whole species (Bolnick et al., 2003; Atkins and Travis, 2010; Fitzpatrick and Keller, 2015; Hällfors et al., 2016). PDMs on the other hand, do capture the response of quantitative traits as a function of environmental conditions (Michel et al., 2017; Smith et al., 2017; Lozano-Jaramillo A. et al., 2019). We used PDMs to study variation within quantitative traits in response to the most important set of environmental variables identified by SDMs. The association of these environmental variables with habitat suitability were evaluated for their individual effect on each of the discriminating traits. The relationship between quantitative traits and environmental variables was expected to be non-linear (Zuur et al., 2007; Oddi et al., 2019). The assumptions of classical statistical approaches such as generalized linear models (GLM) are violated when responses are non-linear, variances change with predictors, or ecological processes operate at spatio-temporal scales (Zuur et al., 2009; Bolker et al., 2013).

Exploration of phenotypic and environmental data was initially carried out to understand their distribution, variance structure, and linearity or non-linearity of trend and to choose appropriate analytical methods. GAMs were selected because they are particularly useful for analyzing relationships explained by complicated shapes, such as hump-shaped curves (Crawley,

2012).The R package "*mgcv*" (Wood and Augustin, 2002) was used to fit GAMs (Hastie and Tibshirani, 1990). Model validation was made based on Akaike information criterion (AIC) values.

The response of each quantitative trait was predicted as a function of ecotype, niche, and the six SDM-selected environmental variables. The GAM included ecotypes and their respective niches as linear terms and the environmental covariates as smoothing parameters. The notation for the GAM smoothing in a Gaussian model is as follows (Hastie and Tibshirani, 1990; Wood and Augustin, 2002).

$$g\left(E\left(y_i\right)\right) = \alpha + \beta_j + \gamma_m + f_k(X_{ki}) \ldots,$$

Where $\left(E\left(y_i\right)\right)$ is one of n observations of the response trait, g is the Gaussian distributed exponential family with identity link function, α is the intercept, β_j is a linear parameter for ecotype (1,2,3), γ_m is a linear parameter for environmental niches (1,2,3), f_k are the smoothing terms based on non-parametric predictor covariates X_{ki} (the shape of the predictor functions which will be fully determined by the data structure).

Estimation of smoothing parameters effects (environmental variables) was done by restricted maximum likelihood as random effects (Wiley and Wiley, 2019) with Gaussian process smooth (bs = "gp") in the GAMs model (Wood, 2012).

Partial dependence plots (PDPs; Friedman, 2001) are the most popular approach for visualizing the effects of the predictor variables on the predicted outcome during supervised machine learning applications (Apley and Zhu, 2020). A PDP can show whether the relationship between the target and a feature is linear, monotonic, or more complex. PDPs exhibiting the effects of environmental factors with estimated p-value on a phenotype were produced by using the R package "*mgcViz*" (Fasiolo et al., 2020) at 95% confidence interval.

RESULTS

Environmental Variables Contribute to Habitat Suitability

Optimum Model Parameters

ENMeval identified HQP (Hinge, Quadratic, and Product) features with regularization-multiplier = 3.0 as the best parameter combination. This had the lowest deltaAICc value and was chosen to produce suitability maps by MaxEnt (**Figure 3A**). Compared to the default (**Figure 3B**), the model fit with the optimum parameters predicted larger areas as most suitable for poultry production (**Figure 3C**). The areas least populated by chickens include the extreme lowlands (below 400 m.a.s.l.), with prohibitively high temperature, high solar radiation, low precipitation, and high relative humidity; and the extreme highlands (above 3,400 m.a.s.l), with prohibitively low temperatures. The extreme highlands are frosty and hence not habitable both by livestock and humans. Ethiopian lowland pastoral areas are affected by recurrent drought and have generally sparse livestock population (Tilahun and Schmidt, 2012). Agreement between the results of the present study and the census report (CSA, 2017) and other literature indicating the distribution of livestock (Tilahun and Schmidt, 2012) confirm that those areas in the country which are shown as least suitable in the habitat suitability maps produced by SDMs are indeed unsuitable for the study species. Sedentary systems in central Ethiopia have conducive environmental conditions for chickens while pastoral systems (hot, dry areas, with strong solar radiation) along the borders of the country do not (Getahun, 1978; Bayou and Assefa, 1989; CSA, 2017; Mirkena et al., 2018; Gebrechorkos et al., 2019). The choice of livestock species to rear is also culturally embedded over generations.

Most Contributing Environmental Variables

Species distribution models identified the most important environmental variables associated with distribution of chickens (**Figure 4**). Correlated variables ($|r| > 0.6$) and those with a relative contribution score below 4% were removed to restrict multicollinearity driven effects in projecting species ranges (Dormann et al., 2013; Brun et al., 2019). Out of 34 environmental variables, nine were retained as most important in determining habitat suitability and can be regarded as potential drivers of local adaptation in Ethiopian indigenous chickens. The first five variables with the highest contribution included soil clay content, precipitation of the warmest quarter, precipitation of the coldest quarter, and temperature seasonality.

Jackknife test was run to compare the relative importance of the nine selected environmental variables (**Figure 5**). The test showed that precipitation of the coldest quarter and water vapor pressure in May have the highest gain when used in isolation, and therefore are the most useful variables for predicting the distribution of the species on occurrence data. On the other hand, the environmental variable that decreases gain the most when omitted is solar radiation in May, meaning it has the most important information that is not present in other variables.

Distinct Niches Are Associated With Distinct Ecotypes

Populations of animals adapted to a specific environment or niche are regarded as ecotypes. Clustering of sample chicken populations into phenotypically homogenous groups and an overlap of the clustered populations with niche classification based on their respective environments was used as a basis to define ecotypes. The number of chicken ecotypes was determined through Silhouette method using phenotypic data (**Figure 6A**). The optimal cluster in the present study, the one that maximized the average silhouette from a range of possible k values, was $k = 3$. The same clustering method (Ward, 1963) was used to make a valid comparison of population memberships between dendrograms produced on similarity of niches (**Figure 6B**) and on similarity of phenotypes (**Figure 6C**).

Populations were clustered into three environmental niches based on correlation test (**Figure 6B**). Ward's method had the strongest clustering structure for clustering on niche overlap (Ward = 0.89). The agglomerative coefficients for the other approaches (complete linkage = 0.78; average linkage (UPGMA) = 0.68; and single linkage = 0.36) was lower. At a cutoff value of 5.0, reading the plot from left to right, niche-I comprised 11 sampling locations, while niche-II and niche-III comprised six and nine locations, respectively.

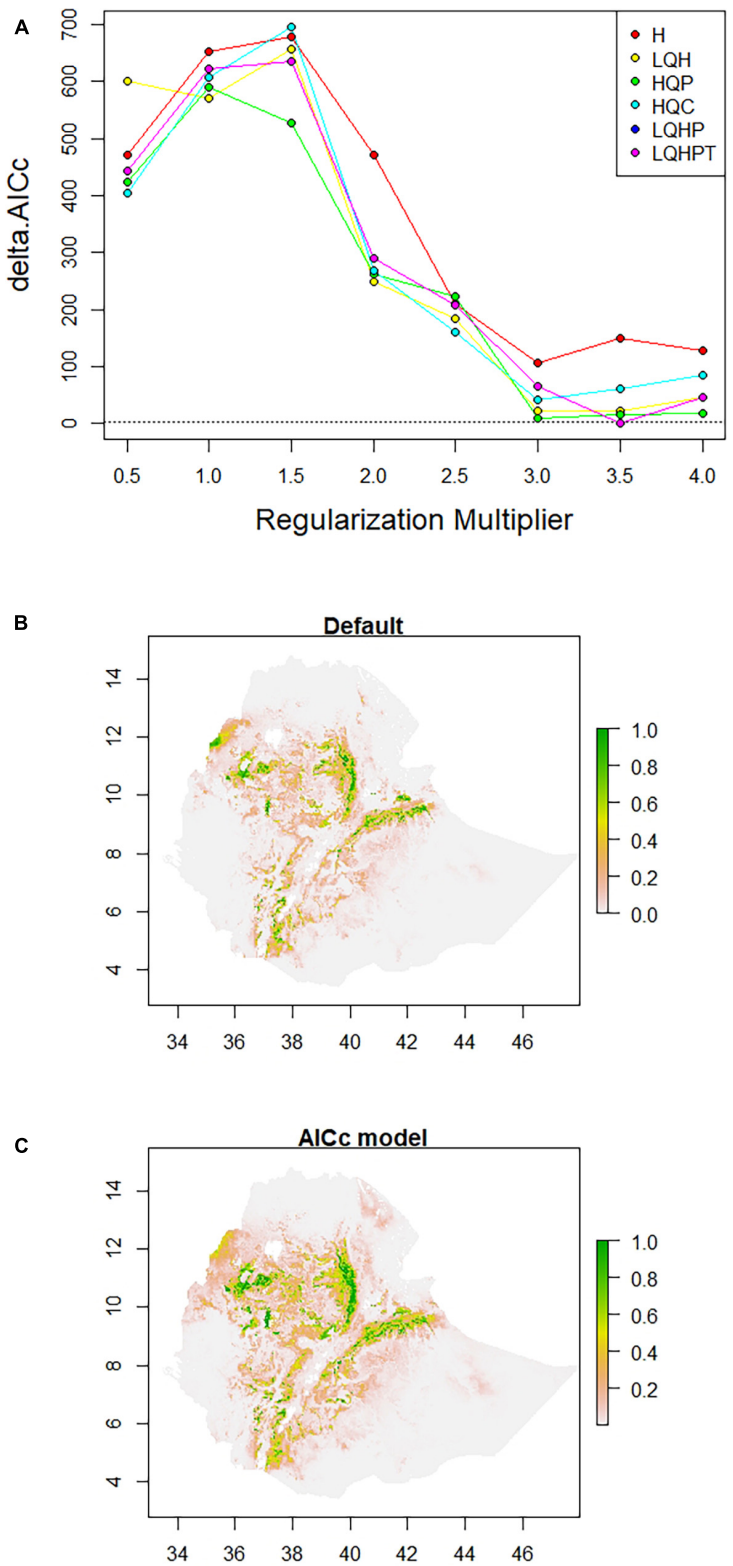

FIGURE 3 | Model configuration and habitat suitability maps for Ethiopian indigenous chicken populations. **(A)** AICc values for analyzed feature combinations using different regularization-multipliers ranging from 0.5 to 4.0. Feature combinations include one or more of the following types: L, linear; Q, quadratic; H, hinge; P, product; and T, threshold. **(B)** Map produced using default settings of MaxEnt. **(C)** Map produced using optimum parameters (HQP features with regularization-multiplier = 3.0) identified by ENMeval.

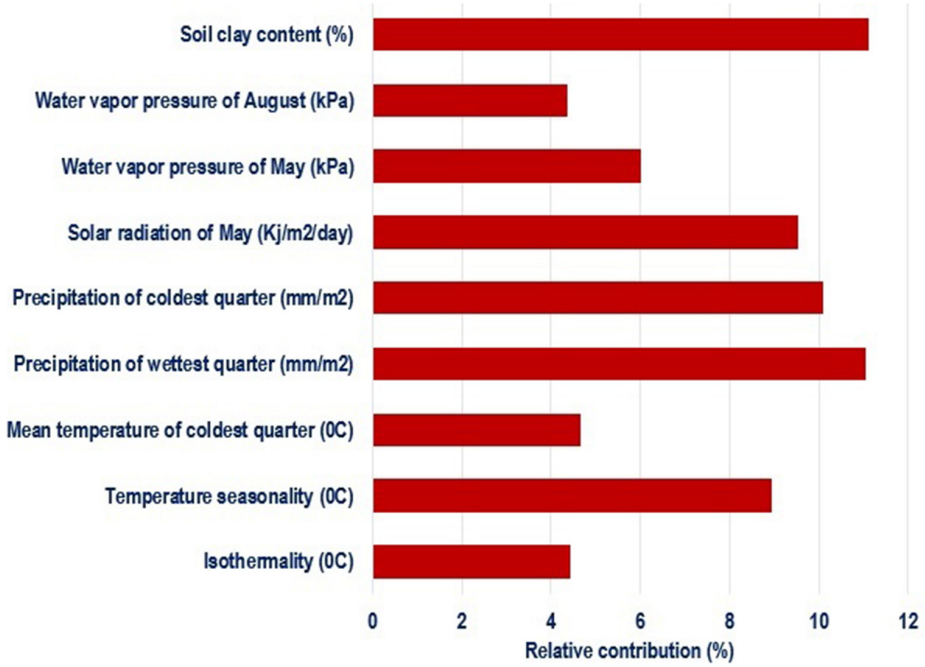

FIGURE 4 | Environmental variables of importance and their percent contribution predicted by MaxentVariableSelection.

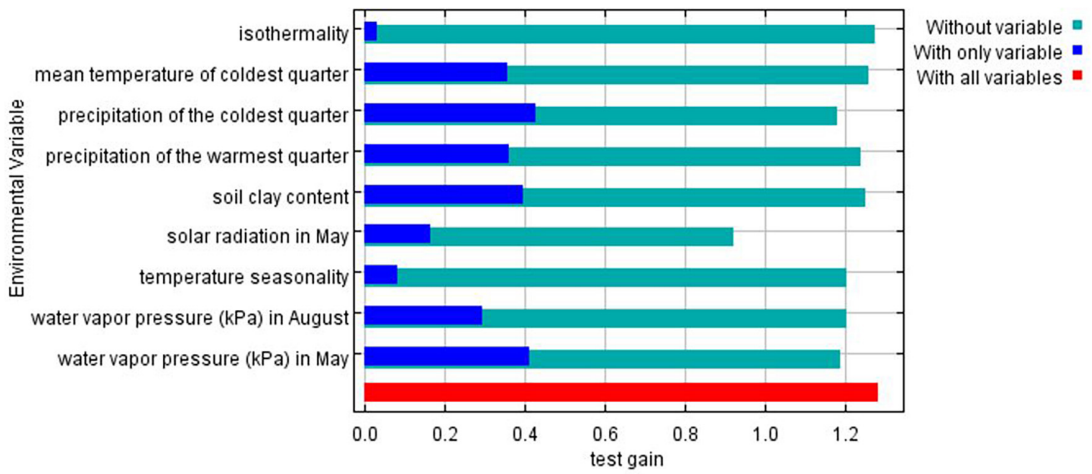

FIGURE 5 | Gains of the variables in the Maxent model (Jackknife test) for Ethiopian indigenous chickens. Turquoise bars: model gain without corresponding variables; blue bars: model gain with only the corresponding variables; red bars: total gain using all the variables.

Variation in Quantitative Traits

Before classifying the 26 sample chicken populations into ecotypes through hierarchical clustering based on similarity for quantitative traits, we reduced the number of traits through discriminant analysis (**Table 1**). Out of 19 quantitative traits (**Supplementary Table 4**), eight (BL, WS, CL, CW, BW, EW, WW, and KL) had the highest discriminant function because of their high variation between populations. Except wattle width ($p < 0.05$), the remaining seven of these eight discriminant traits showed highly significant phenotypic variation ($p < 0.0001$ to $p < 0.01$) between female sample chicken populations. The GLM analysis combining data from both sexes revealed all the discriminating quantitative traits varied significantly between sexes ($p < 0.0001$) except for beak length ($p = 0.1738$). The partial r-square indicates body length (BL) had the highest discriminatory effect out of all traits retained in the models in both sexes. Only two quantitative traits (BL and BW) were found useful ($p < 0.0001$) for classifying male sample chicken populations. This might be related with their lower sample size or a different structure of morphological variation among male sample populations compared to females.

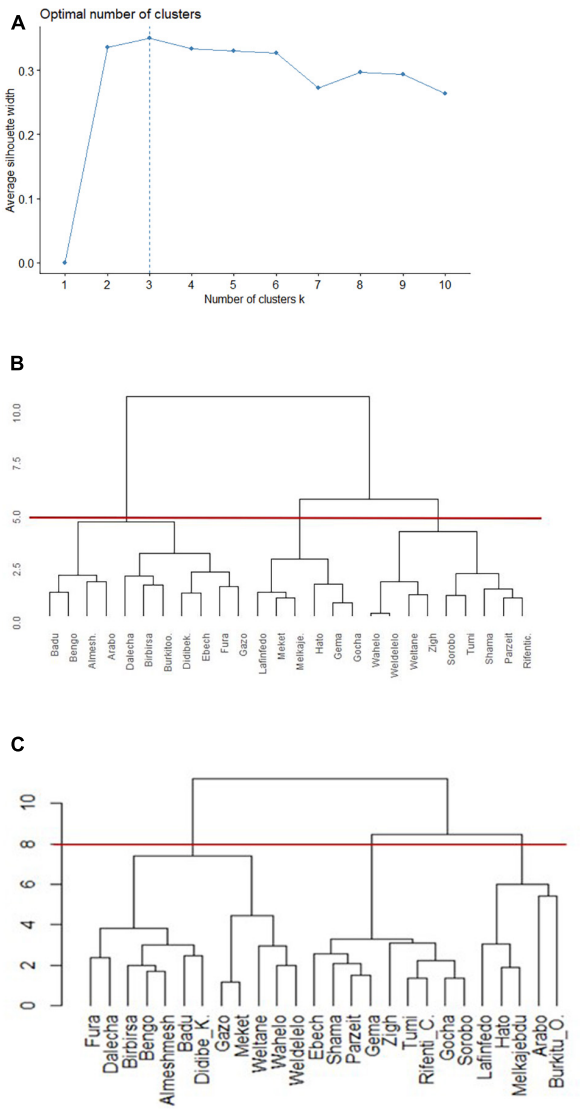

FIGURE 6 | Dendrograms of clusters to group 26 Ethiopian indigenous chicken populations (hierarchical agglomerative clustering, Ward's minimum variance method). **(A)** Plot of statistics for deciding appropriate number of clusters based on phenotype. **(B)** Dendrogram based on niche overlap statistic (*I*) between suitability maps. The red line at a cutoff value of 5.0 produces three niches. **(C)** Dendrogram based on the most discriminating phenotypes. The red line at a cutoff value of 8.0 produces three distinct ecotypes.

A subset of quantitative traits that best revealed the differences among chicken populations (**Table 1**) were then used for clustering. Ward's hierarchical clustering rendered the highest agglomerative coefficient (Ward = 0.81) for clustering of populations on phenotypic similarity compared with the other approaches [complete linkage = 0.71; average linkage (UPGMA) = 0.58; and single linkage = 0.49; **Figure 6C**]. The cutoff value at 8, indicated by horizontal line, resulted in three clusters. A PCA on quantitative trait data showed that the first three PCs explain 75.7% of the phenotypic variation among

TABLE 1 | Stepwise selection summary indicating most discriminating traits for adult male and female Ethiopian indigenous chicken sample populations.

Sex	Quantitative trait	Partial R-sq.	*F* value	Pr > F
Hens	BL	0.4761	13.51	<0.0001
	WS	0.2934	6.15	<0.0001
	CL	0.2274	4.34	<0.0001
	CW	0.1766	3.15	<0.0001
	BW	0.1741	3.08	<0.0002
	EW	0.1677	2.93	<0.0003
	WW	0.1184	1.63	<0.0214
	KL	0.1534	1.93	<0.0014
Cocks	BL	0.7756	14.52	<0.0001
	BW	0.4856	3.9	<0.0001

BL = body length (mm); WS = wingspan (mm); CL = comb length (mm); BW = body weight(g); EW = earlobe width (mm); WW = wattle width (mm); and KL = beak length (mm).

TABLE 2 | Ecotype of Ethiopian indigenous chicken populations defined on phenotype and their respective niches as identified by species distribution models.

Ecotype	Populations	Distributed within the same niche	Distributed outside the niche
I	Fura, Dalecha, Birbirsa, Bengo, Almeshmesh, Badu, Didibe Kistana, Gazo, Meket, Weltane, Wahelo, Weledelelo	Fura, Dalecha, Birbirsa, Bengo, Almeshmesh, Badu, Didibe Kistana, Gazo	Meket, Weltane, Wahelo, Weldelelo
II	Lafinfedo, Hato, Melkajebdu, Arabo, Burkitu Obora	Lafinfedo, Hato, Melkajebdu,	Arabo, Burkitu Obora
III	Ebech, Shama, Parzeit, Gema, Zigh, Tumi, Rifenti Chabir, Sorobo, Gocha,	Shama, Parzeit, Zigh, Tumi, Rifenti Chabir, Sorobo	Ebech, Gema, Gocha

populations (PC1 = 43.1%, PC2 = 19.5%, and PC3 = 13.2%) supporting our grouping of chicken populations into three ecotypes (**Supplementary Table 5**).

A summary of cluster analyses (**Table 2**) shows that most of the populations of a specific ecotype are distributed within the same niche while only a few of them distributed elsewhere. Eight out of 12 populations from ecotype-I, three out of five populations from ecotype-II, and six out of nine populations from ecotype-III were correctly classified into their respective niches.

Matching between chicken ecotypes and different environmental classification methods was performed to establish a logical association between phenotypic distinctiveness and environmental selective pressures (**Table 3**). The environmental classification methods included SDMs, conventional (Dove, 1890), Official (MoA, 1998), and gradient-based agroecological classifications. The highest level of correct classification was performed by SDMs (64.5%), followed by environmental gradient (elevational cline) classification (57.3%). The higher correct classification level obtained by the SDM approach, suggests the potential influence of the selected environmental variables (n = 9) on shaping adaptive variation among Ethiopian indigenous chicken ecotypes.

TABLE 3 | Comparison of methods to classify environments of Ethiopian indigenous chicken ecotypes (n = 3).

Classification method	Criteria for classification	No. of classes	No. of populations correctly classified (%)			Total no. of populations correctly classified (%)
			Ecotype-I	Ecotype-II	Ecotype-III	
SDM	Niche similarity	3	8 (66.7)	3 (60.0)	6 (66.7)	17 (64.5)
*Conventional AEs	Climatic classes (altitude, temperature, precipitation)	3	6 (66.7)	3 (33.3)	4 (50.0)	13 (50.0)
§ Official AEs	Temperature, soil type, plant growing period/moisture condition, land use	13	3 (33.3)	3 (33.3)	2 (25.0)	8 (30.7)
Environmental gradient	Elevational clines in distinct geographies	4	5 (55.5)	6 (66.6)	4 (50.0)	15 (57.3)

Conventional agroecological classes (AEs) comprise three groups measured in m.a.s.l.: I = lowlands (400–1,800); II = 1,800–2,400; and III = 2,400–3,500 (Dove, 1890).
§ Official AEs represent standard agroecologies of Ethiopia (MoA, 1998).

Environmental Variables Contribute to Phenotypic Differentiation

Having noticed that populations have differentiated distinctly in specific environments, we focused on predicting phenotypic values of ecotypes for the most discriminant quantitative traits within their respective niches under the influence of the selected environmental variables. Prediction of quantitative traits with GAMs in each of the three Ethiopian indigenous chicken ecotypes is presented in **Table 4**. Significant p-values were obtained for all the nine SDM identified environmental variables except for soil clay content. Five environmental variables (Bio18, Bio19, WVPM, and WVPA) had significant effect on differentiation of multiple traits. The traits selected by discriminant function for their usefulness in classification of populations into ecotypes had also the highest model fit (R-square adjusted values) explaining their importance in studying the influence of environmental variables on adaptive phenotypic variation.

Ethiopian indigenous chicken ecotypes identified by SDMs showed significant quantitative trait variation (**Table 5**). Populations in ecotype-I had the smallest measurement for all traits while ecotype-II had the largest measurements for most traits. It is not possible to tell from the present results alone whether the performance exhibited by ecotypes is primarily attributable to their niche or their genetic background.

Habitat suitability maps for Ethiopian indigenous chicken ecotypes (**Figure 7**) illustrate ideal environmental conditions that vary spatially between ecotypes. Chickens of ecotype-I (**Figure 7A**) are mainly distributed in central and northwest Ethiopia, ecotype-II (**Figure 7B**) are distributed in the west and southwest, while ecotype-III (**Figure 7C**) are distributed in eastern and northeastern Ethiopia. Areas of the country characterized by adverse environmental conditions due to their extreme temperature, high solar radiation, and low precipitation are shown as least suitable. This result conforms to the available census data which shows regions in the country with more friendly climate to chickens are more populated by the species (CSA, 2017).

The response of adult live body weight (BW) and BL in female indigenous chickens to some of the significant environmental variables ($p < 0.001$) are presented in **Figures 8**, **9**. The relationship between BW and solar radiation, and BW and water vapor pressure in May (kPa) is linear while its relationship with isothermality is non-linear (**Figure 8**). Isothermality quantifies how large the day-to-night temperatures oscillate relative to the annual oscillations. An isothermal value of 100 indicates the diurnal temperature range is equivalent to the annual temperature range, while anything less than 100 indicates a smaller level of temperature variability within an average month relative to the year (O'Donnell and Ignizio, 2012). Our results suggest that BW is less influenced by smaller temperature fluctuations within a month relative to the year. On the other hand, solar radiation above 18,000 kJ/m^2/day is stressful and has negative and linear effect on female BW. The relationship between bodyweight and mean temperature of the coldest quarter is more complex, showing that the mean temperatures during the coldest 3 months of the year is less useful to examine how this variable affects adult live BW.

A non-linear relationship is noted between BL and water vapor pressure in August (kPa), and between BL and precipitation of the coldest quarter (mm/m^2). Temperature seasonality had a negative and linear relationship with this trait. Temperature seasonality is a measure of temperature change over the course of the year. Our result indicates that higher standard deviation in the mean monthly temperature is associated with smaller BL, a trait which is strongly correlated with live BW. Precipitation of the coldest quarter is a quarterly index which approximates the total precipitation that prevails during the 3 months of the year. Accelerated mean change in BL, in the context of the model was seen up to 700 mm/m^2 of precipitation in the coldest quarter. Precipitation above this threshold might be related with less availability of scavenging feed resources and more prevalence of diseases and parasites, having adverse effects on this trait. Biologically speaking, water vapor pressure is a function of temperature and pressure. Negative relation is noted between this environmental variable and BL, probably because of the stressful situation (e.g., lower feed intake) it creates on the animals. A non-linear reduction was observed in BL for higher soil clay content above 20% which may have a relationship with the type of vegetation and land use pattern in those areas (**Figure 9**).

TABLE 4 | Prediction of quantitative traits with Generalized Additive Models (GAMs) in Ethiopian indigenous chicken ecotypes (n = 3).

Trait	[1]Fixed effects/linear term		Random effects/ smoothing term									Model fit			
			Bio3	Bio4	Bio11	Bio18	Bio19	SRM	WVPM	WVPA	SCC	df	AIC	R-sq.(adj)	Deviance explained (%)
BL	Ecotype	Niches	*			*	***	*	***	***		14.9	1538.7	0.65	66.7
WS	Ecotype	Niches			***	***	*	*	***			13.6	1643.65	0.55	56.5
CL	Ecotype	Niches					*					8.0	2583.4	0.21	22.3
CW	Ecotype	Niches							***	***		9.3	2140.3	0.10	11.9
BW	Ecotype	Niches	*			**	*	***	***			12.5	-121.2	0.45	46.5
EW	Ecotype	Niches		***	.				***	***		16.3	1656.8	0.25	28.2
WW	Ecotype	Niches	*			**				**		10.0	2034.7	0.12	14.3
KL	Ecotype	Niches				*	**	.		**		11.9	1724.0	0.05	7.6

Akaike information criterion (AIC) is a goodness of fit measure (likelihood or log-likelihood) that penalizes for complexity number of parameters or degree of freedoms). BL = body length; WS = wingspan; CL = comb length CW = comb width; BW = body weight; EW = earlobe width; WW = wattle width; and KL = beak length. Bio3, Isothermality; Bio4, Temperature seasonality; Bio11, Mean temperature of coldest quarter; Bio18, Precipitation of warmest quarter; Bio19, Precipitation of coldest quarter; SRM, solar radiation of May; WVPM, water vapor pressure of May; WVPA, water vapor pressure of August; SCC, soil clay content.
[1]Linear effect of ecotype is significant for all discriminating phenotypes.
*Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 ".".*

TABLE 5 | Quantitative trait variation in Least Square Mean (Standard Error) among adult female Ethiopian indigenous chickens of different ecotypes defined by integrating SDMs with PDMs.

Ecotype*	LSMean (S.E.)									
	Hens (n = 380)								Cocks (n = 133)	
	BW	BL	WS	CL	CW	EW	WW	KL	BW	BL
I	1.01 (0.01)^{b}	35.46 (0.22)^{c}	38.78 (0.22)^{b}	21.3 (0.62)^{c}	7.95 (0.36)^{b}	8.96 (0.2)^{b}	16.89 (0.31)^{c}	16.32 (0.20)	1.31 (0.05)^{b}	38.89 (0.52)^{b}
II	1.31 (0.02)^{a}	39.13 (0.23)^{a}	41.88 (0.24)^{a}	30.21 (0.65)^{a}	10.18 (0.37)^{a}	10.9 (0.2)^{a}	19.4 (0.32)^{a}	16.46 (0.21)	1.78 (0.05)^{a}	44.34 (0.54)^{a}
III	1.28 (0.02)^{a}	38.48 (0.24)^{b}	42.03 (0.24)^{a}	25.22 (0.66)^{b}	8.82 (0.38)^{b}	10.4 (0.20)^{a}	18.48 (0.33)^{b}	16.65 (0.22)	1.82 (0.05)^{a}	44.49 (0.57)^{a}

*BW = body weight (g); BL = body length(mm); WS = wingspan(mm); CL = comb length(mm); CW = comb width(mm); EW = earlobe width(mm); WW = wattle width(mm); and KL = beak length(mm). ^{a,b,c}Means with different superscripts within the same column are significantly ($p < 0.05$) different. *Ecotypes were highly significant from each other ($p < 0.0001$) for all phenotypic measurements except for KL in hens ($p = 0.5393$).*

DISCUSSION

Sustainable livestock production particularly in the tropics requires adaptive genotypes which can withstand the undesirable effects of climate change and produce optimally (Fleming et al., 2017; Bettridge et al., 2018). Ecological variables vary in terms of their influences on organisms as inducers of local adaptation. Knowledge of ecological factors responsible for adaptive variation should be the first step to design selective breeding programs on indigenous livestock, plan crossbreeding with improved genotypes, or introduce new genotypes from a different environment (Fleming et al., 2017; Bettridge et al., 2018; Birhanu et al., 2021; Gheyas et al., 2021).

We have applied distribution models to identify the most important environmental factors associated with habitat suitability and phenotypic differentiation in indigenous populations of chickens. Previous studies indicated that populations differentiate phenotypically and genetically in response to the environment (Schmid and Guillaume, 2017; Smith et al., 2017). A tight relation is expected between environmental elements (e.g., precipitation, temperature, radiation, and elevation) and livestock population dynamics (Alemayehu and Getu, 2016; Getachew et al., 2016) in Ethiopia.

Precipitation of the warmest and the coldest quarters, soil clay content, temperature seasonality, solar radiation, water vapor pressure, and mean temperature of the coldest quarter, were identified by SDMs as the most important variables associated with habitat suitability in Ethiopian indigenous chickens. Precipitation is associated with types and amounts of crops cultivated; availability of scavenging feed resources and edible soil fauna; disease prevalence, and predation. Precipitation and temperature were also identified as most important contributors to local adaptation in African chickens (Fleming et al., 2017; Bettridge et al., 2018; Gheyas et al., 2021). The BW of Horro, Koekoek, Sasso, and SRIR chickens distributed to different regions of Ethiopia was best predicted by variables associated with temperature and precipitation (Lozano-Jaramillo A. et al., 2019). Clay content is a proxy for soil fertility and has impacts on feed availability for poultry. Through their physical and chemical properties, clay minerals can be expected to have more nutrient reserves in the tropics (Landon, 2014; Kome et al., 2019).

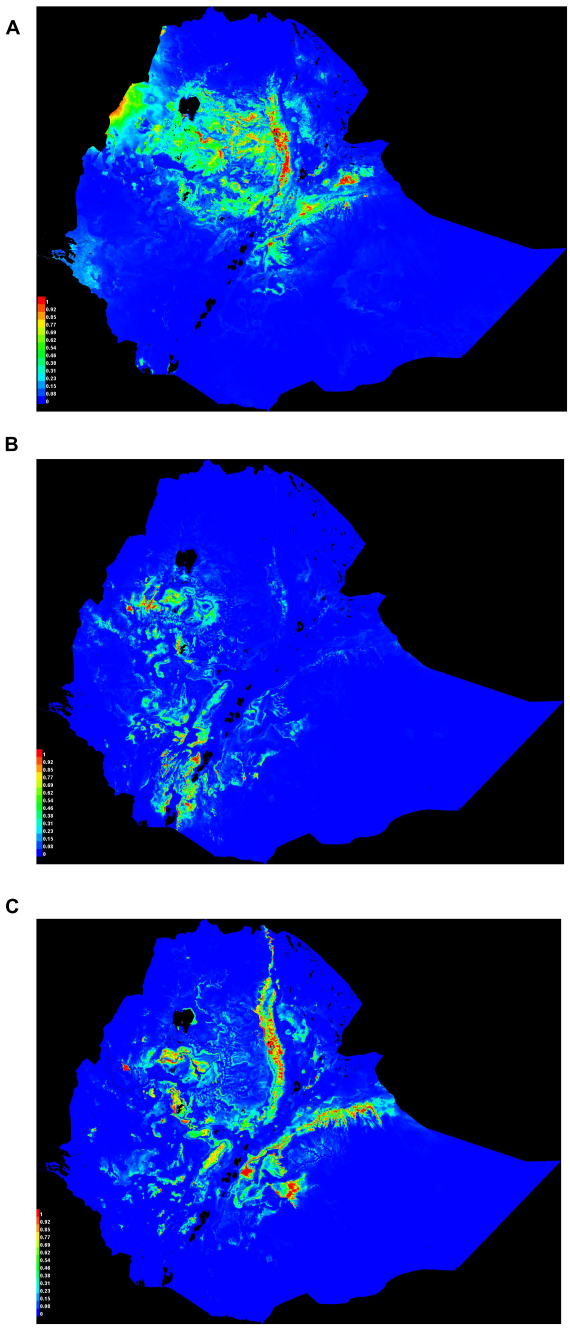

FIGURE 7 | Suitability maps of three Ethiopian chicken ecotypes. Colors toward red spectrum indicate more suitable conditions. **(A)** Ecotype-I. **(B)** Ecotype-II. **(C)** Ecotype-III.

All the nine environmental variables selected for their association with habitat suitability by SDMs had significant effects on differentiation of quantitative traits. The influence of isothermality (Bio3), temperature (Bio4 and Bio11), precipitation (Bio19), solar radiation, and water vapor pressure on trait differentiation may be related with adaptive physiology of chickens, in terms of their biological response to extremes in relative humidity and heat stress. Lozano-Jaramillo A. et al. (2019) and Alemu et al. (2021) have also observed effects of precipitation and temperature on improved chicken breeds introduced to smallholder farmers in Ethiopia.

We classified the Ethiopian indigenous chicken sample populations into three ecotypes and compared their respective performances. Homogenous clusters for measured quantitative traits and their overlaps with distinct niches were used to define ecotypes. Unlike previous efforts made to group Ethiopian indigenous chicken populations on qualitative phenotypes such as comb shape, and feather color (Melesse and Negesse, 2011; FAO, 2012; Negassa et al., 2014; Getachew et al., 2016; Overdijk, 2019), the definition of ecotypes in the present study integrated phenotypic and environmental information. This process included identification of the most contributing environmental variables for habitat suitability, grouping of sample locations into specific niches based on their environmental similarity, and selection of the most useful quantitative traits for population classification purposes.

Phenotypic distribution models, in a form of non-linear GAMs were demonstrated as an innovative approach to integrate environmental and phenotypic information and study their relationships. GAMs relax the assumptions of linear models such as GLMs and achieve acceptable goodness of fit. Such a non-linear data structure would have been missed otherwise (Wiley and Wiley, 2019). PDMs were used to complement predictions of SDMs in studying responses of prairie grass to climate change (Smith et al., 2017).

The use of SDMs is unchartered territory for livestock scientists. Limitations are expected in their use on domesticated species because of human interference influencing the natural distribution of the study populations. While existing SDMs alone do not seem appropriate to study breeds recently introduced into a new environment artificially, the models are applicable for those studying local adaptation among indigenous populations of livestock which have lived in their environment for hundreds of generations or more and have experienced significant selective pressures. Predictive ability of machine learning algorithms on domesticated species can be improved if they incorporate more data in addition to presence–absence information and harness sophisticated algorithms. Boosted regression trees and random forests as well as generalized additive and linear mixed models have improved prediction of SDMs in other species (Shirk et al., 2018).

Several evolutionary processes shape genetic and phenotypic differentiation, including the joint effects of environment (phenotypic plasticity), gene flow, and natural selection (Schmid and Guillaume, 2017). It is not clear from the present study whether the phenotypic differentiation that ensued between indigenous chicken ecotypes is the result of differentiation in allele frequencies. An integrated framework including environmental, phenotypic, and genomic analysis is needed to unravel the genetic basis of phenotypic differentiation among populations and ecotypes of these chickens. If the phenotype is directly influenced by the environment, genetic, and phenotypic differentiations can be decoupled (Crispo, 2008; Schmid and Guillaume, 2017). Improvements in predictive ability of models

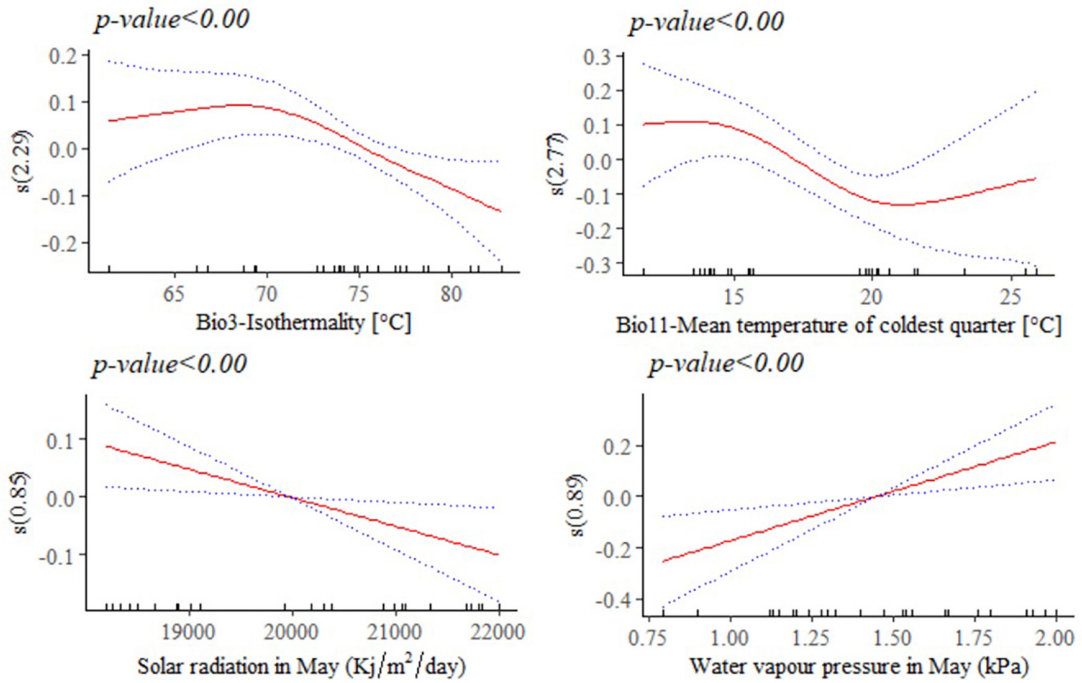

FIGURE 8 | Generalized additive model partial dependence plots for live body weight (kg) in female indigenous chickens. Each plot shows a covariate and the partial dependence of adult live body weight in the context of the model. The *y* axis shows the mean of observed change in live body weight and the *x* axis the covariate interval. The blue line represents the 95% confidence interval; Red line, mean of observed change in live body weight; s, smoothed variable; and *( )*, effective degrees of freedom.

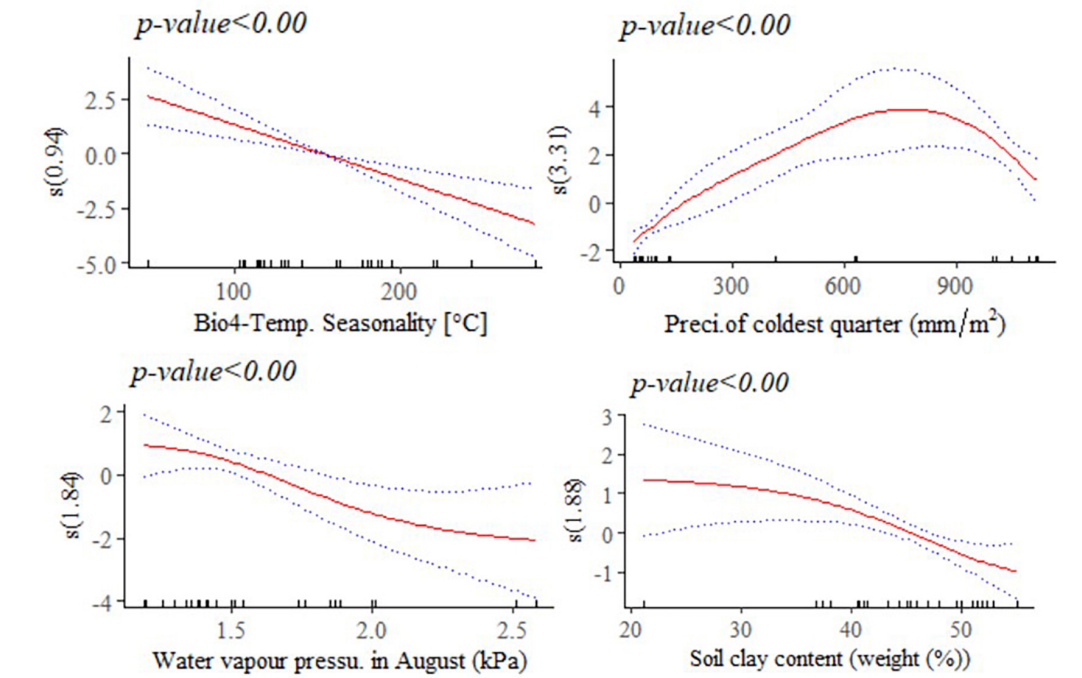

FIGURE 9 | Generalized additive model partial dependence plots for body length (mm) in female indigenous chickens. Each plot shows a covariate and the partial dependence of adult live body weight in the context of the model. The *y* axis shows the mean of observed change in live body length and the *x* axis the covariate interval. The blue line represents the 95% confidence interval; Red line, mean of observed change in live body weight; s, smoothed variable; and *( )*, effective degrees of freedom.

is also achieved when SDMs are used along with phenotypic and genomic information in landscape genetics and genomics studies (Joost et al., 2007; Gotelli and Stanton-Geddes, 2015; Razgour, 2015).

The present study demonstrated how SDM-identified environmental information can be integrated with PDMs to define ecotypes, predict quantitative traits, and understand the ecological roots of phenotypic differentiation. Considering the environmental influences of economically important quantitative traits, such as live BW, improves the estimation of breeding values and assists in the development of improved breeds suited to smallholder farmers. Differences in performance among ecotypes in the different niches will also mean evaluations of performance and yield stability across environments are pertinent in breeding and development programs designed for low- and medium-input poultry production systems of the tropics. Prospects of further use for SDMs and PDMs in livestock include definition of agroecologies, estimation of genotype by agroecology interactions, multi-environment performance evaluations, and prediction of performance under present and future production scenarios (e.g., climate change).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee (IACUC), International Livestock Research Institute (ILRI).

AUTHOR CONTRIBUTIONS

FGK, JB, HK, OH, and TD conceived the ideas and designed the study. FGK selected sample populations, collected metadata, performed the phenotyping of chickens and data analysis, and wrote the manuscript. JB and SA supported in the GAMs analysis. HK and JB provided useful comments and suggestions and helped improve the data analysis and draft the manuscript. TD, HK, OH, and JB secured funding. All authors critically revised the manuscript and gave final approval for publication.

ACKNOWLEDGMENTS

We thank the African Chicken Genetic Gains (ACGG) program of the International Livestock Research Institute (ILRI), funded by the Bill and Melinda Gates Foundation (Grant Agreement OPP1112198); and Animal Breeding and Genomics of Wageningen University & Research (WUR) for sponsoring this research. ILRI and WUR have provided a Ph.D. scholarship to the FGK.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.723360/full#supplementary-material

Supplementary Table 1 | Geographic distribution of 26 Ethiopian indigenous chicken sample populations. *Traditional agroecological classes (AEs) comprise three groups measured in m.a.s.l.: I = lowlands (400–1,800); II = 1,800–2,400; and III = 2,400–3,500 (Dove, 1890). § Official AEs represent standard agroecologies of Ethiopia (MoA, 1998).

Supplementary Table 2 | Environmental (climatic, soil, vegetation type) variables obtained for the locations of indigenous Ethiopian chicken sample populations.

Supplementary Table 3 | ENMeval table results for all combinations of features and betamultipliers. *Feature classes: L, linear; Q, quadratic; H, hinge; P, product; and T, threshold. § RM, Regularization multiplier.

Supplementary Table 4 | Quantitative traits measured on individual hens and cocks (n = 513) in indigenous Ethiopian chicken sample populations.

Supplementary Table 5 | Principal Component Analysis (PCA) results for female quantitative traits.

REFERENCES

Alemayehu, K., and Getu, A. (2016). Impacts of climate variability on livestock population dynamics and breed distribution patterns in selected districts of Western Amhara, Ethiopia. *Anim. Genet. Resour.* 59, 113–121. doi: 10.1017/s2078633616000230

Alemu, S. W., Hanotte, O., Kebede, F. G., Esatu, W., Abegaz, S., Bruno, J. E., et al. (2021). Evaluation of live-body weight and the number of eggs produced for introduced and local chickens in Ethiopia. *Acta Agric. Scand. A Anim. Sci.* 70, 71–77. doi: 10.1080/09064702.2021.1891278

Apley, D. W., and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 82, 1059–1086. doi: 10.1111/rssb.12377

Atkins, K., and Travis, J. (2010). Local adaptation and the evolution of species' ranges under climate change. *J. Theor. Biol.* 266, 449–457. doi: 10.1016/j.jtbi.2010.07.014

Austin, M. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Modell.* 157, 101–118. doi: 10.1016/s0304-3800(02)00205-3

Bayou, T., and Assefa, A. (1989). Solar radiation maps fot Ethiopia. *Zede J.* 8, 7–15.

Bergmann, C. (1848). *Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse*. Göttingen: Vandenhoeck und Ruprecht.

Bettridge, J. M., Psifidi, A., Terfa, Z. G., Desta, T. T., Lozano-Jaramillo, M., Dessie, T., et al. (2018). The role of local adaptation in sustainable production of village chickens. *Nat. Sustain.* 1, 574–582. doi: 10.1038/s41893-018-0150-9

Billi, P. (2015). *Landscapes and Landforms of Ethiopia*. Dordrecht: Springer.

Birhanu, M. Y., Alemayehu, T., Bruno, J. E., Kebede, F. G., Sonaiya, E. B., Goromela, E. H., et al. (2021). Technical efficiency of traditional village chicken production in Africa: entry points for sustainable transformation and improved livelihood. *Sustainability* 13:8539. doi: 10.3390/su13158539

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., et al. (2015). *Package 'rgdal'. Bindings for the Geospatial Data Abstraction Library*. Available online at: https://cran.r-project.org/web/packages/rgdal/index.html (accessed October 15, 2017).

Bivand, R., Lewin-Koh, N., Pebesma, E., Archer, E., Baddeley, A., Bearman, N., et al. (2021). *Package 'maptools'*.

Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K. O., and Bivand, M. R. (2017). *Package 'rgeos'. The Comprehensive R Archive Network (CRAN)*.

Bolker, B. M., Gardner, B., Maunder, M., Berg, C. W., Brooks, M., Comita, L., et al. (2013). Strategies for fitting nonlinear ecological models in R, AD M odel B uilder, and BUGS. *Methods Ecol. Evol.* 4, 501–512. doi: 10.1111/2041-210x.12044

Bolnick, D. I., Svanbäck, R., Fordyce, J. A., Yang, L. H., Davis, J. M., Hulsey, C. D., et al. (2003). The ecology of individuals: incidence and implications of individual specialization. *Am. Nat.* 161, 1–28. doi: 10.2307/3078879

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., et al. (2019). Model complexity affects species distribution projections under climate change. *J. Biogeogr.* 47, 130–142. doi: 10.1111/jbi.13734

Cain, A. J., and Sheppard, P. M. (1954). Natural selection in Cepaea. *Genetics* 39:89. doi: 10.1093/genetics/39.1.89

Clausen, J., Keck, D. D., and Hiesey, W. M. (1940). *Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants*. Washington, DC: Carnegie Institution of Washington.

Crawley, M. J. (2012). *The R book*. Hoboken, NJ: John Wiley & Sons.

Crispo, E. (2008). Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J. Evol. Biol.* 21, 1460–1469. doi: 10.1111/j.1420-9101.2008.01592.x

CSA (2017). *Livestock Characteristics, Agricultural Sample Survey*. Addis Ababa: Central Staistical Authority.

Dana, N., Van der Waaij, L. H., Dessie, T., and van Arendonk, J. A. (2010). Production objectives and trait preferences of village poultry producers of Ethiopia: implications for designing breeding schemes utilizing indigenous chicken genetic resources. *Trop. Anim. Health Prod.* 42, 1519–1529. doi: 10.1007/s11250-010-9602-6

Darwin, C., Wallace, A. R., Lyell, S. C., and Hooker, J. D. (1858). *On the Tendency of Species to form Varieties: And on the Perpetuation of Varieties and Species by Natural Means of Selection*. London: Linnean Society of London.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., et al. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22, 1383–1399. doi: 10.1111/mec.12182

Decker, B. L. (1986). "World geodetic system 1984," in *Proceedings of the 4th International Geodetic Symposium on Satellite Positioning*, Austin, TX.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Dove, K. (1890). *Kulturzonen von Nord-Abessinien*. Gotha: J. Perthes.

Elith, J., and Franklin, J. (2013). "Species distribution modeling," in *Encyclopedia of Biodiversity*, 2nd Edn, ed. S. M. Scheiner (Amsterdam: Elsevier Inc), 692–705.

Elith, J., and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Ethiopian Mapping Authority (1988). *National Atlas of Ethiopia*. Addis Abeba: Ethiopian Mapping Authority.

Etterson, J. R., and Shaw, R. G. (2001). Constraint to adaptive evolution in response to global warming. *Science* 294, 151–154. doi: 10.1126/science.1063656

FAO (2012). *Phenotypic Characterization of Animal Genetic Resources*. Rome: FAO.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Rev. Geophys.* 45, 1–33.

Fasiolo, M., Nedellec, R., Goude, Y., and Wood, S. N. (2020). Scalable visualization methods for modern generalized additive models. *J. comput. Graph. Stat.* 29, 78–86. doi: 10.1080/10618600.2019.1629942

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Fitzpatrick, M. C., and Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* 18, 1–16. doi: 10.1111/ele.12376

Fleming, D. S., Weigend, S., Simianer, H., Weigend, A., Rothschild, M., Schmidt, C., et al. (2017). Genomic comparison of indigenous African and Northern European chickens reveals putative mechanisms of stress tolerance related to environmental selection pressure. *G3* 7, 1525–1537. doi: 10.1534/g3.117.041228

Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.

Gebrechorkos, S. H., Hülsmann, S., and Bernhofer, C. (2019). Long-term trends in rainfall and temperature using high-resolution climate datasets in East Africa. *Sci. Rep.* 9:11376.

Getachew, F., Abegaz, S., Assefa, A., Misganaw, M., Emshaw, Y., Hailu, A., et al. (2016). Multivariate analyses of morphological traits in indigenous chicken populations of Metekel zone, Northwestern Ethiopia. *Anim. Genet. Resour.* 59, 15–25.

Getahun, A. (1978). Agricultural systems in Ethiopia. *Agric. Syst.* 3, 281–293. doi: 10.1016/0308-521x(78)90014-8

Gheyas, A. A., Vallejo-Trujillo, A., Kebede, A., Lozano-Jaramillo, M., Dessie, T., Smith, J., et al. (2021). Integrated environmental and genomic analysis reveals the drivers of local adaptation in African indigenous chickens. *Mol. Biol. Evol.* [E pub ahead of print]

Gotelli, N. J., and Stanton-Geddes, J. (2015). Climate change, genetic markers and species distribution modelling. *J. Biogeogr.* 42, 1577–1585. doi: 10.1111/jbi.12562

Hällfors, M. H., Liao, J., Dzurisin, J., Grundel, R., Hyvärinen, M., Towle, K., et al. (2016). Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecol. Appl.* 26, 1154–1169. doi: 10.1890/15-0926

Hampe, A. (2004). Bioclimate envelope models: what they detect and what they hide. *Glob. Ecol. Biogeogr.* 13, 469–471. doi: 10.1111/j.1466-822x.2004.00090.x

Hastie, T., and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman & Hall, 335.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12:e0169748. doi: 10.1371/journal.pone.0169748

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., et al. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10:e0125814. doi: 10.1371/journal.pone.0125814

Hijmans, R. J., Guarino, L., Cruz, M., and Rojas, E. (2001). Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genet. Resour. Newslett.* 127, 15–19.

Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., and Hijmans, M. R. J. (2017). Package 'dismo'. *Circles* 9, 1–68.

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., et al. (2015). *Package 'raster'. R package 734*.

Hurni, H. (1998). *Agroecological Belts of Ethiopia. Explanatory Notes on Three Maps at a Scale of 1*. Addis Ababa: Soil Conservation Research Program.

Joost, S., Bonin, A., Bruford, M. W., Despreìs, L., Conord, C., Erhardt, G., et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969. doi: 10.1111/j.1365-294x.2007.03442.x

Jueterbock, A., Smolina, I., Coyer, J. A., and Hoarau, G. (2016). The fate of the Arctic seaweed *Fucus distichus* under climate change: an ecological niche modeling approach. *Ecol. Evol.* 6, 1712–1724. doi: 10.1002/ece3.2001

Kassambara, A., and Mundt, F. (2017). *Package 'factoextra'. Extract and Visualize the Results of Multivariate Data Analyses 76*.

Kaufman, L., and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken,NJ: John Wiley & Sons.

Klecka, W. R., Iversen, G. R., and Klecka, W. R. (1980). *Discriminant Analysis*. Thousand Oaks, CA: Sage.

Knüpffer, H., Terentyeva, I., Hammer, K., and Kovaleva, O. (2003). "Ecogeographical Diversity–a Vavilovian approach," in *Diversity in Barley*, eds R. Invon Bothmer, T. Van Hintum, H. Knüpffer, and K. Sato (Amsterdam: Elsevier).

Kome, G. K., Enang, R. K., Tabi, F. O., and Yerima, B. P. K. (2019). Influence of clay minerals on some soil fertility attributes: a review. *Open J. Soil Sci.* 9, 155–188. doi: 10.4236/ojss.2019.99010

Landon, J. R. (2014). *Booker Tropical Soil Manual: A Handbook for Soil Survey and Agricultural Land Evaluation in the Tropics and Subtropics*. Milton Park: Routledge.

Langerhans, R. B. (2008). Predictability of phenotypic differentiation across flow regimes in fishes. *Integr. Comp. Biol.* 48, 750–768. doi: 10.1093/icb/icn092

Leinonen, T., Cano, J. M., Mäkinen, H., and Merilä, J. (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J. Evol. Biol.* 19, 1803–1812. doi: 10.1111/j.1420-9101.2006.01182.x

Lotterhos, K. E., and Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24, 1031–1046. doi: 10.1111/mec.13100

Lozano-Jaramillo, A., Alemu, S., Dessie, T., Komen, H., and Bastiaansen, J. W. M. (2019). Using phenotypic distribution models to predict livestock performance. *Sci. Rep.* 9:15371.

Lozano-Jaramillo, M., Bastiaansen, J. W. M., Dessie, T., and Komen, H. (2019). Use of geographic information system tools to predict animal breed suitability for different agro-ecological zones. *Animal* 13, 1536–1543. doi: 10.1017/s1751731118003002

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., and Hines, J. E. (2017). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Amsterdam: Elsevier.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., and Studer, M. (2013). *Package 'cluster'. Dosegljivo na*.

Maloney, K. O., Schmid, M., and Weller, D. E. (2012). Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. *Methods Ecol. Evol.* 3, 116–128. doi: 10.1111/j.2041-210x.2011.00124.x

Mayr, E. (1942). *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York, NY: Columbia University Press, 334.

McConnell, R. (1967). The east African rift system. *Nature* 215, 578–581.

Melesse, A., and Negesse, T. (2011). Phenotypic and morphological characterization of indigenous chicken populations in southern region of Ethiopia. *Anim. Genet. Resour.* 49, 19–31. doi: 10.1017/s2078633611000099

Michel, M. J. (2011). Spatial dependence of phenotype-environment associations for tadpoles in natural ponds. *Evol. Ecol.* 25, 915–932. doi: 10.1007/s10682-010-9441-y

Michel, M. J., Chien, H., Beachum, C. E., Bennett, M. G., and Knouft, J. H. (2017). Climate change, hydrology, and fish morphology: predictions using phenotype-environment associations. *Clim. Change* 140, 563–576. doi: 10.1007/s10584-016-1856-1

Mirkena, T., Walelign, E., Tewolde, N., Gari, G., Abebe, G., and Newman, S. (2018). Camel production systems in Ethiopia: a review of literature with notes on MERS-CoV risk factors. *Pastoralism* 8:30.

MoA (1998). *Agro-Ecological Zones of Ethiopia*. Addis Ababa: Ministry of Agriculture.

Muchadeyi, F. C., and Dzomba, E. F. (2017). "Genomics tools for the characterization of genetic adaptation of low input extensively raised chickens," in *Poultry Science*, ed. M. Manafi (London: IntechOpen).

Müntzing, A. (1971). Göte wilhelm turesson. *Taxon* 20, 773–775. doi: 10.2307/1218603

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., et al. (2014). ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity forMaxentecological niche models. *Methods Ecol. Evol.* 5, 1198–1205. doi: 10.1111/2041-210x.12261

Negassa, D., Melesse, A., and Banerje, S. (2014). Phenotypic characterization of indigenous chicken populations in Southeastern Oromia Regional State of Ethiopia. *Anim. Genet. Resour.* 55, 101–113. doi: 10.1017/s2078633614000319

Oddi, F. J., Miguez, F. E., Ghermandi, L., Bianchi, L. O., and Garibaldi, L. A. (2019). A nonlinear mixed-effects modeling approach for ecological data: using temporal dynamics of vegetation moisture as an example. *Ecol. Evol.* 9, 10225–10240. doi: 10.1002/ece3.5543

O'Donnell, M. S., and Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States. *US Geol. Survey Data Ser.* 691, 4–9.

Overdijk, S. (2019). *The Effect of Ecological Factors on the Morphology of Ethiopian Chickens*. Ph. D. thesis. Wageningen: Wageningen University and Research.

Pebesma, E., Bivand, R., Pebesma, M. E., RColorBrewer, S., and Collate, A. (2012). *Package 'sp'. The Comprehensive R Archive Network*.

Phillimore, A. B., Hadfield, J. D., Jones, O. R., and Smithers, R. J. (2010). Differences in spawning date between populations of common frog reveal local adaptation. *Proc. Natl. Acad. Sci.* 107, 8292–8297. doi: 10.1073/pnas.0913792107

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., and Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., et al. (2009). Sample selection bias and presenceâĂŘonly distribution models: implications for background and pseudo-absence data. *Ecol. Appl.*, 19, 181–197.

Radosavljevic, A., and Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* 41, 629–643. doi: 10.1111/jbi.12227

Rasband, W. S. (1997). *ImageJ 1.52a*. Bethesda, MD: National Institute of Health.

Razgour, O. (2015). Beyond species distribution modeling: a landscape genetics approach to investigating range shifts under future climate change. *Ecol. Inform.* 30, 250–256. doi: 10.1016/j.ecoinf.2015.05.007

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

SAS (2002). *User Installation Guide for the SAS® System Version 9 for Microsoft® Windows®* . Cary, NC: SAS Institute Inc.

Schmid, M., and Guillaume, F. (2017). The role of phenotypic plasticity on population differentiation. *Heredity* 119, 214–225. doi: 10.1038/hdy.2017.36

Selmoni, O., Vajana, E., Guillaume, A., Rochat, E., and Joost, S. (2020). Sampling strategy optimization to increase statistical power in landscape genomics: a simulation-based approach. *Mol. Ecol. Resour.* 20, 154–169. doi: 10.1111/1755-0998.13095

Shirk, A. J., Cushman, S. A., Waring, K. M., Wehenkel, C. A., Leal-Sáenz, A., Toney, C., et al. (2018). Southwestern white pine (*Pinus strobiformis*) species distribution models project a large range shift and contraction due to regional climatic changes. *For. Ecol. Manag.* 411, 176–186. doi: 10.1016/j.foreco.2018.01.025

Smith, A. B., Alsdurf, J., Knapp, M., Baer, S. G., and Johnson, L. C. (2017). Phenotypic distribution models corroborate species distribution models: a shift in the role and prevalence of a dominant prairie grass in response to climate change. *Glob. Chang. Biol.* 23, 4365–4375. doi: 10.1111/gcb.13666

Storz, J. F. (2002). Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Mol. Ecol.* 11, 2537–2551. doi: 10.1046/j.1365-294x.2002.01636.x

Tadesse, M., Alemu, B., Bekele, G., Tebikew, T., Chamberlin, J., and Benson, T. (2006). *Atlas of the Ethiopian Rural Economy*. Washington, DC: International Food Policy Research Institute.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series. B Methodol.* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Tilahun, H., and Schmidt, E. (2012). *Spatial Analysis of Livestock Production Patterns in Ethiopia. Ethiopia Strategy Support Program II (ESSP II) Working Paper 44*. Addis Ababa: Ethiopia Strategy Support Program II.

Venables, W., and Ripley, B. (2002). *Modern Applied Statistics with S*, 4th Edn. New York, NY: Springer-Verlag New York.

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845

Warren, D. L., Glor, R. E., and Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883. doi: 10.1111/j.1558-5646.2008.00482.x

Warren, D. L., Glor, R. E., and Turelli, M. (2010). ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33, 607–611.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., and Snyder, M. A. (2009). Niches, models, and climate change: assessing the assumptions and uncertainties. *Proc. Natl. Acad. Sci.* 106, 19729–19736. doi: 10.1073/pnas.0901639106

Wiley, M., and Wiley, J. F. (2019). *Advanced R Statistical Programming and Data Models: Analysis, Machine Learning, and Visualization*. Berkeley, CA: Apress.

Williams, G. (1966). *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Woldekiros, H., and D'Andrea, A. (2017). Early evidence for domestic chickens (*Gallus gallus* domesticus) in the Horn of Africa. *Int. J. Osteoarchaeol.* 27, 329–341. doi: 10.1002/oa.2540

Wood, S. (2012). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation.Bath, UK: University of Bath.

Wood, S. N., and Augustin, N. H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecol. Modell.* 157, 157–177. doi: 10.1016/s0304-3800(02)00193-x

Xiong, J., Thenkabail, P., Tilton, J., Gumma, M., Teluguntla, P., Congalton, R., et al. (2017). *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2015 Africa 30 m V001*.

Zuur, A., Ieno, E. N., and Smith, G. M. (2007). *Analyzing Ecological Data*. Sioux Falls, U.S.A.: U.S. Geological Survey (USGS). New York, NY: Springer.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer Science & Business Media.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fgene.2021.670390

Local Ancestry to Identify Selection in Response to Trypanosome Infection in Baoulé x Zebu Crossbred Cattle in Burkina Faso

Bernadette Yougbaré[1,2], Dominique Ouédraogo[1,3], Arnaud S. R. Tapsoba[3], Albert Soudré[4], Bienvenue L. Zoma[1,3], Pablo Orozco-terWengel[5], Sanou Moumouni[2], Salifou Ouédraogo-Koné[3], Maria Wurzinger[1], Hamidou H. Tamboura[2], Amadou Traoré[2], Okeyo Ally Mwai[6], Johann Sölkner[1], Negar Khayatzadeh[1,7], Gábor Mészáros[1] and Pamela A. Burger[8]*

[1] Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences Vienna (BOKU), Vienna, Austria, [2] Institut de l'Environnement et de Recherches Agricoles (INERA), Ouagadougou, Burkina Faso, [3] Institut du Développement Rural, Université Nazi Boni, Bobo-Dioulasso, Burkina Faso, [4] Unité de Formation et de Recherche en Sciences et Technologies, Université Norbert Zongo, Koudougou, Burkina Faso, [5] School of Biosciences, Cardiff University, Cardiff, United Kingdom, [6] International Livestock Research Institute (ILRI), Nairobi, Kenya, [7] SUISAG, Sempach, Switzerland, [8] Research Institute of Wildlife Ecology, Vetmeduni Vienna, Savoyenstraße 1, Vienna, Austria

The genomes of crossbred (admixed) individuals are a mosaic of ancestral haplotypes formed by recombination in each generation. The proportion of these ancestral haplotypes in certain genomic regions can be responsible for either susceptibility or tolerance against pathogens, and for performances in production traits. Using a medium-density genomic marker panel from the Illumina Bovine SNP50 BeadChip, we estimated individual admixture proportions for Baoulé x Zebu crossbred cattle in Burkina Faso, which were tested for trypanosome infection by direct ELISA from blood samples. Furthermore, we calculated local ancestry deviation from average for each SNP across 29 autosomes to identify potential regions under selection in the trypanotolerant Baoulé cattle and their crossbreds. We identified significant deviation from the local average ancestry (above 5 and 10% genome-wide thresholds) on chromosomes 8 and 19 in the positive animals, while the negative ones showed higher deviation on chromosomes 6, 19, 21, and 22. Some candidate genes on chromosome 6 (*PDGFRA*) and chromosome 19 (*CDC6*) have been found associated to trypanotolerance in West African taurines. Screening for F_{ST} outliers in trypanosome positive/negative animals we detected seven variants putatively under selection. Finally, we identified a minimum set of highly ancestry informative markers for routine admixture testing. The results of this study contribute to a better understanding of the genetic basis of trypanotolerance in Baoulé cattle and their crossbreeds. Furthermore, we provide a small informative marker set to monitor admixture in this valuable indigenous breed. As such, our results are important for conserving the genetic uniqueness and trypanotolerance of Baoulé cattle, as well as for the improvement of Baoulé and Zebu crossbreds in specific community-based breeding programs.

Keywords: admixture, local ancestry deviation, selection signature, SNP, *F*st, cattle, Burkina Faso

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman

Reviewed by:
Pablo Fonseca, University of Guelph, Canada
Hussain Mahdi Bahbahani, Kuwait University, Kuwait
Mohammed Kotb Abo-Ismail, California Polytechnic State University, United States

***Correspondence:**
Negar Khayatzadeh
kh.negar@gmail.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *21 February 2021*
Accepted: *09 August 2021*
Published: *27 September 2021*

Citation:
Yougbaré B, Ouédraogo D, Tapsoba ASR, Soudré A, Zoma BL, Orozco-terWengel P, Moumouni S, Ouédraogo-Koné S, Wurzinger M, Tamboura HH, Traoré A, Mwai OA, Sölkner J, Khayatzadeh N, Mészáros G and Burger PA (2021) Local Ancestry to Identify Selection in Response to Trypanosome Infection in Baoulé x Zebu Crossbred Cattle in Burkina Faso. Front. Genet. 12:670390. doi: 10.3389/fgene.2021.670390

INTRODUCTION

The *Bos taurus taurus* and *Bos taurus indicus* subspecies are the two most important cattle types in West Africa (Belemsaga et al., 2005; Okeyo Mwai et al., 2015). These animals have been raised in complex social and political processes, and they have adapted to harsh environmental conditions over the centuries (Dueppen, 2012). The adaptive traits include the tolerance to diseases and drought, ability to walk long distances, and capacity to survive on poor pastures (Okeyo Mwai et al., 2015). Yet, this valuable diversity is increasingly threatened by genetic dilution due to changes in production systems, livestock herders' preferences for specific breeds and/or traits, market conditions and other opportunities (Hanotte et al., 2010). The taurine cattle, more specifically the Baoulé cattle have existed in the tsetse fly (Glossina spp.) challenged zones for long and therefore acquired trypanotolerance, an immunology phenomenon that has a genetic basis (Naessens et al., 2002; Agyemang, 2005). These animals have a capacity to rid themselves of trypanosome parasites and maintain low parasitemia. Thus, trypanotolerant animals have been introduced in other tsetse affected countries of Africa to make use of their genetic advantage in purebred populations or crossed to other types like Zebu. Several studies revealed admixture among the taurine and Zebu subspecies (Hanotte, 2002; Freeman et al., 2004; Flori et al., 2014) as the result of the continuous genetic flow that occurs every year during seasonal cross-border livestock movements from the drier Sahelian zones in the north to the more humid zones in the south of West and East Africa.

The south-western region of Burkina Faso is the original habitat of Baoulé cattle. In this area, production systems are mixed crop-livestock and agroforestry, with the Lobi ethnic group concentrating on subsistence crop production while the transhumant people tend to keep their lifestyle of pastoral livestock production. The cattle population in this region is estimated at 343,000 heads, representing about 4% of the estimated national stock of 9 million according to the Ministry of Animal Resources (Ministere des ressources animales, 2014). The livestock system is extensive in all studied departments (Zoma et al., 2020), with 7 to 100 cattle per farmers. The indigenous Baoulé cattle, despite its small size and lower growth rate, is well adapted to the local environment of West Africa. It has gained cultural importance due to its social roles and tolerance to trypanosomiasis (Zoma et al., 2020). However, the continued crossbreeding with Zebu cattle because of its large size threatens the integrity of the Baoulé breed (Yougbaré et al., 2020). Recently, community-based breeding programs have been implemented in the south-western region of Burkina Faso to conserve and improve the local genetic resources of the indigenous Baoulé cattle as well as Baoulé x Zebu advanced crosses (Ouédraogo et al., 2020; Zoma et al., 2020).

Since the advent of high-throughput single nucleotide polymorphism (SNP) genotyping, inferring selection signatures from differences in local admixture levels has received considerable attention in human genetics (Tang et al., 2007; Jin et al., 2012; Bhatia et al., 2014). Similar studies in livestock investigated local ancestry levels of New World Creole cattle (Gautier and Naves, 2011; Flori et al., 2014; Pitt et al., 2018) and selection signatures in dairy cattle in East Africa, resulting from admixture of European breeds (Kim and Rothschild, 2014), as well as in East African short horn Zebu (Bahbahani et al., 2015).

The genomic ancestry proportions between trypano-susceptible indicine Zebu and the trypano-tolerant taurine Baoulé cattle in Burkina Faso were assessed based on microsatellites and 155 SNPs in 23 candidate regions (Smetko et al., 2015). In this study, we followed up and extended on the previous work using dense genomic marker data. Our study aimed to estimate the individual local ancestry proportions for each SNP to identify potential regions under selection in Baoulé x Zebu crossbreds tested positive or negative for trypanosomosis and finding a small set of ancestry informative SNP for routine admixture testing. Estimating the proportional contributions of ancestral populations in admixed (crossbred) individuals is important to clarify the population structure, historical background, and pattern of admixture along the genome of admixed individuals.

MATERIALS AND METHODS

Study Areas and Sample Collection

This study was carried out in the province of Poni in the South-western administrative region of Burkina Faso. We selected three study sites with different management and breeding systems of Baoulé cattle and Baoulé x shorthorn Zebu crossbreds, including 27 villages from the Bouroum-Bouroum, Kampti and Loropéni departments (**Figure 1**). In the Bouroum-Bouroum department, we worked with 55 sedentary farmers of the ethnic group of Lobi, who keep purebred Baoulé and are the owners of these animals. In Kampti, we included 18 farms with mostly Baoulé x Zebu crossbreds and some pure Baoulé owned by the Mossi ethnic group, but herded by the transhumant Fulani people. Finally, in Loropéni 15 farms were included mainly with crossbred animals kept by Lobi and Djan breeders. As all samples were collected within close geographic distances (< 50 km) in the tsetse infested province of Poni, it is reasonable to assume that all individuals were exposed to the same trypanosome infection challenge.

A total of 737 blood samples, including 387 Baoulé and Baoulé x Zebu crossbreds from Bouroum-Bouroum, and 350 crossbreeds from Kampti and Loropéni were collected in EDTA tubes during the health monitoring activities of the "Characterization and Sustainable Utilization of Local Cattle Breeds" (LoCaBreed) project. DNA extraction from EDTA blood samples was performed with the MasterPureTM DNA Purification Kit for Blood Version II (Biozym Scientific, Oldendorf, Germany) following the manufacturer's protocol. The trypanosomosis status was recorded by indirect ELISA test to diagnose positive or negative trypanosome infection in the blood samples (Desquesnes et al., 2003) resulting in a total of 377 positive and 360 negative animals (**Table 1**).

Genotype Data

The genotyping of the 737 DNA samples with the Illumina Bovine SNP50 BeadChip was performed at Neogen (Lincoln,

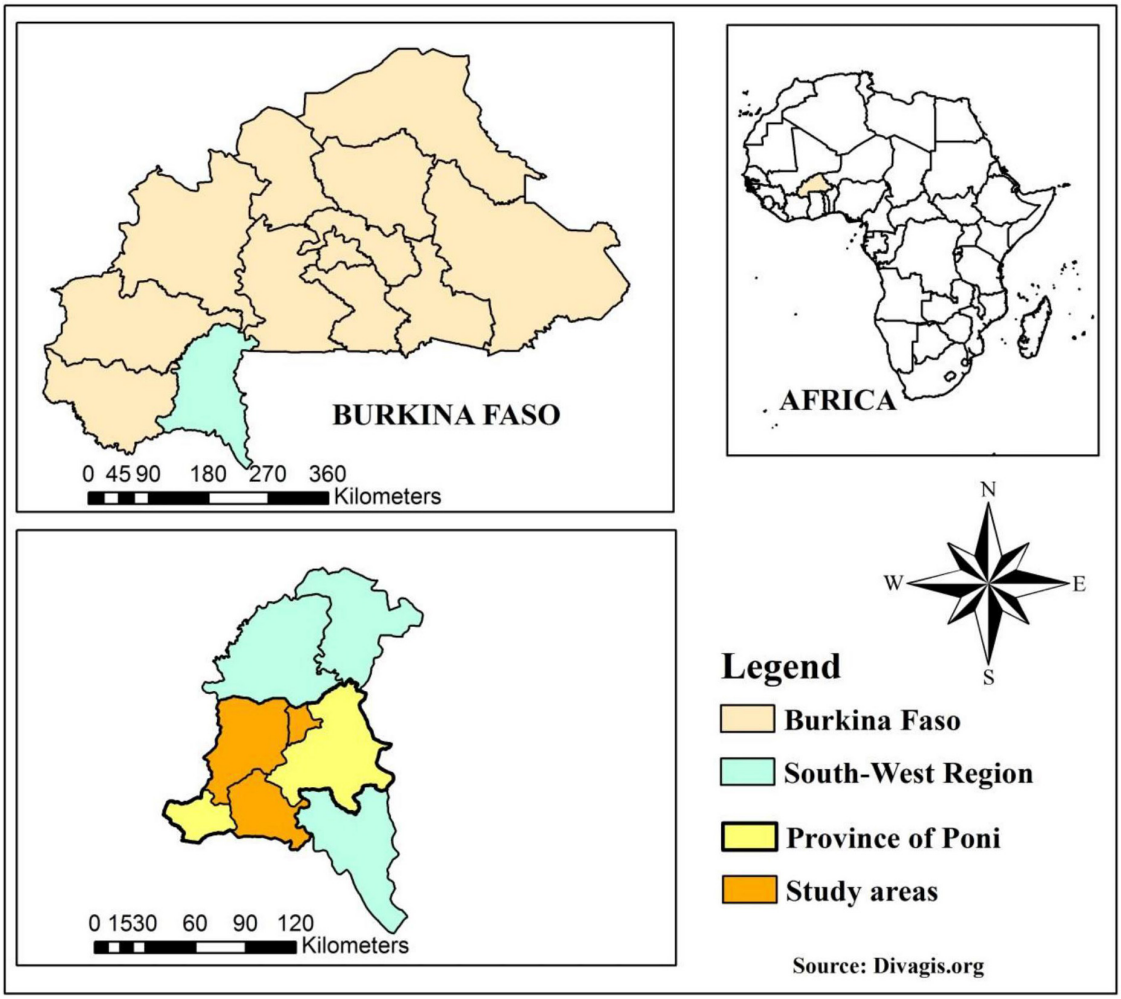

FIGURE 1 | Map of Burkina Faso showing the three study areas.

United States). Additional genotypes from 30 purebred Zebu and 35 crossbreds (Pérez et al., 2014) were included in the study to represent the two ancestral populations (Baoulé and Zebu) and to increase the number of crossbreds, summing to a total of 802 animals. Stringent quality filtering of the data was performed with PLINK 1.9 (Purcell et al., 2007; Chang et al., 2015). Specifically, the dataset was controlled to exclude non-autosomal SNPs, and SNPs with a minor allele frequency (MAF) lower than 0.05, a call rate less than 90% and those that deviated from Hardy Weinberg equilibrium with Fisher's exact test with P-value $1 \times 10 - 6$. After quality control, 28,034 SNPs and 776 animals were available for subsequent analyses.

Global Admixture Analysis

Unsupervised global ancestry estimation was performed with the full set of quality controlled SNPs using ADMIXTURE software (Alexander et al., 2009) with the number of ancestral populations (Baoulé and Zebu) fixed at two (K = 2). The admixture bar plots for ancestry proportions were created in R with the *barplot* function (The R Development Core Team, 2020). We calculated the frequencies of the admixture levels for all animals in Excel and plotted them in categories of 0.1 steps. We assigned 30 purebred Baoulé (global admixture levels $\geq$ 0.999 Baoulé) and 30 purebred Zebu (global admixture levels $\geq$ 0.987) as reference populations to investigate local admixture levels in 716 animals that were considered as potential crossbreds based on the sampling information. Animals found to be purebred (global admixture levels $\geq$ 0.995 Baoulé) were removed from the pool of crossbreds.

TABLE 1 | The 737 genotyped cattle with the trypanosomosis status.

Regions	Negative animals	Positive animals	Total
Bouroum-Bouroum	204	183	387
Kampti	87	89	176
Loropéni	69	105	174
Total	360	377	737

Local Ancestry Estimation in Admixed Populations

Local Ancestry in Mixed Populations (LAMP) is a program for estimating locus-specific ancestries in admixed individuals, using allele frequencies of the reference populations (Sankararaman et al., 2008). We applied the LAMP-ANC mode implemented in LAMP and provided the estimated allele frequencies files for Baoulé and Zebu as the purebred ancestral populations. LAMP-ANC is a modification of the LAMP mode and shows higher accuracy allowing triple mixing to be estimated, while LAMP cannot determine frequencies for more than two ancestral populations (Sankararaman et al., 2008). The following parameters were set: admixture proportions (alpha) = 0.8 for Baoulé and $\geq$ 0.2 for Zebu based on the global ancestry estimation using ADMIXTURE program, number of generations since admixture (g) = 2 and recombination rate (r) = 10^{-8}. We estimated the local ancestry proportion, as well as the "delta ancestry" with R in trypanosome positive and negative trypanosomosis animals following Khayatzadeh et al. (2016) using a custom script (see section "Data Availability"). The "delta ancestry" reflects the extreme fluctuations in ancestry differences across the genome, which are calculated by subtracting the genome-wide ancestry from locus-specific ancestry for each ancestry component. Such extreme fluctuations in ancestry differences are unlikely to have occurred by random genetic drift and potentially exhibit a selection signature in the admixed individuals (Tang et al., 2007). To identify significant deviations from the genome-wide average ancestry, we performed permutation tests (Doerge and Churchill, 1996) of the local ancestry proportions over the whole genome of admixed animals as proposed and carried out by Tang et al. (2007) in an admixed human population (Puerto Ricans) and replicated by Gautier and Naves (2011), Flori et al. (2014) in African Taurine, and Khayatzadeh et al. (2016) in composite cattle breed (Swiss Fleckvieh) to find significant thresholds for the deviations of local genetic ancestries from global ancestries. Separating animals with positive and negative trypanosomosis status, for each animal we concatenated the local ancestry estimations of all 29 autosomes and then permuted the circularized genome by cutting at a random location and rearranging the two resulting pieces of the genome for each individual independently. This type of permutation largely preserves the extent of Linkage Disequilibrium (LD), assuming that it is homogeneously distributed over the whole genome. We implemented 1,000 permutations. The distributions of maximum and minimum over all permutations were then used to define the 5 and 10% genome-wide thresholds levels that indicated significant deviation of the observed local ancestries from the genome-wide average ancestry (Tang et al., 2007; Gautier and Naves, 2011; Khayatzadeh et al., 2016).

F_{ST} Outlier Analysis

We applied BayeScan 2.1 (Foll and Gaggiotti, 2008) to identify F_{ST} outlier loci putatively under selection between the trypanosome positive (n = 244) and negative (n = 266) crossbred animals using a cut-off at $p < 0.05$ corrected for a false discovery rate [FDR; (Benjamini and Hochberg, 1995)]. BayeScan uses a Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) algorithm to obtain posterior distributions, with 100,000 iterations and a Burn-in length of 50,000 iterations. The regions within $\pm$ 0.5 Mb of the most significant SNPs were searched for any potential associated genes based on the ARS UCD1.2 Bos Taurus Genome Assembly on the NCBI database.

Identification of Ancestry Informative SNPs for Effective Hybrid Detection

We aimed to identify SNPs with the highest F_{ST} differentiation between the 30 pure Baoulé and 30 pure Zebu. We re-filtered the original dataset for MAF < 10%, individual and genotype missingness < 10%, respectively. The F_{ST} values were calculated in PLINK following Weir and Cockerham (1984). With these we were able to provide a set of top 200 F_{ST} markers, which were then used as a starting point to manually remove markers less than 5 Mb to each other—preference given to higher F_{ST} markers. Based on this, we selected the top 15, 25, 50, and 100 SNPs, and extracted these for the crossbred animals and repeated the global admixture analysis (K = 2). We used the *cor* function in R to calculate the Pearson correlation coefficient (Pearson's r) for pairwise determining the linear association between admixture levels (ancestry proportion) estimated based on the different sets of ancestry informative SNPs [all SNPs (35,952 SNPs) versus the top 100, 50, 25, 15 SNPs]. Significance of the Pearson's r was assessed with the P-value from the Pearson Correlation Coefficient Calculator (Social Science Statistics, 2021).

RESULTS

Global Admixture of All Animals

The individual admixture proportions using the full set of SNPs were estimated for all pure and admixed animals and are presented in **Figure 2**. The distribution of the global admixture proportions for the 802 animals is presented in **Figure 3**. Notably, we detected 91 cattle with a Baoulé ancestry > 0.995 among the presumed crossbreds, which we excluded from the subsequent analysis of local ancestry in admixed animals.

Local Ancestry and the Delta Ancestry Across Chromosomes in Trypanosome Positive and Negative Crossbreds

The average ancestry estimation for every single SNP was performed across 29 autosomes for trypanosome positive and negative Baoulé x Zebu crossbreds, respectively. The permutation tests over all chromosomes indicated significant local ancestry deviation from the average (above the 5 and 10% genome-wide thresholds) in chromosomes 8 and 19 for trypanosome positive crossbreds (**Figure 4**), and in chromosomes 6, 19, 21, and 22 for trypanosome negative animals, respectively (**Figure 5**). The local admixture proportions for these chromosomes are presented in **Figure 6** and for

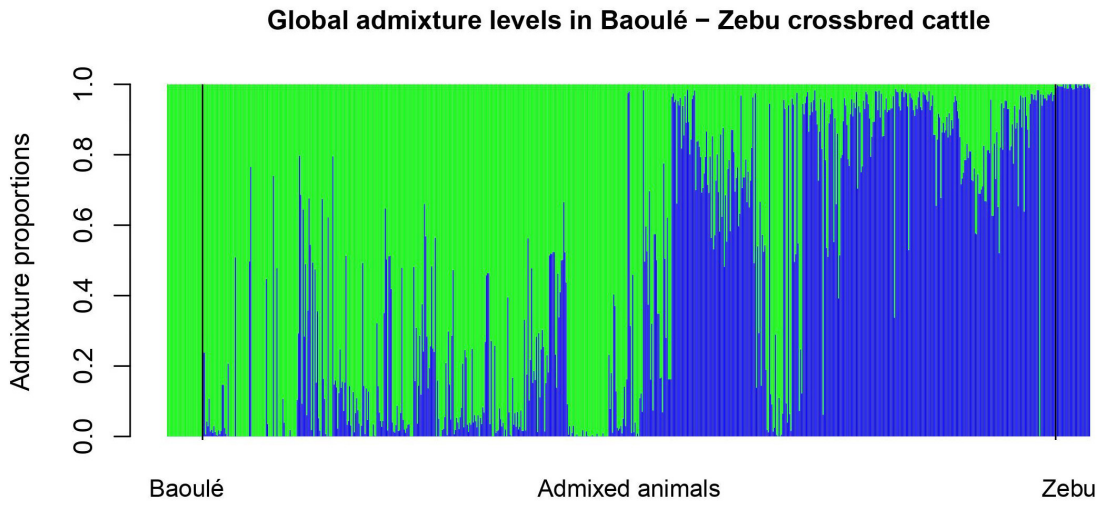

FIGURE 2 | Global Admixture in Baoulé—Zebu crossbred cattle with the quality-filtered set of 28,034 single nucleotide polymorphisms (SNPs).

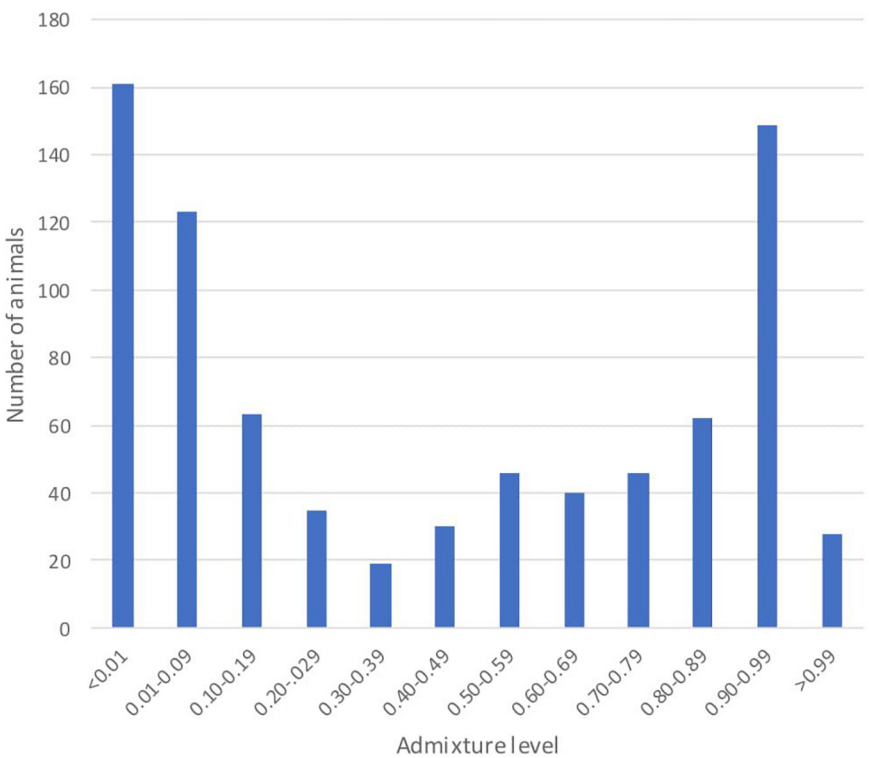

FIGURE 3 | Distribution of the global admixture proportions for 802 animals with the full set of 31,612 single nucleotide polymorphisms (SNPs).

all other chromosomes in **Supplementary Figures 1**, **2**. We further visualized the deviations from the average ancestry in the respective chromosomes and identified regions of higher delta ancestry (wide peaks) on chromosome 8 between 35–50 Mb and in chromosome 21 between 20–35 Mb and 40–50 Mb, respectively (**Figure 7**). These genomic regions might harbor candidate genes associated to tolerance or susceptibility of trypanosomosis.

F_{ST} Outliers Between Trypanosome Positive and Negative Crossbreds

We screened the genomes of the Baoule and Zebu crossbred animals for outlier SNPs with high F_{ST} values and disregarded the pure-bred Baulé and Zebu. Among these crossbred animals we grouped them in trypanosome positive and negative animals to avoid detection of breed differences unrelated to trypanosome

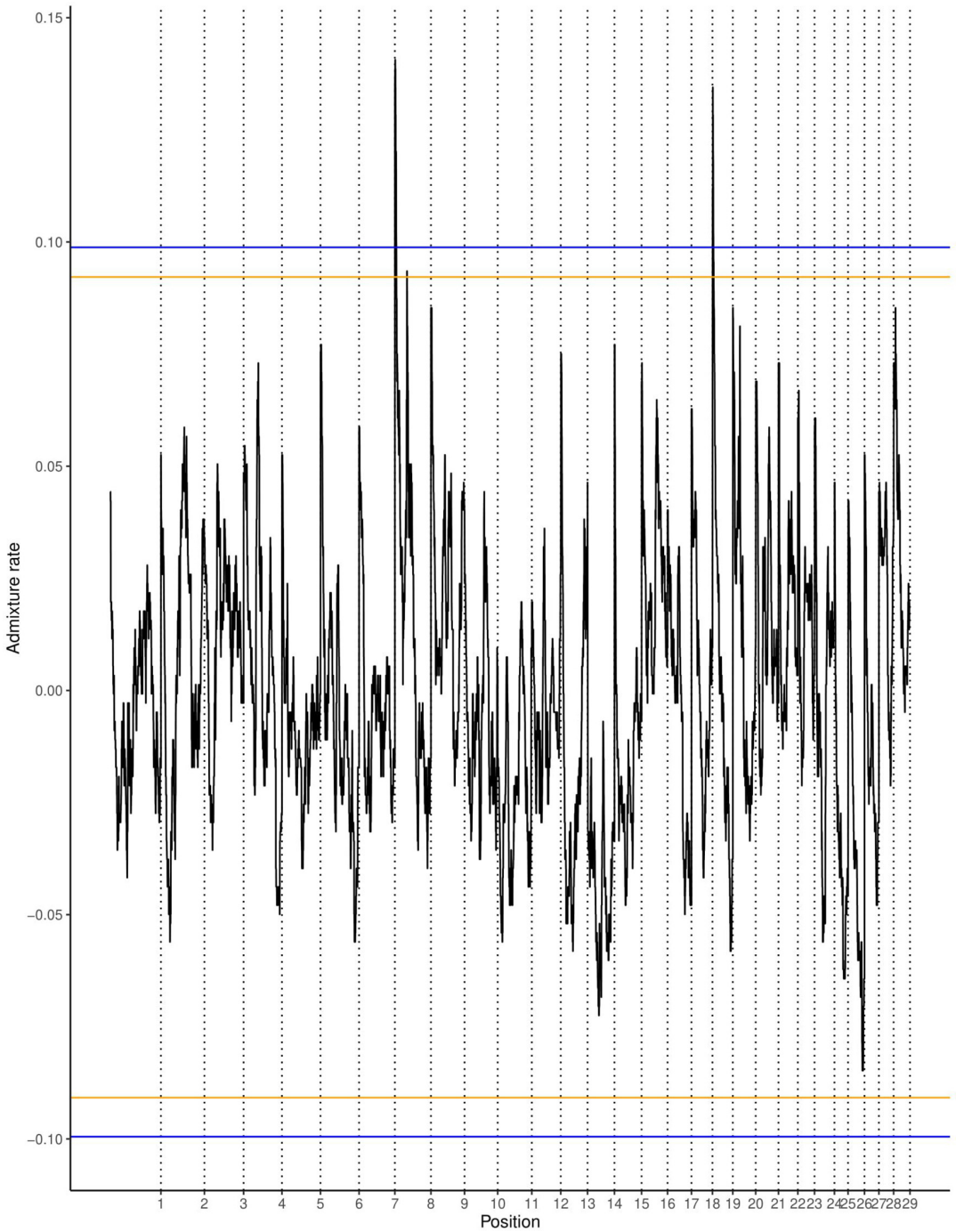

FIGURE 4 | Local ancestry deviations based on the permutation threshold for the 244 positive crossbreds animals. Orange and blue lines signify the 5 and 10% genome-wide threshold.

tolerance. We detected seven variants with a FDR corrected threshold of $p < 0.05$ (**Figure 8**). The seven outlier SNPs with the highest levels of F_{ST} values were found in chromosomes 2, 3, 5, 20, 21, and 23, and are presented in **Table 2**, together with their neighboring genes. The positions of the SNPs were not located in regions with higher delta ancestry.

Identification of the Most Ancestry Informative Markers

To reliably detect hybrids even with a small set of SNPs applicable for routine genetic monitoring, we selected the most ancestry informative markers resulting in the highest differentiation between Baoulé and Zebu cattle. The 100 SNPs

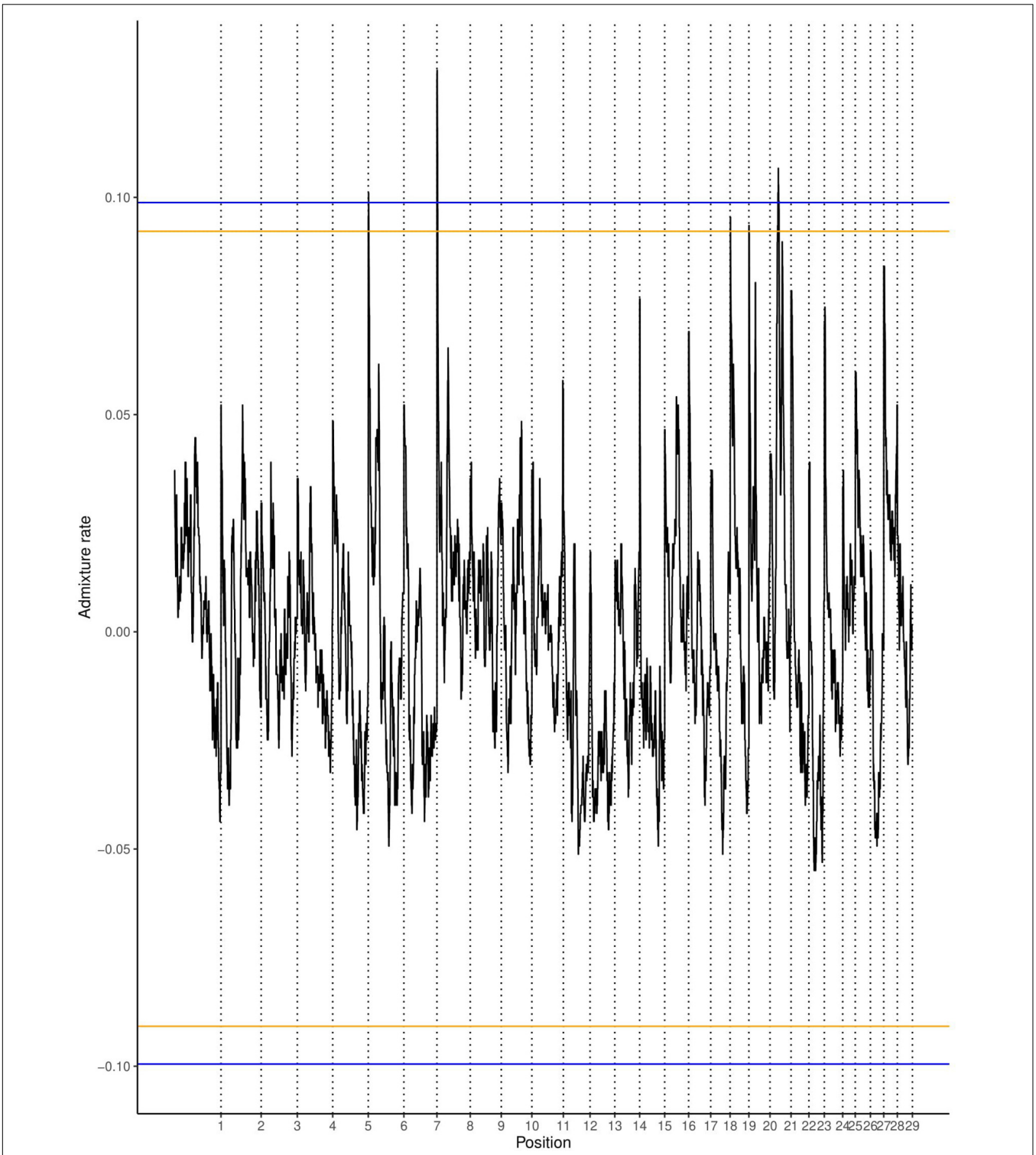

FIGURE 5 | Local ancestry deviations based on the permutation threshold for the 266 negative crossbreds animals. Orange and blue lines signify the 5 and 10% genome-wide threshold.

with the highest divergence presented F_{ST} values ranging between 0.98 and 0.79 (**Supplementary Table 1**). We estimated admixture levels of the crossbred individuals using the top 15, 25, 50, and 100 SNPs (**Supplementary Figure 5**). The Pearson correlation coefficients *r* between the levels of admixture obtained with the full dataset of 35,952 SNPs and each of

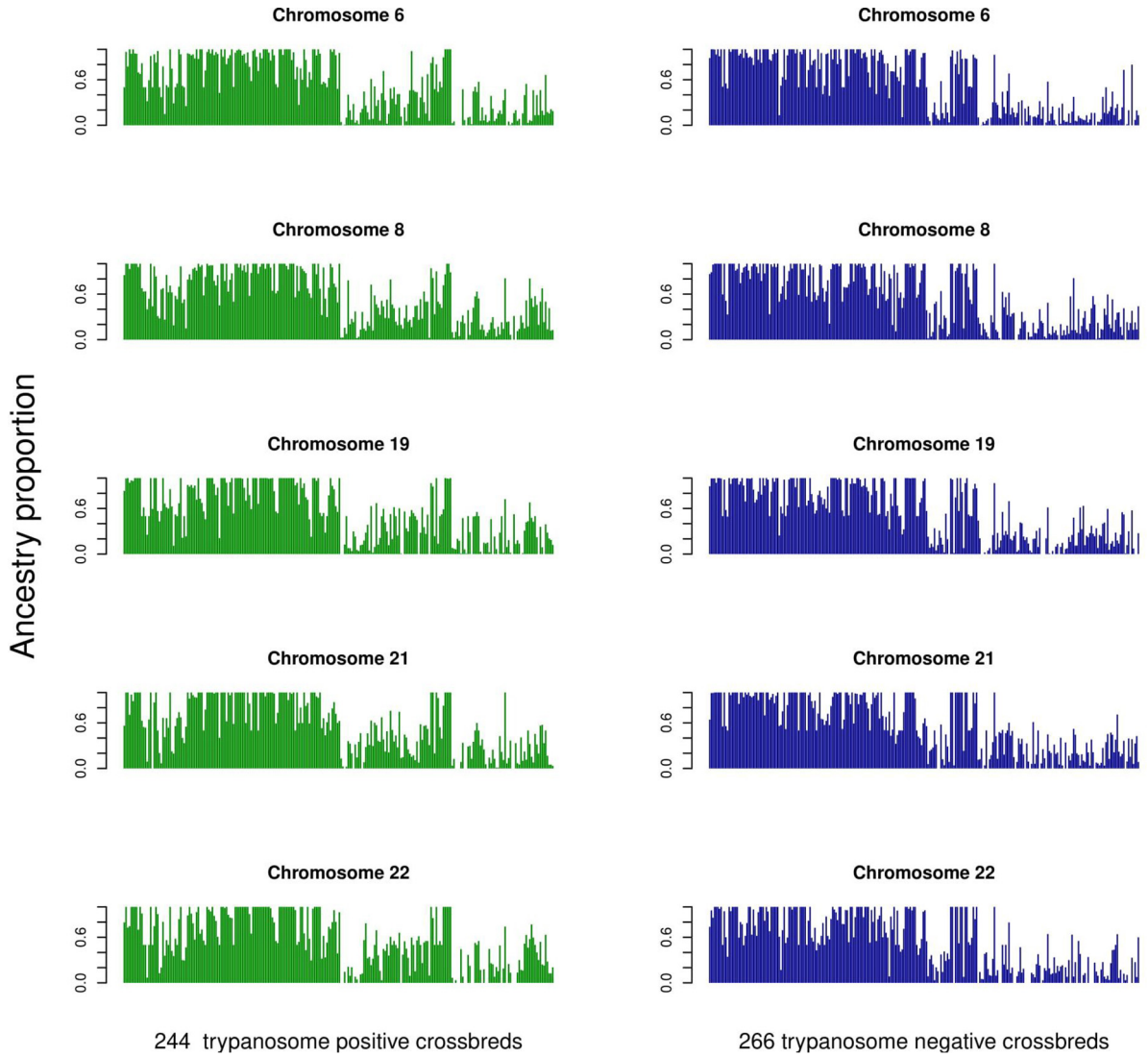

FIGURE 6 | Individual admixture proportions across chromosomes 6, 8, 19, 21, and 22 for the 244 trypanosomose positive and 266 negative crossbreds as determined by LAMP.

the sets of top SNPs were generally high and ranged between 0.949 (allSNPs/top15) and 0.990 (allSNPs/top100) (**Table 3**). All Pearson correlation coefficients were statistically significant ($p < 0.001$).

DISCUSSION

Global Admixture in the South-Western Taurine Cattle Population of Burkina Faso

The high amounts of global admixture observed in the taurine cattle population in the three studied departments of Burkina Faso indicated mixed genetic backgrounds of the cattle in Bouroum-Bouroum, Kampti and Loropeni (**Figures 2**, **3**). The observed admixture levels within the departments are likely due to unrestricted mating among cattle of different genetic backgrounds. Long-distance migrations within and across countries, utilization of communal pastures, exchange of breeding animals, and uncontrolled mating facilitate constant gene flow. Houessou et al. (2019) explained this situation by lack of selection and high levels of gene flow due to cyclical cross-border cattle herd movements known as "transhumance" and extensive commercial transactions of cattle in the West African region.

The uncontrolled mating in extensive production systems, which are typically practiced in West Africa, can lead to the introgression of Zebu genes in the small taurine cattle population, which represents a threat to their genetic integrity (Dossa and Vanvanhossou, 2016), and might lead to a potential dilution of

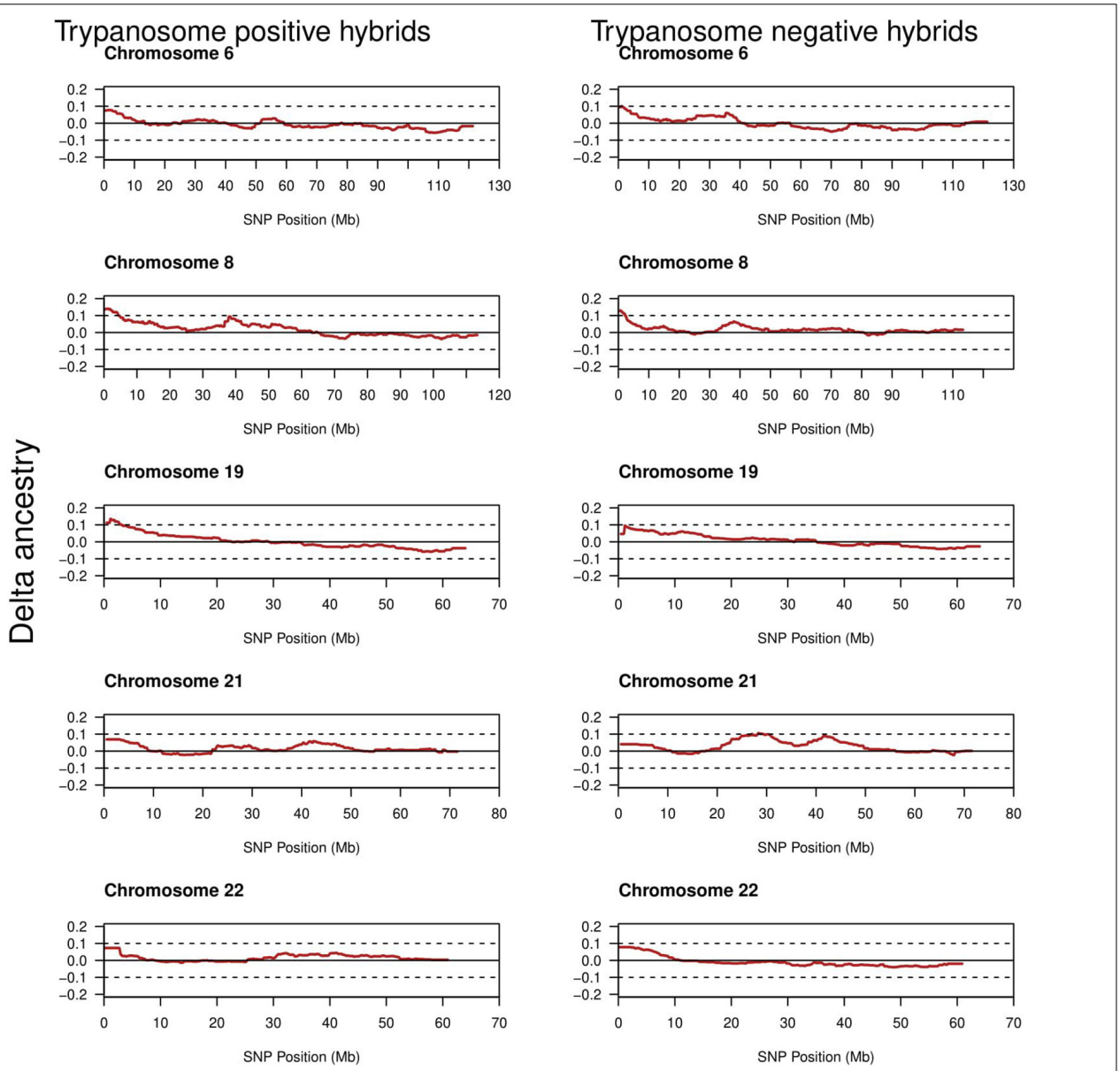

FIGURE 7 | The delta ancestry across chromosomes 6, 8, 19, 21, and 22 for the 244 positive and 266 negative crossbreds trypanosomosis status. The red line shows the deviation.

their trypanotolerance (Traoré et al., 2015; Albert et al., 2019). The increasing importance of Zebu in the South-western region of Burkina Faso might endanger Baoulé cattle in the long term. As the North, which is the preferred area for Zebu cattle, is hit by drought, increasing numbers of Zebu cattle breeders looking for pasture are moving to the South-west where the climate is quite favorable and grass is still abundant. Thus, suitable management is required for the sustainable use of local breeds, and recently community-based breeding programs (CBBP) for Baoulé cattle and their crossbreds have been implemented (Ouédraogo et al., 2020). Within the CBBP, Zoma et al. (2020) identified four distinct types of cattle production systems sedentary Lobi farms, sedentary crossbreed farms, semi-transhumant Fulani Zebu farms, and transhumant Fulani Zebu farms. The admixture between Zebus and Baoulé cattle observed in this study could be due to differences in the production systems. Furthermore, notable size differences between purebred and crossbred Baoulé were confirmed (Yougbaré et al., 2020) and breeders prefer to have large animals like Zebu cattle. As shown in **Figure 2**, we identified several purebred Baoulé cattle that had been considered as admixed based on the sampling information. These animals originated from the populations of Loropeni and Bouroum-Bouroum where the farmers have a preference for breeding purebred Baoulé (Zoma et al., 2020).

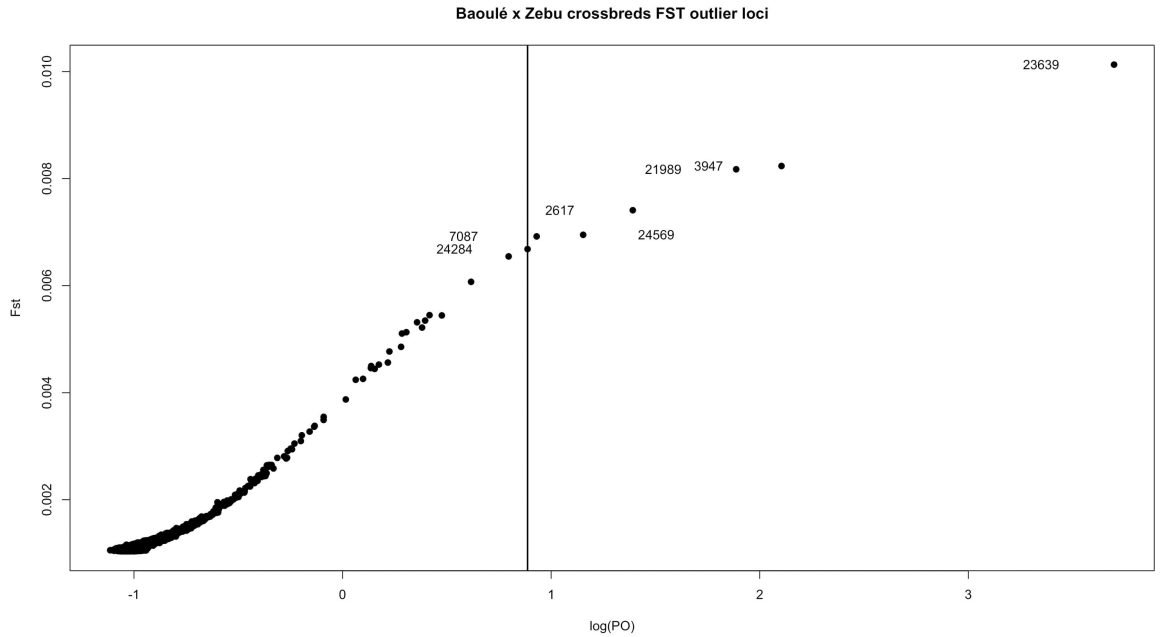

FIGURE 8 | F_{ST} outliers between trypanosome positive and negative crossbreds. The vertical line shows the FDR corrected cut-off ($p < 0.05$); the outlier SNP names and positions are provided in **Table 2**.

Different Local Ancestry in Trypanosome Positive and Negative Baoulé x Zebu Crosses

In a recently admixed population, ancestral populations have been mixing for a relatively small number of generations, resulting in a new population with different proportions of their genome derived from the original parental populations (Khayatzadeh et al., 2016). Local ancestry analysis of admixed populations has been successfully used to detect recent selection in admixed Swiss Fleckvieh cattle (Khayatzadeh et al., 2016), as well as selection for Zebu introgressed regions in Colombian creole taurine cattle (Pitt et al., 2018). In our study, we applied this approach to identify significantly different local admixture levels and detected five chromosomes with higher deviation from the average ancestries, with an excess of Baoulé ancestry, which might account for a higher tolerance to trypanosomiasis. Similarly, Decker et al. (2014) investigated the population structure of domesticated cattle and calculated Asian indicine (*B. t. indicus*), Eurasian taurine, and African taurine (both *B. t. taurus*) ancestry proportions.

We applied an approach of significance testing and performed a permutation test of circularizing the genome by concatenating the SNPs of all autosomes in a single string, cutting this string once and rearranging the two resulting segments, as proposed by Tang et al. (2007). The permutation approach removes not only the effects of selection, but also the local effects of genetic drift; the threshold is considered non-conservative. Nevertheless, based on simulations (Tang et al., 2007) outliers are unlikely to be due to genetic drift. Therefore, this procedure is considered robust to find significant signals for selection while accounting for confounding effects of demographic history of the admixed cattle.

We found regions deviating from the average ancestry with a higher amount of Baoulé proportions on chromosomes 6, 8, and 19 in trypanosome negative individuals. A previous study (Noyes et al., 2011) identified *VAV1, PIK3R5, RAC1, VAV2, GAB2,* and *INPP5D* genes in chromosome 8 to be genes under selection in Muturu and N'Dama cattle breeds in response to trypanosomes infection. Surprisingly, we also found higher Baoulé ancestry in chromosome 8 (35–50 Mb) also in trypanosome positive cattle, which could indicate that these regions harbor beneficial Baoulé haplotypes, which are not connected to trypanosomosis tolerance. These regions might harbor genes of general importance for adaptation to the environment. Some canadidate genes on chromosome 6 at 71373513-71421283 (*PDGFRA*) and chromosome 19 at 41185975-41196948 (*CDC6*) for trypanotolerance in West African taurines have been found on these chromosomes (Tijjani, 2019) overlapping with the regions identified in our study. Furthermore, Smetko et al. (2015) identified chromosomes 7 and 22 as regions with the highest Baoulé ancestry proportion, similar to our results.

Genes Under Potential Selection Identified by F_{ST} Outlier Tests

Identifying recent positive selection signatures in domesticated animals can provide information on beneficial mutations and their underlying biological pathways for economically important traits. Global F_{ST} values are one useful method to detect selection signatures across breeds (Biswas and Akey, 2006). The seven

TABLE 2 | The 7 outlier SNPs with the highest F_{ST} values.

SNP	CHR	SNP name	Position	Genes
2617	2	BovineHD0200021582	75210246	LOC100138101, LOC101902632
3947	3	BovineHD0300017052	56576857	HS2ST1, LMO4, ENSBTAG00000054817, and ENSBTAG00000052091
7087	5	ARS-BFGL-NGS-110363	108172899	CACNA1C, DCP1B, CACNA2D4, LRTM2, ADIPOR2, LOC101903199, ERC1, WNT5B, WNK1, RAD52, and FBXL14
21989	20	BovineHD2000008166	27620224	ISL1
23639	21	ARS-BFGL-NGS-22971	11067328	LOC107131341, NR2F2, and LOC101907985
24284	23	BovineHD4100016034	20992806	CD2AP, ADGRF2, ADGRF4, OPN5, and PTCHD4
24569	23	BovineHD2300012802	44142970	PHACTR1, HIVEP1, ADTRP, and EDN1

outlier SNPs, which we identified between trypanosome positive and negative crossbreds, were on chromosome 2, 3, 5, 20, 21, and 23. The chromosomes BTA 2, 3, 5, and 23 have previously been identified harboring common candidate genes in Muturu and N'Dama breeds linked to trypanotolerance in West African taurine population as well as selected candidate genes in Muturu cattle only (Tijjani et al., 2019). Functional annotation and enrichment analyses based on Reactome pathways in PANTHER ver 13.1 (Thomas et al., 2003) confirmed their relevance in response to trypanosome infection pathways. In our study, we identified other genes (**Table 2**) such as *LOC100138101, LMO4, LTRM2, ISL1, PTCHD4, and HIVEP1* as genes potentially responsible for trypanotolerance.

From previous studies genes such as *TICAM1, ARHGAP15, SLC40A1, GFM1,* and *INHBA* have been proposed as candidate genes for trypanotolerance on chromosomes 2, 3, and 5 (Dayo et al., 2009; Noyes et al., 2011). Bahbahani et al. (2018) identified the genes *LTA4H*, *IL7*, *IL15*, *FCN*, *LTA4H,* and *NFAM1* as potential targets of natural selection related to immunity in Sheko cattle, which are a mixture of Asian zebu and African taurine ancestry and considered a trypanotolerant breed with high potential for milk production.

Ancestry Informative Markers to Detect Admixture for Routine Genetic Monitoring

The indigenous cattle breeds are disappearing because of indiscriminate crossbreeding by individual farmers, and schemes for genetic improvement developed without concern for preserving locally adapted breeds (Belemsaga et al., 2005). Many breeding programs or genetic improvement strategies in developing countries failed due to the lack of involvement of farmers in the different steps of implementation (Wurzinger et al., 2011). In many developing countries, livestock crossbreeding has been implemented with poor or no pedigree recording. Thus, ancestry informative markers would provide a great opportunity to estimate the level of admixture in a cost-effective way. Sölkner et al. (2010) proposed that individual admixture levels were estimated more accurately based on the genomic data using panels of pure reference animals, compared to estimation based on pedigree. (Getachew et al., 2017) indicated that the Ovine 50KSNP array is a powerful tool to identify small sets of AIMs for admixture studies in crossbred sheep populations in Ethiopia.

TABLE 3 | Pearson's correlation coefficient matrix displaying r^2 values between the levels of admixture using the most ancestry informative markers compared to the full data set of 35,952 SNPs (allSNPs).

	Top100	Top50	Top25	Top15
allSNPs	0.990	0.984	0.970	0.949
Top100		0.994	0.980	0.962
Top50			0.986	0.966
Top25				0.984

The minimum set of the 25 highest differentiating SNPs (**Supplementary Table 1**) can be used to develop an efficient competitive allele-specific PCR (KASP^{TM}, LGC Group, United States) genotyping assay. Such an easy and fast genotyping array can be implemented at any laboratory equipped with Real-Time PCR machine and can be used for routine monitoring of hybridization in Baoulé cattle. Getachew et al. (2017) identified a total of 74 SNPs from the Ovine 50K SNP data as AIMs. The SNPs were selected based on their F_{ST} values showing the highest levels of allele frequency differentiation between the two parental breeds similar to our methodology. These AIMs provided close estimation with pedigree information. Correlation coefficient between breed level based on admixture estimates from 25 SNP data obtained in this study (r = 0.99; **Table 3**) was higher compared to the correlation value of 0.96 obtained from ∼500 AIMs suggested to predict breed composition in cattle (Frkonja et al., 2012) or the correlation values in the range of 0.89 to 0.96 reported for different human populations in prediction of admixture levels (Halder et al., 2008). Other studies (Judge et al., 2017) recommended at least 300 informative SNPs identified based on similar diversity parameters to be used for genomic-based breed composition prediction. However, as the purpose of our AIM set was to differentiate between only two ancestral breeds (Baoulé and Zebu) the number of 25 highest differentiating SNPs was sufficient to detect admixed individuals (**Table 3** and **Supplementary Figure 5**). The existence of such a 25 SNP set allows their genotyping locally in Burkina Faso, providing a sustainable and low-cost solution to monitor admixture rates in these populations. We will further validate the 25 most AIMs in a larger group of confirmed crossbreds in Burkina Faso. Understanding the relationship between genetic admixture and performances is crucial for the success for local cattle breed conservation and crossbreeding programs. Ideally,

a combination of pedigree and genomic information is used in breeding programs. Applying small sets of AIMs is a cost-effective option to estimate the levels of admixture in situations where pedigree recording is difficult like in Burkina Faso.

CONCLUSION

In this study, we identified local ancestry proportions in genomic regions potentially related to trypanotolerance in front of a global admixture background. Based on a 10% genome-wide threshold exploring extreme deviations from the average distribution of delta ancestry, the chromosomes 6, 8, 19, 21, and 22 contained higher ancestral proportions of Baoulé cattle. Furthermore, we identified genes such as *LOC100138101*, *LMO4*, *LTRM2*, *ISL1*, *PTCHD4*, and *HIVEP1* as genes potentially responsible for trypanotolerance. Identification of genomic regions harboring genes related to trypanotolerance is a strong argument for conservation not only of Baoulé cattle, but all trypanotolerant breeds. The subsequent integration of these regions to genomes of non-trypanotolerant breeds *via* admixture provides a sustainable and effective use of these breeds, despite their lower production characteristics. As such, our study contributes to a better understanding of the genetic mechanism underlying trypanotolerance and will allow building a suitable breeding strategy for Baoulé cattle and their crossbreds in the south-western region of Burkina Faso.

The results indicate that the Bovine 50KSNP array is a powerful tool to identify small sets of AIMs as a cost-effective option to estimate the levels of admixture in situations where pedigree recording is difficult like in Burkina Faso. The minimum set of the 25 highest differentiating SNPs can be used to develop an efficient competitive allele-specific PCR assay.

DATA AVAILABILITY STATEMENT

Quality controlled Bovine 50k SNP chip data, including 31,612 SNPs of the 802 animals included in this study, were uploaded to DRYAD. The dataset has been assigned a unique identifier (doi: 10.5061/dryad.547d7wm7f).

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because Samples were collected during the official health monitoring activities of the APPEAR Project "Characterization and Sustainable Utilization of Local Cattle Breeds in Burkina Faso" approved by the Ministry of Agriculture and Irrigation Development (Ministère de l'Agriculture et des Aménagements Hydro-agricoles), Burkina Faso.

AUTHOR CONTRIBUTIONS

JS conceived the original idea of the study and together with GM, PB, BY, and NK further developed the idea and decided on the set of analysis. BY and PB did the statistical analysis and wrote the text. BY, DO, BZ, AS, SO-K, SM, HT, and ATr collected the data for the analysis and together with JS, GM, PB, NK, ATa, PO-W, MW, ATr, and OM critically reviewed the text. All authors approved the final version of the manuscript.

FUNDING

This study was funded by Austrian Partnership in High Education and Research for Development (APPEAR) through project 120 LoCaBreed: Local cattle breed of Burkina Faso, characterization and sustainable use.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.670390/full#supplementary-material

Supplementary Figure 1 | The local ancestry estimation plot for 29 autosomes chromosomes for 266 trypanosome negative, in Baoulé X Zebu crossbred cattle, excluding CHR 6, 8, 19, 21, and 22 which are presented in **Figure 6**.

Supplementary Figure 2 | The local ancestry estimation plot for 29 autosomes chromosomes for 244 trypanosome positive, in Baoulé X Zebu crossbred cattle, excluding CHR 6, 8, 19, 21, and 22 which are presented in **Figure 6**.

Supplementary Figure 3 | The delta ancestry for 29 autosomes chromosomes for 244 trypanosome positive, in Baoulé X Zebu crossbred cattle excluding CHR 6, 8, 19, 21, and 22, which are presented in **Figure 7**.

Supplementary Figure 4 | The delta ancestry for 29 autosomes for 266 trypanosome negative Baoulé X Zebu crossbred cattle excluding CHR 6, 8, 19, 21, and 22 which are presented in **Figure 7**.

Supplementary Figure 5 | Admixture graphs of the top 100, 50, 25, and 15 ancestry informative SNPs.

Supplementary Table 1 | The 100 highest differentiating SNPs according the F_{ST} value.

REFERENCES

Agyemang, K. (2005). *Trypanotolerant Livestock in the Context of Trypanosomiasis Intervention Strategies*. Rome: Food & Agriculture Org.

Albert, S., Moustapha, G., Stéphane, A. R. T., and Amadou, T. (2019). Is the introgression of Lobi/Baoul cattle by zebuine genes in Burkina Faso Lobi cattle threatened? *Afr. J. Biotechnol.* 18, 77–85. doi: 10.5897/AJB2018.16467

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Bahbahani, H., Afana, A., and Wragg, D. (2018). Signatures génomiques de l'introgression adaptative et de l'adaptation environnementale chez les bovins Sheko du sud-ouest de l'Éthiopie. *PLoS One* 13:e0202479. doi: 10.1371/journal.pone.0202479

Bahbahani, H., Clifford, H., Wragg, D., and Mbole-Kariuki, M. N. (2015). Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Sci. Rep.* 5:11729. doi: 10.1038/srep11729

Belemsaga, D. M. A., Lombo, Y., Thevenon, S., and Sylla, S. (2005). "Inventory Analysis of West African Cattle Breeds," in *Applications of Gene-Based Technologies for Improving Animal Production and Health in Developing Countries*, eds H. P. S. Makkar and G. J. Viljoen (Berlin: Springer-Verlag), 167–173. doi: 10.1007/1-4020-3312-5_15

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B. Meth.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bhatia, G., Tandon, A., Patterson, N., and Aldrich, M. C. (2014). Genome-wide Scan of 29,141 African Americans Finds No Evidence of Directional Selection since Admixture. *Am. J. Med. Genet.* 95, 437–444. doi: 10.1016/j.ajhg.2014.08.011

Biswas, S., and Akey, J. M. (2006). Genomic insights into positive selection. *Trends Genet.* 22, 437–446. doi: 10.1016/j.tig.2006.06.005

Chang, C. C., Chow, C. C., Tellier, L. C., and Vattikuti, S. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7.

Dayo, G.-K., Thevenon, S., Berthier, D., Moazami-Goudarzi, K., Denis, C., Cuny, G., et al. (2009). Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. *Mol. Ecol.* 18, 1801–1813. doi: 10.1111/j.1365-294X.2009.04141.x

Decker, J. E., McKay, S. D., Rolf, M. M., and Kim, J. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254

Desquesnes, M., Bengaly, Z., and Dia, M. L. (2003). Evaluation de la persistance des anticorps détectés par Elisa-indirect Trypanosoma vivax après traitement trypanocide chez des. *Rev. Elev. Med. Vet. Pays. Trop.* 56, 141–144. doi: 10.19182/remvt.9855

Doerge, R. W., and Churchill, G. A. (1996). Permutation Tests for Multiple Loci Affecting a Quantitative Character. *Genetics* 142, 285–294. doi: 10.1093/genetics/142.1.285

Dossa, L. H., and Vanvanhossou, F. U. S. (2016). The indigenous Somba cattle of the hilly Atacora region in North-West Benin: threats and opportunities for its sustainable use. *Trop. Anim. Health Prod.* 48, 349–359. doi: 10.1007/s11250-015-0958-5

Dueppen, S. A. (2012). Cattle in the West African savanna: evidence from 1st millennium CE Kirikongo, Burkina Faso. *J. Archaeol. Sci. Rep.* 39, 92–101. doi: 10.1016/j.jas.2011.09.005

Flori, L., Thevenon, S., Dayo, G.-K., and Senou, M. (2014). Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. *Mol. Ecol.* 23, 3241–3257. doi: 10.1111/mec.12816

Foll, M., and Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221

Freeman, A. R., Meghen, C. M., MacHUGH, D. E., and Loftus, R. T. (2004). Admixture and diversity in West African cattle populations. *Mol. Ecol.* 13, 3477–3487. doi: 10.1111/j.1365-294x.2004.02311.x

Frkonja, A., Gredler, B., Schnyder, U., Curik, I., and Sölkner, J. (2012). Prediction of breed composition in an admixed cattle population. *Anim. Genet.* 43, 696–703. doi: 10.1111/j.1365-2052.2012.02345.x

Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294x.2011.05163.x

Getachew, T., Huson, H. J., Wurzinger, M., and Burgstaller, J. (2017). Identifying highly informative genetic markers for quantification of ancestry proportions in crossbred sheep populations: implications for choosing optimum levels of admixture BMC. *Genet* 18:80. doi: 10.1186/s12863-017-0526-2

Halder, I., Shriver, M., Thomas, M., Fernandez, J. R., and Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum. Mutat.* 29, 648–658. doi: 10.1002/humu.20695

Hanotte, O. (2002). African Pastoralism: Genetic Imprints of Origins and Migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878

Hanotte, O., Dessie, T., and Kemp, S. (2010). Time to Tap Africa's Livestock Genomes. *Science* 328, 1640–1641. doi: 10.1126/science.1186254

Houessou, S. O., Dossa, L. H., Diogo, R. V. C., and Ahozonlin, M. C. (2019). Confronting pastoralists' knowledge of cattle breeds raised in the extensive production systems of Benin with multivariate analyses of morphological traits. *PLoS One* 14:e0222756. doi: 10.1371/journal.pone.0222756

Jin, W., Xu, S., Wang, H., and Yu, Y. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* 22, 519–527. doi: 10.1101/gr.124784.111

Judge, M. M., Kelleher, M. M., Kearney, J. F., Sleator, R. D., and Berry, D. P. (2017). Ultra-low-density genotype panels for breed assignment of Angus and Hereford cattle. *Animal* 11, 938–947. doi: 10.1017/S1751731116002457

Khayatzadeh, N., Mészáros, G., Utsunomiya, Y. T., and Garcia, J. F. (2016). Locus-specific ancestry to detect recent response to selection in admixed Swiss Fleckvieh cattle. *Anim. Genet.* 47, 637–646. doi: 10.1111/age.12470

Kim, E.-S., and Rothschild, M. F. (2014). Genomic adaptation of admixed dairy cattle in East Africa. *Front. Genet.* 5:443. doi: 10.3389/fgene.2014.00443

Ministere des ressources animales (2014). *Annuaire des statistiques d'élevage 2014 (Burkina Faso)*. Ouagadougou: MINISTERE DES RESSOURCES ANIMALES.

Naessens, J., Teale, A. J., and Sileghem, M. (2002). Identification of mechanisms of natural resistance to African trypanosomiasis in cattle. *Vet. Immunol. Immunopathol.* 87, 187–194. doi: 10.1016/s0165-2427(02)00070-3

Noyes, H., Brass, A., Obara, I., and Anderson, S. (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. *PNAS* 108, 9304–9309. doi: 10.1073/pnas.1013486108

Okeyo Mwai, A., Hanotte, O. H., Kwon, Y.-J., and Cho, S. (2015). African indigenous cattle: Unique genetic resources in a rapidly changing world. Asian-Australas. *J. Anim. Sci.* 2015:0002R. doi: 10.5713/ajas.15.0002R

Ouédraogo, D., Soudré, A., Ouédraogo-Koné, S., and Zoma, B. L. (2020). Breeding objectives and practices in three local cattle breed production systems in Burkina Faso with implication for the design of breeding programs. *J. Livsci.* 232:103910. doi: 10.1016/j.livsci.2019.103910

Pérez, O. A. M., Mészáros, G., Utsunomiya, Y. T., and Sonstegard, T. S. (2014). Linkage disequilibrium levels in Bos indicus and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. *J. Livsci.* 166, 121–132. doi: 10.1016/j.livsci.2014.05.007

Pitt, D., Bruford, M. W., Barbato, M., and Orozco-terWengel, P. (2018). Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evolut. Applicat.* 12, 105–122. doi: 10.1111/eva.12641

Purcell, S., Neale, B., Todd-Brown, K., and Thomas, L. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Med. Genet.* 81, 559–575. doi: 10.1086/519795

Sankararaman, S., Sridhar, S., Kimmel, G., and Halperin, E. (2008). Estimating Local Ancestry in Admixed Populations. *Am. J. Med. Genet.* 82, 290–303. doi: 10.1016/j.ajhg.2007.09.022

Smetko, A., Soudre, A., Silbermayr, K., and Müller, S. (2015). Trypanosomosis: potential driver of selection in African cattle. *Front. Genet.* 6:137. doi: 10.3389/fgene.2015.00137

Social Science Statistics (2021). *P Value from Pearson (R) Calculator*. London: Pearson.

Sölkner, J., Frkonja, A., Raadsma, H. W., and Jonas, E. (2010). Estimation of Individual Levels of Admixture in Crossbred Populations from SNP Chip Data: Examples with Sheep and Cattle Populations. *Interbull Bull.* 10, 62–66.

Tang, H., Choudhry, S., Mei, R., and Morgan, M. (2007). Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *Am. J. Med. Genet.* 81, 626–633. doi: 10.1086/520769

The R Development Core Team (2020). *A Language and Environment for Statistical Computing*. Geneva: R Core Team.

Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., et al. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* 31, 334–341. doi: 10.1093/nar/gkg115

Tijjani, A. (2019). *Genome diversity and adaptation of African Taurine and Zebu cattle*. nottingham: eprints.

Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome sequence analysis reveals selection signatures in endangered trypanotolerant West African muturu cattle. *Front. Genet.* 10:442. doi: 10.3389/fgene.2019.00442

Traoré, A., Koudandé, D. O., Fernández, I., and Soudré, A. (2015). Geographical assessment of body measurements and qualitative traits in West African cattle. *Trop. Anim. Health Prod.* 47, 1505–1513. doi: 10.1007/s11250-015-0891-7

Weir, B. S., and Cockerham, C. C., 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641

Wurzinger, M., Sölkner, J., and Iñiguez, L. (2011). Important aspects and limitations in considering community-based breeding programs for low-input smallholder livestock systems. *Small. Rumin. Res.* 98, 170–175. doi: 10.1016/j.smallrumres.2011.03.035

Yougbaré, B., Soudré, A., Ouédraogo, D., and Zoma, B. L. (2020). Morphometric characterization of purebred and crossbred Baoulé cattle in Burkina Faso. *Acta Agric. Scand. A Anim. Sci.* 0, 1–10. doi: 10.1080/09064702.2020.1825785

Zoma-Traoré, B., Soudré, A., Ouédraogo-Koné, S., and Khayatzadeh, N. (2020). From farmers to livestock keepers: A typology of cattle production systems in south-western Burkina Faso. *Trop. Anim. Health Prod.* 52, 2179–2189. doi: 10.1007/s11250-020-02241-6

Conflict of Interest: NK was employed by company SUISAG, Switzerland.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fgene.2021.723706

High Genetic Diversity but Absence of Population Structure in Local Chickens of Sri Lanka Inferred by Microsatellite Markers

*Amali Malshani Samaraweera[1], Ranga Liyanage[2], Mohamed Nawaz Ibrahim[3], Ally Mwai Okeyo[3], Jianlin Han[3,4] and Pradeepa Silva[5]**

[1] Department of Animal Science, Faculty of Animal Science and Export Agriculture, Uva Wellassa University, Badulla, Sri Lanka, [2] Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka, [3] International Livestock Research Institute (ILRI), Nairobi, Kenya, [4] CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China, [5] Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka

Local chicken populations belonging to five villages in two geographically separated provinces of Sri Lanka were analyzed using 20 microsatellite markers to determine the genetic diversity of local chickens. Population genetic parameters were estimated separately for five populations based on geographic locations and for eight populations based on phenotypes, such as naked neck, long legged, crested or crown, frizzle feathered, Giriraj, commercial layer, crossbreds, and non-descript chicken. The analysis revealed that there was a high genetic diversity among local chickens with high number of unique alleles, mean number of alleles per locus (MNA), and total number of alleles per locus per population. A total of 185 microsatellite alleles were detected in 192 samples, indicating a high allelic diversity. The MNA ranged from 8.10 (non-descript village chicken) to 3.50 (Giriraj) among phenotypes and from 7.30 (Tabbowa) to 6.50 (Labunoruwa) among village populations. In phenotypic groups, positive inbreeding coefficient (F_{IS}) values indicated the existence of population substructure with evidence of inbreeding. In commercial layers, a high expected heterozygosity He = 0.640 ± 0.042) and a negative F_{IS} were observed. The positive F_{IS} and high He estimates observed in village populations were due to the heterogeneity of samples, owing to free mating facilitated by communal feeding patterns. Highly admixed nature of phenotypes was explained as a result of rearing many phenotypes by households (58%) and interactions of chickens among neighboring households (53%). A weak substructure was evident due to the mating system, which disregarded the phenotypes. Based on genetic distances, crown chickens had the highest distance to other phenotypes, while the highest similarity was observed between non-descript village chickens and naked neck birds. The finding confirms the genetic wealth conserved within the populations as a result of the breeding system commonly practiced by chicken owners. Thus, the existing local chicken populations should be considered as a harbor of gene pool, which can be readily utilized in developing locally adapted and improved chicken breeds in the future.

Keywords: local chicken, microsatellite marker, genetic diversity, population structure, tropical climate

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman

Reviewed by:
Giri Athrey, Texas A&M University, United States
Coralie Danchin, Institut de l'Elevage, France

****Correspondence:***
Pradeepa Silva
pradeepas@pdn.ac.lk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 June 2021*
Accepted: *25 August 2021*
Published: *28 September 2021*

Citation:
Samaraweera AM, Liyanage R, Ibrahim MN, Okeyo AM, Han J and Silva P (2021) High Genetic Diversity but Absence of Population Structure in Local Chickens of Sri Lanka Inferred by Microsatellite Markers. Front. Genet. 12:723706. doi: 10.3389/fgene.2021.723706

INTRODUCTION

The choice of chickens by rural small-scale poultry farmers in Sri Lanka is often indigenous chickens for both egg and meat production, owing to numerous advantages they bring, such as disease resistance, high fertility, good maternal quality, longevity, ability to utilize poor-quality feeds, and most, importantly, the ability to manage them under a harsh environment condition with low level of management inputs despite their poor productivity compared with commercial layers and broilers (Silva et al., 2014). Moreover, high ash content in eggs of some phenotypes (non-descript village chicken, naked neck, long legged, and crown) and high fat content in egg yolk of local chickens in Sri Lanka were reported in a study by Sanjeewa et al. (2011). In the same study, the internal egg quality was found to be preserved in local chicken eggs during storage compared with commercial chicken eggs (Sanjeewa et al., 2011). In Korean native chicken, considerable amount of health-promoting compounds, such as carnosine, anserine, betaine, and carnitine, were identified compared with commercial broilers (Jayasena et al., 2015). Therefore, there is a special preference for native chicken meat due to its unique taste and texture, especially after cooking (Wattanachant et al., 2004).

However, such valuable indigenous genetic pool with undiscovered genetic potential has been eroding due to various reasons, especially in developing counties. Furthermore, some of the poultry breeds in Sri Lanka have already been lost or are at the risk of extinction (Punyawardena, 2010). Conservation of genetic resources is important from a global perspective as genetic variability underlying the adaptability and potential of animal genetic resources is essential to meet the changes in the preference and demand of consumers and to diminish the challenges posed by climate change and emerging diseases (FAO, 2007). Hence, it is frequently highlighted that the characterization of animal genetic resources is an essential initiative for sustainable utilization of animal genetic resources (AnGR). The Global Strategy for the Management of Farm Animal Genetic Resources coordinated by the Food and Agricultural Organization (FAO) aims to identify and propose sustainable genetic improvement plan for indigenous AnGR. Characterizing AnGR both phenotypically and genotypically is essential to understand and describe it properly, and then to propose a rational action plan for sustainable utilization. Accordingly, the attempts made on phenotypic characterization of indigenous chickens in Sri Lanka identified that the populations consisted of diverse phenotypes (Bett et al., 2014), while Liyanage et al. (2015) described seven distinct phenotypic groups, including naked neck, long legged, crest/crown, Giriraj, commercial crosses, frizzle feathered, and the non-descript group of multiple crosses of other groups.

So far, a large number of studies have been conducted to characterize the chicken populations in Asia and Africa using microsatellite markers due to their highly polymorphic nature and abundance throughout the entire genome (Berthouly et al., 2009; Bodzsar et al., 2009; Cuc et al., 2010; Eltanany et al., 2011; Leroy et al., 2012). With recent advances in DNA sequencing, single-nucleotide polymorphisms (SNPs) have been used extensively to characterize the genetic diversity and animal identification systems for closely related species/breeds/types (Samaraweera et al., 2011). However, since SNPs are biallelic and less informative, around 1.7–5.56 SNPs were needed to achieve the same information content as that obtained with one microsatellite marker (Gärke et al., 2012). Moreover, if short tandem repeats (STRs) have been selected based on a high minor allele frequency in a restricted number of breeds, this may underestimate the diversity in other breeds (Lenstra et al., 2012). Therefore, the highly polymorphic microsatellite markers are still valid for assessing the genetic diversity of AnGR.

In this context, this study was conducted to determine the genetic diversity of local chicken populations in two selected areas and among identified phenotypes in Sri Lanka using microsatellite markers.

MATERIALS AND METHODS

Selection of Sampling Locations and Households

Two veterinary divisions, Thirappane (80.5039–80.6331 E, 8.1185–8.2202 N) in the North Central Province and Karuwalagaswewa (79.5395–80.5042 E, 8.0047–8.0692 N) in the North Western Province were selected for sample collection based on the highest density of local chickens in Sri Lanka (**Figure 1**). According to the distribution of farms, three villages of the Thirappane site, namely, Dematagama (DM), Labunoruwa (LA), and Ooththupitiya (OT), and two villages in the Karuwalagaswewa site, namely, Tabbowa (TB) and Thewanuwara (TH), were sampled.

Sample Collection

Altogether 818 blood samples of chickens older than 6 months of age were collected on the Whatman FTA® filter paper (Whatman Bio-Science, Maidstone, United Kingdom) and stored at room temperature until it was used for the analysis (i.e., 216, 69, 67, 219, and 247 from DM, LA, OT, TB, and TH, respectively). Among 818 samples, 192 samples were purposively selected for genotyping based on locations, households, and phenotypes (**Table 1**). Ethical permission for the project was obtained from the Institutional Research and Ethics Committee (IREC) and the Institutional Animal Care and Use Committee (IACUC) of the International Livestock Research Institute (ILRI), Nairobi, Kenya.

DNA extraction and genotyping using microsatellite markers were carried out at the CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, CAAS, Beijing, China.

DNA Extraction, PCR Amplification, and Genotyping

DNA extraction was carried out from blood collected on the FTA filter paper, and around 15 discs from each FTA filter paper were added to 100 μl of distilled water and then heated at 90°C for 10 min. The resulting solution was used for PCR amplification with 20 microsatellite markers recommended by the International Society of Animal Genetics

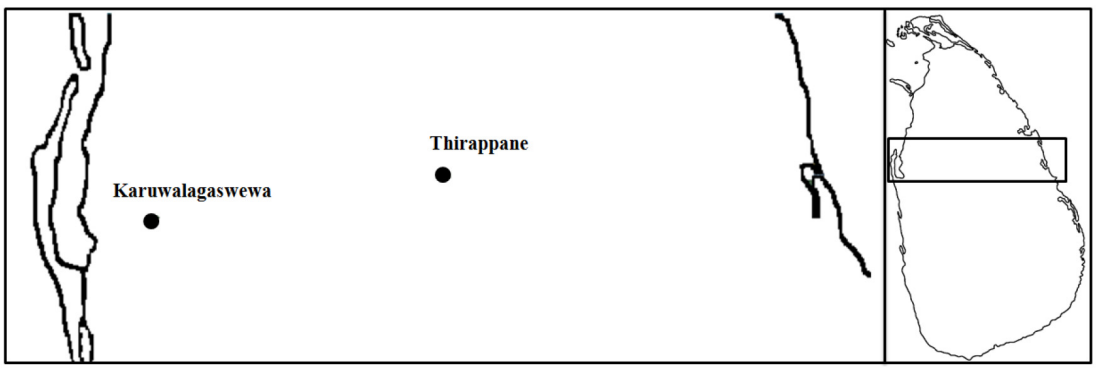

FIGURE 1 | The Sri Lankan map showing the two sampling locations, Thirappane and Karuwalagaswewa.

TABLE 1 | The number of village chicken ecotypes selected from each household of the five villages in two sites (Thirappane and Karuwalagaswewa) for genotyping analysis.

Site	Village	Village chicken (VC)	Naked neck (NN)	Long leg (LL)	Crown (CC)	Frizzled feathered (FF)	Giriraj (GR)	Commercial layers (CL)	Cross (CR)	Total number of birds from each village
Thirappane (Site 1)	Dematagama (DM)	10	10	6	4	1	0	6	0	37
	Labunoruwa (LA)	19	8	0	2	3	5	2	0	39
	Ooththupitiya (OT)	20	17	3	0	0	0	0	0	40
Karuwalagaswewa (Site 2)	Tabbowa (TB)	12	13	12	0	1	0	0	1	39
	Thewanuwara (TH)	13	7	10	0	1	0	0	6	37
Total		74	55	31	6	6	5	8	7	192

(ISAG)–FAO Advisory Group on Animal Genetic Diversity (**Table 2**; FAO, 2011). The PCR amplicons were separated by size using the 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, United States). Sizing and genotyping of the alleles were carried out using the GeneMapper 3.7 software (Applied Biosystems).

Data Analysis

The total number of samples used in the analysis was divided into five populations according to the geographical locations (villages) of sampling as DM, LA, OT, TB and TH (**Supplementary Table 1**), and to eight phenotypic groups; Naked Neck (NN), Long Legged (LL), Crested or Crown (CC), Frizzle Feathered (FF) Giriraj (GR), Commercial Layer (CL), Crossbreds (CR) and non-descript village chicken (VC) (**Figure 2**, **Supplementary Table 2**).

Deviations from Hardy–Weinberg equilibrium (HWE) for all locus-population combinations and linkage disequilibria (LD) between all pairs of loci were determined using the program Genepop version 4.1.3 (Raymond, 1995). HWE was assessed based on the Markov chain algorithm to estimate the unbiased exact *p*-value of the test (Guo and Thompson, 1992) for each locus in each population. Fisher's exact test was used to test for LD with unbiased *p*-values derived from the Markov chain randomization method.

Allelic frequency, mean number of alleles per locus (MNA), observed heterozygosity (Ho), and expected heterozygosity (He) were calculated to quantify the genetic variation within populations using Microsatellite Toolkit version 3.1 (Park, 2001) and FSTAT version 2.9.3.2. MNA is the average number of alleles observed in a population over all loci genotyped. F-statistics for each locus (Weir and Cockerham, 1984) was calculated and tested using the FSTAT.

Population structure and the level of admixture of individuals were analyzed using the Structure version 2.3.4 (Pritchard et al., 2000). Population structure was analyzed assuming the most probable number of the clusters (K) values at 1–6 with 30 simulations for each K, assuming an admixture model with correlated allele frequencies using a burn-in of 100,000 followed by 100,000 Markov Chain Monte Carlo iterations to check the level of admixture and substructuring of the populations. The most probable number of the clusters (K) was selected based on the highest averaged log probability of data [Ln Pr(X/K)] with the lowest variation {Var[Ln Pr(X/K)]} among the 30 runs. *Ad hoc*

TABLE 2 | Primer name, primer sequence, allele size, and annealing temperature of 20 microsatellite markers used for the analysis.

	Name	Primer sequence (5′ - > 3′) forward reverse	Label	Allele range (bp)	Annealing temperature (°C)
1	LEI0094	GATCTCACCAGTATGAGCTGC TCTCACACTGTAACACAGTGC	FAM	240–300	58.5
2	MCW0069	GCACTCGAGAAAACTTCCTGCG ATTGCTTCAGCAAGCATGGGAGGA	HEX	150–180	
3	ADL0268	CTCCACCCCTCTCAGAACTA CAACTTCCCATCTACCTACT	FAM	100–120	53
4	MCW0034	TGCACGCACTTACATACTTAGAGA TGTCCTTCCAATTACATTCATGGG	HEX	210–250	
5	LEI0166	CTCCTGCCCTTAGCTACGCA TATCCCCTGGCTGGGAGTTT	FAM	340–380	60
6	MCW0248	GTTGTTCAAAAGAAGATGCATG TTGCATTAACTGGGCACTTTC	FAM	210–230	
7	MCW0216	GGGTTTTACAGGATGGGACG AGTTTCACTCCCAGGGCTCG	FAM	135–150	
8	LEI0234	ATGCATCAGATTGGTATTCAA CGTGGCTGTGAACAAATATG	HEX	210–370	54.3
9	ADL0278	CCAGCAGTCTACCTTCCTAT TGTCATCCAAGAACAGTGTG	HEX	110–130	
10	MCW0222	GCAGTTACATTGAAATGATTCC TTCTCAAAACACCTAGAAGAC	FAM	200–230	55
11	MCW0016	ATGGCGCAGAAGGCAAAGCGATAT TGGCTTCTGAAGCAGTTGCTATGG	HEX	135–160	
12	MCW0295	ATCACTACAGAACACCCTCTC TATGTATGCACGCAGATATCC	FAM	85–110	54.3
13	MCW0037	ACCGGTGCCATCAATTACCTATTA GAAAGCTCACATGACACTGCGAAA	FAM	150–160	
14	MCW0206	CTTGACAGTGATGCATTAAATG ACATCTAGAATTGACTGTTCAC	FAM	220–250	58.6
15	MCW0111	GCTCCATGTGAAGTGGTTTA ATGTCCACTTGTCAATGATG	FAM	95–115	
16	MCW0067	GCACTACTGTGTGCTGCAGTTT GAGATGTAGTTGCCACATTCCGAC	HEX	150∼200	56
17	MCW0183	ATCCCAGTGTCGAGTATCCGA TGAGATTTACTGGAGCCTGCC	FAM	290–330	
18	MCW0014	TATTGGCTCTAGGAACTGTC GAAATGAAGGTAAGACTAGC	FAM	160–200	54
19	MCW0330	TGGACCTCATCAGTCTGACAG AATGTTCTCATAGAGTTCCTGC	HEX	250–300	
20	MCW0081	GTTGCTGAGAGCCTGGTGCAG CCTGTATGTGGAATTACTTCTC	FAM	100∼120	54.6

FIGURE 2 | Chicken phenotypes in Sri Lanka. **(A)** Naked neck; **(B)** long legged; **(C)** crested or crown; **(D)** Giriraj; **(E)** non-descript chicken; **(F)** frizzle feathered; **(G)** commercial layer.

quantity (ΔK) based on the rate of change in the log probability of data between successive K values was calculated to verify the best K using the web-based STRUCTURE HARVESTER (Earl, 2012). Estimated cluster membership coefficient matrices of multiple runs for each K were permuted using the CLUMPP (Jakobsson and Rosenberg, 2007) to obtain close match in all simulations with the Greedy algorithm at 1,000 repeats of random inputs of the data and the pairwise matrix similarity statistic to be G′. Then the aligned membership coefficients were displayed using Microsoft Office Excel 2007. Principal component analysis (PCA) was performed using the SPSS version 13.0[1] with arcsine-transformed allele frequencies. Factorial correspondence analysis (FCA) was performed using GENETIX 4.05 (Belkhir, 2004).

An unweighted pair-group method using arithmetic average (UPGMA) dendrogram between populations was constructed from Nei's standard genetic distances (D_S; Nei, 1987) using

[1]https://spss.software.informer.com/13.0/

DISPAN (Ota, 1993) with 1,000 bootstrap values. Analysis of molecular variance (AMOVA), pairwise F_{ST} (Reynolds' genetics distance), and correlation between distance matrices (Mantel test with 10,000 permutations) were computed using Arlequin version 3.5.2.2 (Excoffier et al., 2005). The Splits Tree4 version 4.14.2 was used to draw a NeighborNet tree based on pairwise F_{ST} values (Huson and Bryant, 2006).

RESULTS

The present study assessed the genetic diversity and genetic structure within Sri Lankan local chicken populations, which have been defined using two different boundaries of geographical boundaries (five villages) and phenotype-based boundaries (eight phenotypes). The results are also presented accordingly.

Hardy–Weinberg Equilibrium and Linkage Disequilibria in Different Local Chicken Populations

It was observed that among 20 loci across five geographic populations used in the analysis, several loci in individual populations (4 from DM, 8 from LA, 10 from OT and 6 each from TB and TH) were deviated from HWE ($p < 0.05$) while no pair of loci rejected LD ($p > 0.05$).

Genetic Diversity

Village Populations

From 818 samples, more than 15% of birds in each of the five villages were selected for the genotyping analysis. Accordingly, 192 samples across five geographic populations were genotyped, and a total of 185 microsatellite alleles were detected at 20 loci, indicating a considerably rich allelic diversity. The number of alleles across loci within populations ranged from 130 alleles in LA to 146 alleles in TB. The number of alleles per locus per population ranged from 3 (MCW0248 in DM; MCW0014 in LA) to 16 (LEI0234 in DM). Moreover, 27 private alleles, which were unique to one population, were found among five populations, e.g., 6 in DM, 4 in LA, 6 in OT, 9 in TB, and 2 in TH. Furthermore, both the mean frequency of individual private alleles across the five populations (1.97%) and the population-specific frequency of the private alleles (1.3%) were very low, including a high number of migrants per population (8.33), similar to what was reported by Barton and Slatkin (1986).

Estimates of F_{IS} indicates important properties of the mating system within populations (Holsinger and Weir, 2009). Accordingly, a deficit of heterozygotes (positive F_{IS} values; **Tables 3**, **4**) was observed for all populations, showing certain levels of inbreeding.

The MNA per locus for populations ranged from 6.5 in LA to 7.3 in TB. Therefore, a high MNA per locus was observed, indicating a high level of polymorphisms. MNA is also a sensitive measure of genetic variability in comparison with heterozygosity measures. Ho ranged from 0.594 in LA to 0.689 in DM, while He ranged from 0.678 in LA to 0.715 in DM.

TABLE 3 | Mean number of alleles per locus (MNA), expected heterozygosity (He), observed heterozygosity (Ho), and average inbreeding coefficient (F_{IS}) estimated from the 20 microsatellite loci for each village population.

Village	n	MNA	He	Ho	F_{IS}
DM	37	7.10 ± 3.35	0.715 ± 0.023	0.689 ± 0.017	0.036
LA	39	6.50 ± 2.48	0.678 ± 0.027	0.594 ± 0.018	0.126
OT	40	7.05 ± 2.65	0.709 ± 0.024	0.628 ± 0.017	0.116
TB	39	7.30 ± 2.74	0.715 ± 0.022	0.653 ± 0.017	0.088
TH	37	7.10 ± 2.53	0.706 ± 0.020	0.635 ± 0.018	0.102

TABLE 4 | Mean number of alleles per locus (MNA), expected heterozygosity (He), observed heterozygosity (Ho), and average inbreeding coefficient (F_{IS}) estimated from the 20 microsatellite loci for each phenotype.

Phenotype	n	MNA	He	Ho	F_{IS}
VC	74	8.10 ± 2.94	0.714 ± 0.021	0.626 ± 0.013	0.123
NN	55	7.80 ± 3.27	0.716 ± 0.023	0.646 ± 0.014	0.099
LL	31	6.75 ± 2.55	0.722 ± 0.025	0.686 ± 0.019	0.051
GR	5	3.50 ± 1.36	0.611 ± 0.044	0.580 ± 0.049	0.051
FF	6	3.90 ± 1.29	0.656 ± 0.041	0.650 ± 0.043	0.009
CR	7	4.40 ± 1.57	0.662 ± 0.040	0.557 ± 0.042	0.158
CL	8	4.30 ± 1.26	0.640 ± 0.042	0.650 ± 0.038	−0.016
CC	6	3.70 ± 1.26	0.652 ± 0.035	0.625 ± 0.044	0.041

Phenotype-Based Populations

From 818 samples, more than 12% of birds from each of the eight phenotypes were selected for the genotyping analysis. The number of alleles across loci in the eight phenotypes ranged from 162 in CR to 70 in FF. A total of 29 private alleles were detected in VC (12), NN (13), LL (3), and CL (1). The MNA per locus per phenotype ranged from 8.1 in VC to 3.5 in GR (**Table 4**).

Genetic Differentiation

The genetic relationships between populations were determined using Reynolds' genetic distances assuming that the genetic differentiation occurs solely due to genetic drift (Reynolds et al., 1983).

Village Populations

The two villages of TB and TH at Karuwalagaswewa clustered into one group separating other three villages (DM, LA, and OT) at Thirappane. The largest (0.0240) and the smallest (0.0090) pairwise genetic distances were observed between TH and OT and between TH and TB, respectively (**Table 5**). Furthermore, **Figure 3** shows the unrooted UPGMA dendrogram summarizing the Nei's standard genetic distances between villages, confirming the differentiation in local chickens between the two veterinary divisions with a high bootstrap value at 88%. The result of AMOVA for villages is given in **Table 6**, in which most of the genetic variability based on the 20 microsatellite markers (89%) was found to be present among alleles within individuals, followed by the one observed among individuals within the villages (9%).

TABLE 5 | Pairwise F_{ST} estimates between the five chicken populations based on the 20 microsatellite loci.

Populations	DM	OT	LA	TB	TH
DM					
OT	0.0161				
LA	0.0184	0.0219			
TB	0.0128	0.0123	0.0235		
TH	0.0160	0.0240	0.0216	0.0090	

Phenotype-Based Populations

The highest genetic distance (Reynolds' genetic distances in **Table 7**) was observed between GR and CC (0.0914), but the lowest was between the NN and VC (0.0023). The relationship between the eight phenotypes is presented in **Figure 4** as a NeighborNet tree derived from pairwise F_{ST} estimates. Two main clusters were identified, where VC, NN, LL, and FF, were clustering together, while CL, GR, and CR formed a separate cluster, leaving CC as a unique phenotype. Similarly, the highest genetic distance in CC to other phenotypes was found in NeighborNet tree (**Figure 4**). The result of AMOVA for phenotypes is given in **Table 8**, in which the highest genetic variability (89%) was found among alleles within individuals, followed by that present among individuals within the phenotypes.

Furthermore, a high genetic similarity between NN and VC was further confirmed by the population structure analysis (**Figure 5**). The structure analysis indicated that VC, NN, LL, CC, and FF shared a higher proportion of genotypes in common compared with GR, CL, and CR.

Population Structure

Graphic displays of the estimated membership coefficients of each individual to each population based on the phenotypes and based on villages at $2 \leq K \geq 4$ are given in **Figures 5**, **6**, respectively. The STRUCTURE HARVESTER software was used to graphically illustrate the mean estimates of log probability of data and ΔK to select the best K value (**Figure 7**). Accordingly, the rate of change in the log probability of data between successive K values (ΔK) against each K indicated that a higher ΔK value at $K = 2$. Therefore, $K = 2$ was selected as the most probable number of the clusters to reveal the population structure in both situations. Structure analysis indicated the absence of population genetic structure among local chicken populations in Sri Lanka.

Compared with the phylogenetic reconstruction methods, PCA provided a better understanding of the genetic relationship among the local chicken populations, precisely the level of admixture (**Figure 8**). Similar to the structure result, the FCA analysis also indicated the absence of population structure (**Figure 9**).

DISCUSSION

The present study investigated the genetic diversity and population genetic structure of two categories of indigenous chicken populations defined by geographical boundaries and phenotypic boundaries in two geographically distant regions of Sri Lanka.

The result showed that some loci were deviated from HWE, probably due to the deficiencies of heterozygotes. Several factors, like the within-population fragmentation created by clusters of households, which were involved in our stratified sampling (Samaraweera et al., 2014), could bring the Wahlund effect, leading to an overall deficiency of heterozygotes. The inbreeding observed in all five village chicken populations (**Table 3**) could have also contributed to such deviations even though there was no significant LD between all pairs of the 20 loci across the five chicken populations.

Genetic Diversity Within the Populations

The total number of alleles per locus per population, the number of private alleles, and MNA per locus across the populations all indicated a high level of genetic diversity within the indigenous chicken populations of Sri Lanka. Furthermore, the highest MNA was exhibited by VC, NN, and LL, followed by CL, CR, FF, CC, and GR. Comparatively, low MNAs were recorded from other local chicken populations in the world, for example, Vietnamese local chickens (5.1) (Pham et al., 2013), Hungarian indigenous chicken breeds (2.9–4.2) (Bodzsar et al., 2009), French local breeds (6.6) (Berthouly et al., 2008), and Egyptian strains (4.92) (Eltanany et al., 2011). The higher number of alleles in VC, NN, and LL was due to the presence of private alleles that occurred at low frequencies within the populations and also due to the high number of observed alleles, owing to the free range management system, which allowed mixing of chickens among neighboring households (Bett et al., 2014).

All the chicken phenotypes from Sri Lanka had a high He, ranging from 0.61 to 0.72, which were similar to those of

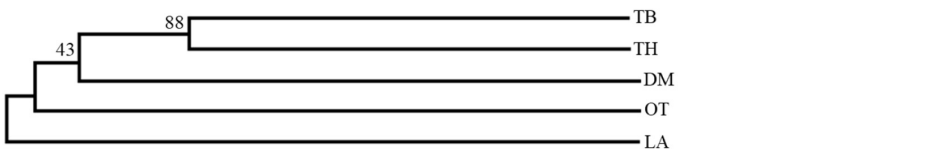

FIGURE 3 | The unweighted pair-group method using arithmetic averages (UPGMA) dendrogram summarizing the standard genetic distances (*Ds*) between village-based populations. The numbers at the nodes are the percentage bootstrap values after 1,000 replications over loci.

TABLE 6 | Analysis of molecular variance (AMOVA) based on 20 microsatellite loci of five populations (villages) assigned into two groups (two sites).

Sources of variation	df	Percentage of variation
Between sites	1	0.34
Between villages, within site	3	1.43
Between individuals, within village	187	9.21
Within individuals	192	89.02

TABLE 7 | Pairwise F_{ST} estimates between the eight chicken phenotypes based on the 20 microsatellite loci.

	VC	NN	LL	CC	FF	GR	CL	CR
VC								
NN	0.0023							
LL	0.0115	0.0121						
CC	0.0280	0.0263	0.0488					
FF	0.0175	0.0301	0.0323	0.0842				
GR	0.0446	0.0364	0.0475	0.0914	0.0843			
CL	0.0326	0.0330	0.0496	0.0602	0.0570	0.0401		
CR	0.0173	0.0200	0.0280	0.0663	0.0471	0.0171	0.0244	

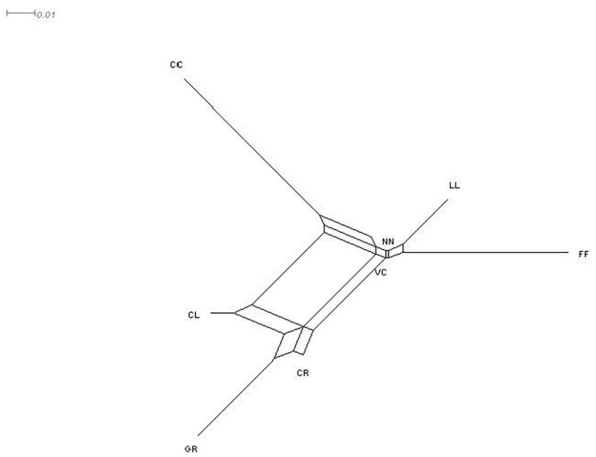

FIGURE 4 | NeighborNet tree summarizing the genetic distances between chicken phenotypes. NN, naked neck; LL, long legged; CC, crested or crown; FF, frizzle feathered; GR, Giriraj, CL, commercial layer; CR, crossbreds; VC, non-descript chicken.

TABLE 8 | Analysis of molecular variance (AMOVA) based on the 20 microsatellite loci of eight populations (phenotypes).

Sources of variation	df	Percentage of variation
Between phenotypes	7	1.5
Between individuals, within phenotype	184	9.4
Within individuals	192	89.1

African and Asian scavenging chicken populations (**Table 4**). For example, the He of African scavenging chicken ranged from 0.53 to 0.66 (Leroy et al., 2012), of Vietnamese local chickens from 0.50 to 0.63 (Pham et al., 2013), and of traditional French breeds from 0.43 to 0.62 (Berthouly et al., 2008). In a study by Leroy et al. (2012), low He were reported for commercial lines compared with scavenging chicken populations, ranging from 0.29 to 0.48. However, a high He (0.639 ± 0.042) but a negative F_{IS} value (-0.016) were observed in CL (**Table 4**). Except in CL, positive F_{IS} values were present across all the village chicken populations, indicating non-random mating or the existence of population substructures with evidence of inbreeding within the populations. Comparatively, the high He and outbreeding of CL population could be due to the introduction of commercial strains to village farmers through interventions by the government and non-government organizations, aiming to uplift the chicken production for rural livelihoods.

Positive F_{IS} and high He estimates present in all village chicken populations can be due to the heterogeneity of the samples since they consisted of a mixture of phenotypes, which could be considered as genetic subdivision within the villages (e.g., the Wahlund effect) and non-random mating (Hedrick, 2013). Such subdivision may be explained by the observation of diverse phenotypes within the villages but specific chicken phenotypes often owned by different households. The 192 samples genotyped were selected from 818 samples representing all the phenotypes from the selected households where around 58% of households in these villages reared chickens with two or more phenotypes. Moreover, irrespective of the way of defining populations, i.e., based on geographical boundaries or based on phenotypic boundaries, a pattern of substructuring was also observed among the households, which is described in detail under the population structure section.

In addition, the high heterozygosity observed in this study indicated the individual variation within populations as a measure of allelic diversity. According to Nei (1987), heterozygosity is hardly affected by infrequent alleles at multi-allele loci. Therefore, the high heterozygosity observed in this study cannot be readily explained by the infrequent private alleles, which ranged from 2 to 9 in five villages and 1–2 in eight phenotypes. Every household had more than one phenotypic representation of birds included in our sampling. Thus, each population was a mixture of phenotypes with considerable allelic frequency. Nevertheless, the management system of free ranging supported a maximum and long run interaction among chickens from the neighboring households (Silva et al., 2014).

However, the expected and observed heterozygosity values were not significantly different ($p > 0.05$) in the five populations or in the eight phenotypes, suggesting a non-selective mating regime practiced among the Sri Lankan local chickens.

Genetic Distance Between the Populations

The estimates of genetic distances clearly matched with the geographical distances among different village chicken populations. For example, village populations of TH and TB, which were located only around 43 km far apart showed a low genetic distance supported by a high bootstrap value (88%) compared with village populations of TH and OT, which were located around 120 km far apart and separated by the highest

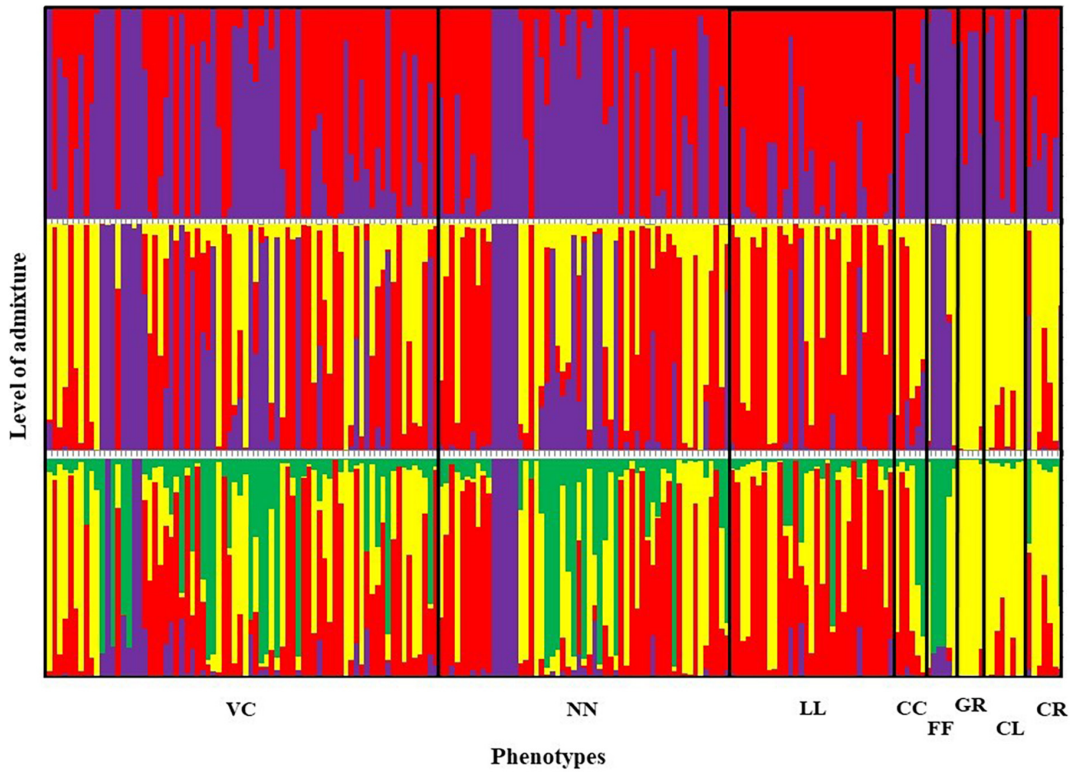

FIGURE 5 | Summary bar plot of estimates of Q (the estimated membership coefficient for each individual in each cluster). Each individual is represented by a single vertical line broken into *K* colored segments with lengths proportional to *K* inferred clusters (best *K* = 2). Chicken phenotypes: NN, naked neck; LL, long legged; CC, crested or crown; FF, frizzle feathered; GR, Giriraj, CL, commercial layer; CR, crossbreds; VC, non-descript chicken.

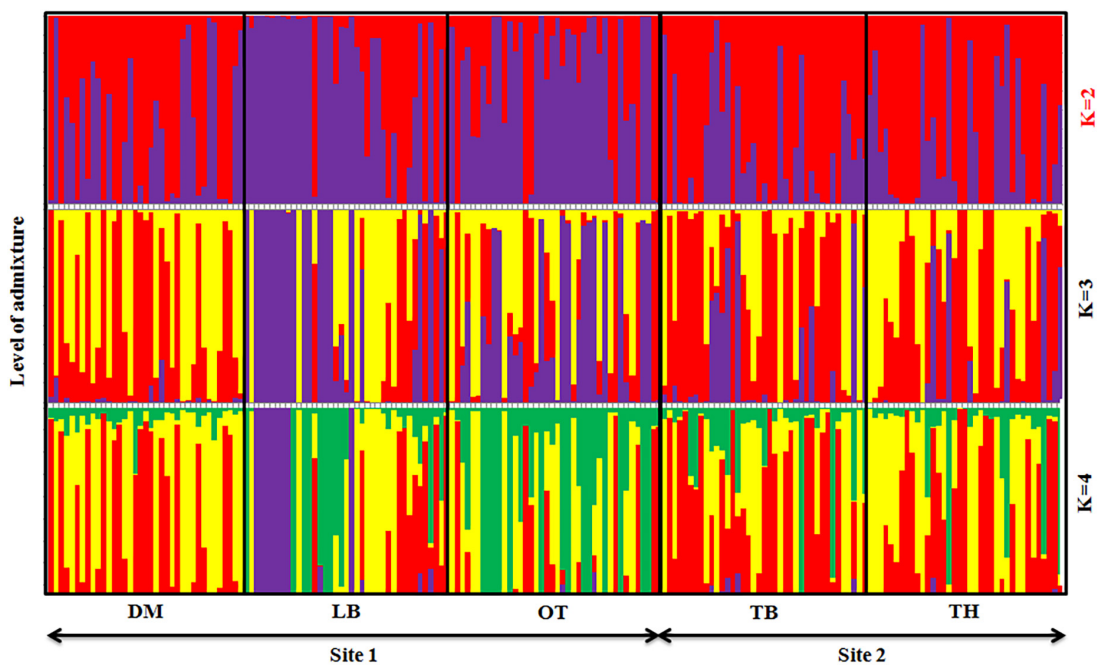

FIGURE 6 | Summary bar plot of estimates of Q (the estimated membership coefficient for each individual in each cluster). Each individual is represented by a single vertical line broken into *K* colored segments with lengths proportional to *K* inferred clusters (best *K* = 2).

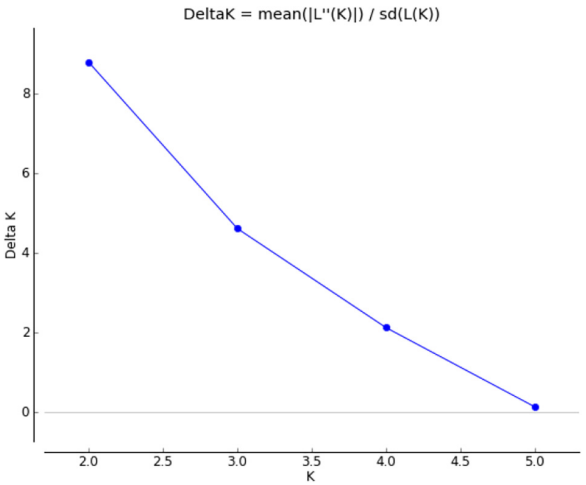

FIGURE 7 | Graphical illustration of delta *K*.

genetic distance (**Figure 3**). However, DM and TB village populations were from the two sites; their genetic distance were lower than the genetic distances between the villages of the same site (i.e., between DM and LA or OT; **Table 5**). Based on the estimates of phenotype-based populations, the NeighborNet tree (**Figure 4**) showed that the birds influenced by exotic chickens clustered together (CL, CR, and GR), while the crown chicken was separated further from all other phenotypic groups. Given the fact that crown (crest) is controlled by a single gene with incomplete dominant mode of inheritance, thus, the preference and selection of this phenotype could have similar influence on its separation as in the case of NN phenotypic group, where naked neck is also controlled by the same mode of inheritance. However, as depicted in the present study, the separation of CC from the remaining phenotypic groups could have been contributed by the crown characteristic as well as the association of crown gene with several other characteristics (Wang et al., 2012), for example, some modifier genes associated with different genetic architecture compared with the other phenotypic groups. However, it is interesting to note that the crown chicken found in the backyard system in Sri Lanka is different from the well-known Polish chicken, which is characterized by a "v"-shaped comb.

Population Structure and Level of Admixture Among the Populations

The graphical illustration based on structure analysis of chicken phenotypes did not show a distinguished population genetic structure, and the birds shared a highly admixed genetic background (**Figure 5**). This was also noticeable with the phenotypic diversity of these local chickens, where no distinguished color or comb pattern was observed (Liyanage et al., 2015). Moreover, a study by Silva et al. (2009) using mitochondrial DNA also revealed a similar finding within local chickens of Sri Lanka. However, the phenotypes of GR, CL, and CR tended to share a similar pattern of membership coefficient at $K = 3$ and 4, as represented by yellow color, where most of them had an exotic genetic background. Interestingly, several birds in naked neck phenotype showed the same clustering pattern along different *K* values, showing some substructure as illustrated in **Figure 5**. It was noticed that these NN birds were from the same farmer. To see a possible population substructuring pattern within the households, households with more than five birds in the sample were identified, and as given in **Figure 10**, several households had clear substructure within their chicken flocks though they had heterogeneous phenotypes. Therefore, a weak subclustering pattern was evident due to the mating system practiced following the restriction of limited or no exchange of breeding birds among the farmers.

Similarly, a high level of admixture was observed among the five geographical populations too (**Figure 6**), though LA showed a higher level of admixture of commercial strains. DM, TB, and TH villages also indicated a similar level of admixture among populations. The population genetic structure analysis suggested that the geographical isolation or closeness has influenced such similarities or differences in genetic structure within the village populations, for example, LA and OT villages in one site located close to each other compared with DM village. At the best *K* ($K = 2$), a similar population structure was observed in the two villages of LA and OT but leaving DM separated.

Similar results were also found in AMOVA where most variability was found at intra-individual level (89%) both within the populations defined by geographical boundaries and phenotypic boundaries. In addition, the observed genetic distances between the populations also confirmed the argument.

Implications for Conservation and Upgrading in the Future

Local chickens of Sri Lanka do not have a distinguished population genetic structure, for example, a specific breed. However, a weak clustering was evident with the households due to the mating system practiced. One of the earlier studies by Gunaratne et al. (1993) has reported that, though backyard chickens were reared under the free range management system, there were limited or no exchange of breeding birds among farms. This could maximize the inbreeding within flocks of individual households, thus, leading to the formation of some substructure among the flocks.

Nevertheless, the local chickens possess a high genetic diversity in terms of total number of alleles and number of private alleles due to free range management coupled with mixed rearing of different phenotypes. According to our recent studies based on mitochondrial DNA and whole genome re-sequencing data generated from a worldwide sampling of domestic chickens, in all five wild Red Jungle Fowl (*Gallus gallus*) subspecies and other three wild Jungle Fowl species, it is evident that the backyard chicken populations of Sri Lanka rooted back to the Red Jungle Fowl that is not inhabitant in Sri Lanka, but not to the endemic species of Ceylon Jungle Fowl (*Gallus lafayetti*) of the country (Silva et al., 2009; Wang et al., 2020); therefore, the genetic diversity observed in Sri Lanka local chicken populations was invariably the result of contributions from different introductions in the past. It is known that Sri Lanka has been exposed

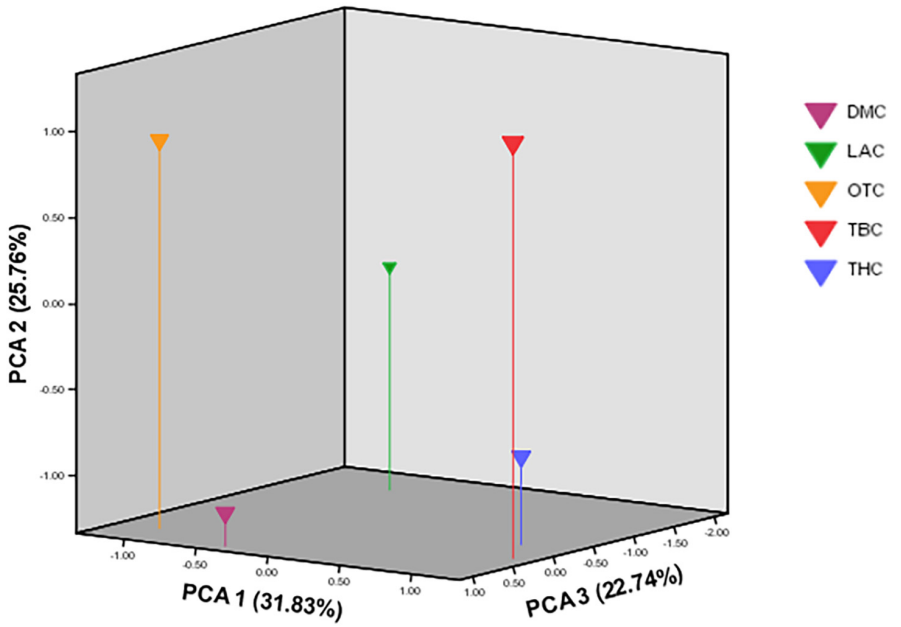

FIGURE 8 | Scatter plot showing the first three principal components over five village-based populations analyzed as revealed by principal component analysis (PCA) implemented in MVSP (Populations: DM, Dematagama; LA, Labunoruwa; OT, Ooththupitiya; TB, Tabbowa; TH, Thewanuwara).

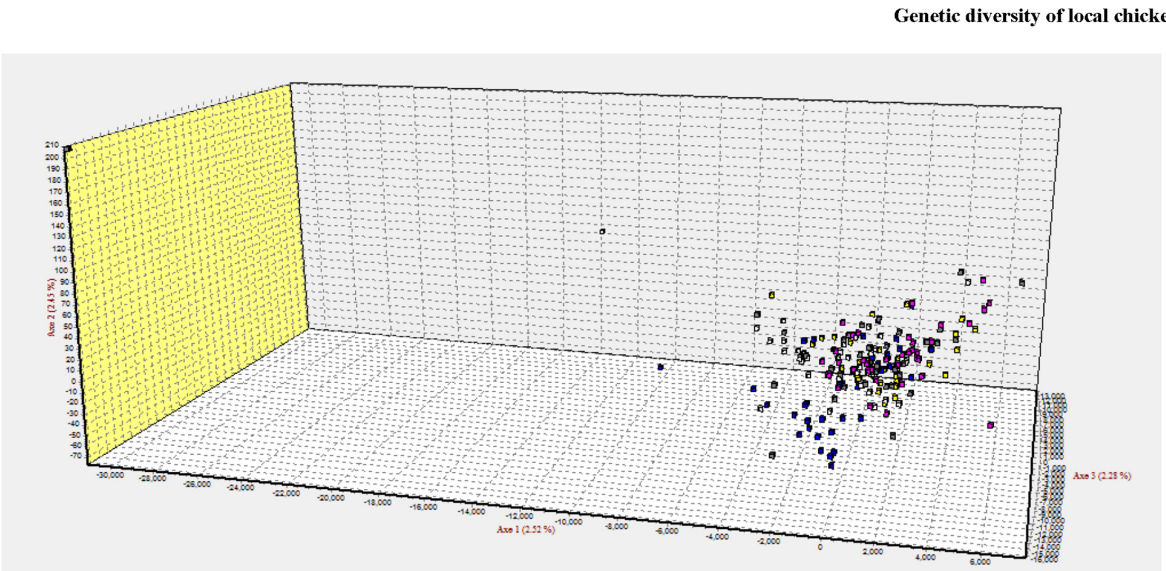

FIGURE 9 | 3D clustering patterns of all individuals analyzed using 20 microsatellite markers as revealed by factorial correspondence analysis (FCA) implemented in GENETIX.

to a variety of domestic animal species transported through trading in the past, owing to its critical location in the middle of the Indian Ocean connecting the sea routes between the East and the West. Thus, domestic chickens would have been a fair commodity for trade as well as a good protein source during long sea journeys since centuries back. Therefore, the evolutionary process of backyard chicken populations of the country has a long historical mixing of different chicken populations originating from several continents that may have contributed to both the high genetic and phenotypic diversity observed in this study.

The naked neck phenotype in study populations was superior in body weight, body circumference, keel length, and drum length (Bett et al., 2014). Furthermore, both naked neck and frizzle feathered genes were known to account for heat tolerance (Yunis and Cahaner, 1999). Nevertheless, farmers preferred the

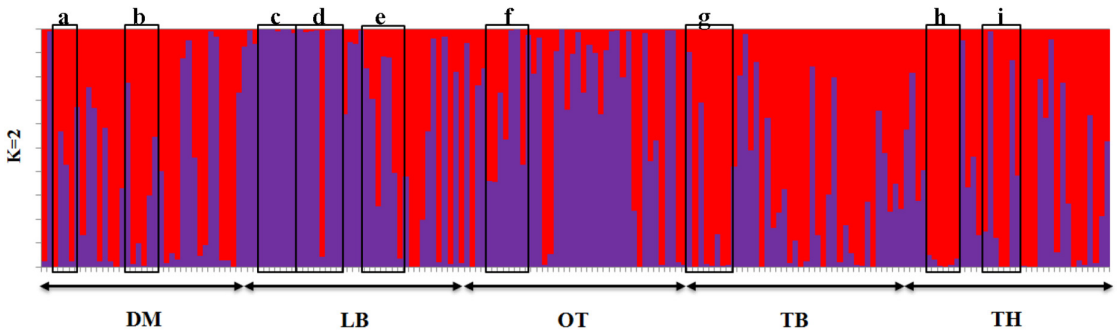

FIGURE 10 | Households having more than five birds in the sampled population (from a–i). Phenotypes in (c) nacked neck and village chicken; (d) village chicken, frizzle feathered, commercial layer; (h) village chicken, naked neck, long legged, cross breds.

NN phenotype due to their higher carcass weight and higher egg production. As reported by Abeykoon et al. (2014), farmers in the study sites expressed their willingness to pay more for frizzle feathered phenotype, followed by crested chicken, thus, indicating their preference for local chickens over commercial strains. However, it is strange to note that, though there was a preference for certain phenotypes, majority of farmers do not practice a selective breeding program. Hence, the populations remained as an admixture group in village production system.

In contrast to the fast genetic progress that could be achieved by upgrading or crossbreeding programs, the diversity in the backyard chickens yields a steady and heterogeneous genetic base adapted to the low input/output smallholder system in Sri Lanka as in many parts of the developing world (FAO, 2010). While agreeing to the fact that the ancestral diversity that existed in the contributing populations leading to the admixture may be lost in the path of evolution of the backyard chicken populations, there is a curbing effect due to the absence of selection pressure or any directional selection, owing to the sociocultural reasons of farmers interwoven in the production system. Thus, the absence of common and directional selection could have been instrumental in preserving the rich diversity of contributing populations to a certain level, which otherwise could have been lost, resulting in high phenotypic and genetic diversity in the different populations studied.

The local chickens in Sri Lanka have been bred in the backyard low-input production systems for generations. Following this process, unique genetic variants could have evolved as adaptations to the climate and management conditions in Sri Lanka. Therefore, these birds can be used as a gene pool to maintain their major, specific genetic variants present among the backyard non-descript chicken populations in Sri Lanka. Since the genetic diversity existing in the populations studied are mainly constituted by individual-level variations, there is no strong population genetic structure formed within any of these populations. Therefore, they all show weak genetic fragmentations, even at household level due to the common management and breeding strategies practiced by farmers for generations. Accordingly, the present populations of local chickens in Sri Lanka serve the purpose of conservation through sustainable utilization, and more importantly, they could be considered as an ideal foundation for genetic improvement to establish breeds/lines to be adapted to particular environments and production systems in Sri Lanka.

The findings of this study confirm the genetic wealth conserved within the local chicken populations. The absence of population genetic structure is a result of the management and breeding regime commonly practiced in different localities of Sri Lanka. Therefore, we recommend that future strategies focus on sustainable development of this valuable resource with interventions appropriate to empower the existing operations of these village flocks in order to ensure that genetic diversity is maintained with time.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Research and Ethics Committee (IREC) and Institutional Animal Care and Use Committees (IACUC) of the International Livestock Research Institute (ILRI), Nairobi, Kenya.

AUTHOR CONTRIBUTIONS

MI, JH, and PS conceptualized and designed the study. RL acquired the data. AS and JH analyzed the data. AS, JH, and PS interpreted the results. AS and PS drafted the article. AO, JH, and PS critically revised the article. MI, AO, JH, and PS gave the final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financed by the Global Environment Facility via the United Nations Environment Programme (GEF-UNEP),

grant name GFL-2328-2715-4A41: Development and application of decision-support tools to conserve and sustainably use genetic diversity in indigenous livestock and wild relatives. The GEF-UNEP-ILRI FAnGR Asia project was implemented in Sri Lanka by the University of Peradeniya.

ACKNOWLEDGMENTS

The authors acknowledge the support of members of the ILRI and the Sri Lankan teams, and also all the farm families in the project sites.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.723706/full#supplementary-material

Supplementary Table 1 | Allele sizes of 20 microsatellite markers for eight phenotype-based populations: naked neck (NN), long legged (LL), crested or crown (CC), frizzle feathered (FF), Giriraj (GR), commercial layer (CL), crossbreds (CR), and non-descript village chicken (VC).

Supplementary Table 2 | Allele sizes of 20 microsatellite markers for five village populations: Dematagama (DM), Labunoruwa (LA), Ooththupitiya (OT), Tabbowa (TB) and Thewanuwara (TH).

REFERENCES

Abeykoon, M., Weerahewa, J., Weligamage, P., and Silva, G. (2014). Willingness to pay for chicks of different indigenous chicken types: an application of experimental auctions. *Trop. Agric. Res.* 26, 162–174.

Barton, N., and Slatkin, M. (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56, 409–415. doi: 10.1038/hdy.1986.63

Belkhir, K. (2004). *GENETIX 4.05, Logiciel Sous Windows TM Pour la Génétique des Populations*. Available online at: http://www.genetix.univ-montp2.fr/genetix/genetix.htm (accessed July 28, 2015).

Berthouly, C., Bed'Hom, B., Tixier-Boichard, M., Chen, C., Lee, Y., Laloë, D., et al. (2008). Using molecular markers and multivariate methods to study the genetic diversity of local European and Asian chicken breeds. *Anim. Genet.* 39, 121–129. doi: 10.1111/j.1365-2052.2008.01703.x

Berthouly, C., Leroy, G., Van, T. N., Thanh, H. H., Bed'Hom, B., Nguyen, B. T., et al. (2009). Genetic analysis of local Vietnamese chickens provides evidence of gene flow from wild to domestic populations. *BMC Genet.* 10:1. doi: 10.1186/1471-2156-10-1

Bett, R., Bhuiyan, A., Khan, M., Silva, G., Islam, F., Abeykoon, M., et al. (2014). Phenotypic variation of native chicken populations in the South and South East Asia. *Int. J. Poult. Sci.* 13:449.

Bodzsar, N., Eding, H., Revay, T., Hidas, A., and Weigend, S. (2009). Genetic diversity of Hungarian indigenous chicken breeds based on microsatellite markers. *Anim. Genet.* 40, 516–523. doi: 10.1111/j.1365-2052.2009.01876.x

Cuc, N., Simianer, H., Eding, H., Tieu, H., Cuong, V., Wollny, C., et al. (2010). Assessing genetic diversity of Vietnamese local chicken breeds using microsatellites. *Anim. Genet.* 41, 545–547. doi: 10.1111/j.1365-2052.2010.02039.x

Earl, D. A. (2012). Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Eltanany, M., Philipp, U., Weigend, S., and Distl, O. (2011). Genetic diversity of ten Egyptian chicken strains using 29 microsatellite markers. *Anim. Genet.* 42, 666–669. doi: 10.1111/j.1365-2052.2011.02185.x

Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform.* 1:11769343050010003. doi: 10.1177/117693430500100003

FAO (2007). *Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration*. Rome: FAO.

FAO (2010). *Chicken Genetic Resources Used in Smallholder Production Systems and Opportunities for Their Development. by P. Sørensen.* FAO Smallholder Poultry Production Paper No. 5. Rome: FAO.

FAO (2011). *Molecular Genetic Characterization of Animal Genetic Resources. FAO Animal Production and Health Guidelines. No. 9.* Rome: FAO.

Gärke, C., Ytournel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., et al. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim. Genet.* 43, 419–428. doi: 10.1111/j.1365-2052.2011.02284.x

Gunaratne, S., Chandrasiri, A., Hemalatha, W. M., and Roberts, J. (1993). Feed resource base for scavenging village chickens in Sri Lanka. *Trop. Anim. Health Prod.* 25, 249–257. doi: 10.1007/BF02250880

Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361–372. doi: 10.2307/2532296

Hedrick, P. W. (2013). High inbreeding in sheep or erroneous estimation? *J. Hered.* 104, 298–299. doi: 10.1093/jhered/ess139

Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nat. Rev. Genet.* 10, 639–650. doi: 10.1038/nrg2611

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233

Jayasena, D. D., Jung, S., Bae, Y. S., Park, H. B., Lee, J. H., and Jo, C. (2015). Comparison of the amounts of endogenous bioactive compounds in raw and cooked meats from commercial broilers and indigenous chickens. *J. Food Compost. Anal.* 37, 20–24. doi: 10.1016/j.jfca.2014.06.016

Lenstra, J., Groeneveld, L., Eding, H., Kantanen, J., Williams, J., Taberlet, P., et al. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim. Genet.* 43, 483–502. doi: 10.1111/j.1365-2052.2011.02309.x

Leroy, G., Kayang, B. B., Youssao, I. A., Yapi-Gnaoré, C. V., Osei-Amponsah, R., N'Goran, E. L., et al. (2012). Gene diversity, agroecological structure and introgression patterns among village chicken populations across North, West and Central Africa. *BMC Genet.* 13:34. doi: 10.1186/1471-2156-13-34

Liyanage, R., Dematawewa, C., and Silva, G. (2015). Comparative study on morphological and morphometric features of village chicken in Sri Lanka. *Trop. Agric. Res.* 26, 261–273.

Nei, M. (ed.). (1987). "Genetic distance between populations," in *Molecular Evolutionary Genetics* (Chichester, NY: Columbia University Press), 208–253.

Ota, T. (1993). *DISPAN: Genetic Distance and Phylogenetic Analysis*. State College, PA: Pennsylvania state university.

Park, S. (2001). *MStools v 3.1 (Excel spreadsheet Toolkit for Data Conversion)*. Dublin: Smurfit Institute of Genetics.

Pham, M., Berthouly-Salazar, C., Tran, X., Chang, W., Crooijmans, R., Lin, D., et al. (2013). Genetic diversity of V ietnamese domestic chicken populations as decision-making support for conservation strategies. *Anim. Genet.* 44, 509–521. doi: 10.1111/age.12045

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945

Punyawardena, B. (2010). "Agro-ecological context and farm animal genetic resources in Sri Lanka," in *Indigenous Animal Genetic Resources in Sri Lanka:*

Status, Potential and Opportunities, ed. P. Silva (Peradeniya: University of Perdeniya), 50–84.

Raymond, M. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249.

Reynolds, J., Weir, B. S., and Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105, 767–779. doi: 10.1093/genetics/105.3.767

Samaraweera, A., Silva, P., Abeykone, N., Ibrahim, M. N., Okeyo Mwai, A., and Jianlin, H. (2014). "Population genetic structure of Sri Lankan backyard chicken flocks: implication for conservation and genetic improvement programs," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*: American Society of Animal Science, Vancouver, BC.

Samaraweera, M., Himali, S., Zeng, S., Jianlin, H., and Silva, P. (2011). Development of molecular tools to differentiate Sri Lankan wild boar (*Sus scrofa* affinis) meat from exotic and village pig (*Sus scrofa* domestica) meat. *Trop. Agric. Res.* 23, 11–20.

Sanjeewa, M., Liyanage, R., Vidanarachchi, J., and Silva, G. (2011). Association between egg production and body morphology of some village chicken ecotypes in Sri Lanka. *Proc. Univ. Res. Ses. Univ. Peradeniya* 16, 44.

Silva, P., Abeykone, N., Samaraweera, A., Han, J., Ibrahim, M., and Okeyo, A. (2014). "Genetic diversity and adaptability exist among backyard poultry populations in Sri Lanka," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, BC.

Silva, P., Guan, X., Ho-Shing, O., Jones, J., Xu, J., Hui, D., et al. (2009). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Anim. Genet* 40, 1–9. doi: 10.1111/j.1365-2052.2008.01783.x

Wang, M.-S., Thakur, M., Peng, M.-S., Jiang, Y., Frantz, L. A. F., Li, M., et al. (2020). 863 genomes reveal the origin and domestication of chicken. *Cell Res.* 30, 693–701. doi: 10.1038/s41422-020-0349-y

Wang, Y., Gao, Y., Imsland, F., Gu, X., Feng, C., Liu, R., et al. (2012). The crest phenotype in chicken is associated with ectopic expression of HOXC8 in cranial skin. *PLoS One* 7:e34012. doi: 10.1371/journal.pone.0034012

Wattanachant, S., Benjakul, S., and Ledward, D. (2004). Composition, color, and texture of Thai indigenous and broiler chicken muscles. *Poult. Sci.* 83, 123–128. doi: 10.1093/ps/83.1.123

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641

Yunis, R., and Cahaner, A. (1999). The effects of the naked neck (Na) and frizzle (F) genes on growth and meat yield of broilers and their interactions with ambient temperatures and potential growth rate. *Poult. Sci.* 78, 1347–1352. doi: 10.1093/ps/78.10.1347

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fgene.2021.666947

Heritability Estimate for Antibody Response to Vaccination and Survival to a Newcastle Disease Infection of Native chicken in a Low-Input Production System

Blaise Arnaud Hako Touko[1], Anold Tatah Kong Mbiydzenyuy[1,2], Tebug Thomas Tumasang[3] and Julius Awah-Ndukum[3]*

[1]Biotechnology and Bioinformatics Research Unit, Department of Animal Production, Faculty of Agronomy and Agricultural Sciences, University of Dschang, Dschang, Cameroon, [2]Animal Research Lab, Department of Animal Sciences, School of Agriculture and Natural Resources, Catholic University Institute of Buea, Buea, Cameroon, [3]Laboratory of Animal Physiology and Health, Department of Animal Sciences, University of Dschang, Dschang, Cameroon

OPEN ACCESS

Edited by:
Olivier Hubert Hanotte, University of Nottingham, United Kingdom

Reviewed by:
Muhammed Walugembe, Iowa State University, United States
Kellie Anne Watson, University of Edinburgh, United Kingdom

***Correspondence:**
Blaise Arnaud Hako Touko
hakoarnaud@gmail.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 February 2021*
Accepted: *20 August 2021*
Published: *30 September 2021*

Citation:
Hako Touko BA, Kong Mbiydzenyuy AT, Thomas Tumasang T and Awah-Ndukum J (2021) Heritability Estimate for Antibody Response to Vaccination and Survival to a Newcastle Disease Infection of Native chicken in a Low-Input Production System. Front. Genet. 12:666947. doi: 10.3389/fgene.2021.666947

The Newcastle disease virus (NDV) is the deadliest chicken pathogen in low-input village poultry, and selecting for NDV resistance has been recommended as a sustainable strategy in backyard poultry production systems. However, selecting for disease resistance needs precision data from either a big population sample size or on many generations with good pedigree records for effective prediction of heritability (*h2*) and breeding values of the foundation stock. Such conditions are almost impossible to meet in low-input backyard production systems. This study aimed at proposing a realistic method for estimating the heritability of the immune response to vaccination and survival of NDV infection in village poultry production to inform a breeding strategy for ND resistance in Cameroon. A 1 and 3% selection intensity of cocks and hens for higher antibody (ab) response (ABR) to vaccination followed by progeny selection of chickens who survived an experimental NDV infection was conducted from an initial population of 1,702 chickens. The selection induced an increase of 1012.47 units/ml ($p < 0.01$) of the NDV antibody of the progeny as well as an effective survival rate (ESR) increase of 11.75%. Three methods were used to estimate the heritability (h^2) of NDV antibody response to vaccination. h^2 was low irrespective of the method with estimates of 0.2227, 0.2442, and 0.2839 for the breeder's equation method, the graphical method, and the full-sib/half-sib nested design, respectively. The mortality rate of infected chickens was high (86%). The antibody response to selection was not influenced by sex and genetic type even though the opposite was observed ($p < 0.05$) for the ESR to NDV infection with naked neck chickens recording an ESR of 14% against 2.25% for the normal feather type. A very low heritability (0.0891) was observed for the survival against NDV infection. We confirm the evidence of disease resistance and the effect of selection for antibody response to vaccination on the improvement of the survival against NDV disease. Although the full sib/half sib nested design is more appropriate in case of availability of pedigree information, the direct methods

are still useful in case of unavailability of full pedigree information. It is recommended that gene expression analysis should be prioritized for disease-resistance assessment and selection of native breeds of poultry.

Keywords: native chicken, heritability, resistance, selection, Newcastle disease

INTRODUCTION

In Cameroon, native local chickens contribute to poverty alleviation and food security of more than 60% of families living in rural areas (Fotsa et al., 2011). However, the breeding stock is almost always renewed every year due to high mortalities and poor management practices. Newcastle disease (ND) is the deadliest among others with 80% of birds in endemic areas in Tanzania (Buza and Mwamhehe, 2001), Southern Africa (Mtileni et al., 2012), and Cameroon (Hako Touko et al., 2015). Vaccination has proven not to be cost-effective in remote areas, and having genotypes that can resist or survive disease outbreaks is believed to be a sustainable strategy. Investigations have been carried out within native chickens of Cameroon, and some genetic markers associated with a high immune response to ND virus (NDV) infection have been identified (Hako Touko et al., 2015). There is a need to design and implement a breeding program in line with the valorization of ND resistance.

The estimation of heritability is a requirement for defining the appropriate improvement strategy. It is well documented that the heritability of antibody response (ABR) to the ND vaccine is low to moderate (Lwelamira et al., 2009; Liu et al., 2014; Rowland et al., 2018). However, that of ND resistance has not yet been estimated. It is paramount to determine the heritability of NDV resistance for local chickens of Cameroon to define the most appropriate strategy for its improvement. The estimation of genetic parameters is quite challenging in low-input poultry production considering the poor management conditions faced. Native chickens are reared as small-size flocks of 3–10 birds per household. They are unselected heterogeneous genetic resources randomly bred with little or no defined management system of feeding, housing, mating, or health management. There are no existing production records or pedigree information kept and no performance evaluation system. It is, therefore, currently impossible to have data on multiple generations for predictive analyses. These peculiarities make it difficult to obtain an initial and homogenous genetic resource for research and estimation of the genetic parameters.

In a previous study, a sample population was selected using LEI158 and MCW371 microsatellite markers of the major histocompatibility complex (MHC) B and quantitative trait loci (QTL) microsatellite alleles of favorable effect on antibody production against ND (Hako Touko et al., 2015). However, due to significant allele diversity, high selection intensity may be required for the improvement of antibody response and survival of NDV infection. This study aimed at proposing a realistic method for estimating the heritability of the immune response to vaccination and survival of NDV infection in village poultry production to inform a breeding strategy for ND resistance.

Abbreviations: ab, Antibody; ABR, Antibody response; APMV1, Avian Paramyxovirus type 1; CBM, Community-based management; ESR, Effective survival rate; GWAS, Genome wide association studies; G0, Parental; population; G1, Offspring of selected population; Gs, Selected parents; h2, Heritability; MHC, Major histocompatibility complex; Na, Naked neck gene; ND, Newcastle disease; NDV, Newcastle disease virus; QTL, Quantitative trait loci; SNP, Single nucleotide polymorphism; TSR, Total survival rate.

MATERIALS AND METHODS

The animal materials used in this research are not subject to any restriction and are approved by the scientific and ethical committee of the Department of Animal Science of the University of Dschang, No E0091/DZOO/FASA/UDs.

Sampling and Experimental Procedure

The study was carried out at the experimental poultry farm of the School of Agriculture and Natural Resources of the Catholic University Institute of Buea, and serological tests were completed in the Laboratory of Animal Health of the Faculty of Agronomy and Agricultural Sciences of the University of Dschang. A total of 1,802 adult local chickens with no pedigree record were collected from farmers of the Community Based Management of Native Chickens Genetic Resources. The birds were 24 weeks of age, including 912 normal feathered (610 hens and 302 cocks) and 890 naked necks (570 hens and 300 cocks). The birds were kept in the poultry farm for 28 days for stabilization and acclimatization.

G0, Gs, G1, and Control

After acclimatization, 1,000 eggs were collected from each parental stock (naked neck and normal feathered) and incubated to produce Parental; population (G0) day-old chicks for each group. G0 chicks were vaccinated against other viral diseases, including infectious bronchitis and Gumboro disease. Other breeding conditions, including prophylaxis, housing, and management were the same. Birds were fed *ad libitum*. Selected parents (Gs) consisted of two groups of equal size: The first group consisted of the top 1% of cocks and 3% of hens, including three cocks and 18 hens for the normal feathered and three cocks and 17 hens for the naked neck, selected for higher NDV antibody mean titer. This group was used to produced Offspring of selected population (G1; Gs offspring hatched from 900 fertile eggs, including 450 from normal feathered and 450 for naked neck). The second group was used for the NDV experimental infection of Gs. The control birds were considered for experimental infection and consisted of 50 normal feathered and 50 naked necks not vaccinated against NDV.

Antibody Responsiveness to Vaccination

For antibody responsiveness to vaccination, G0, Gs, and G1 were tested at 28 weeks of age. Blood samples were collected before vaccination against ND and 14 days after vaccination. Sera from collected samples were isolated in 1.5-ml microtubes after sedimentation and screened for antibody titer. The vaccine used was Inmugal V.P. Hitchener B1 of Ovejero Laboratorios, Leon-Spain. An indirect ELISA method using Biochek Poultry Immunoassays for ND Antibody Test kit (catalogue code CK 116, www.asineh.com/PDF/Biochek; Product catalogue. PDF, pages 16–19) was used according to the strict instructions of the manufacturer. The relative amounts of antibodies in chicken samples were calculated by reference to the positive control and expressed as sample to positive (S/P) ratio and calculated as follows:

$$S/P = (\text{Mean of Test Sample} - \text{Mean of negative control})/(\text{Mean of Test sample-Mean of negative control})$$

For interpretation, samples with an S/P of 0.350 or greater contain anti-NDV antibodies and are considered positive. For the calculation of the titer, the following equation relates the S/P of a samples:

$$\text{Log10 Titer} = 1.0 * \text{Log(SP)} + 3.52, \text{ with antilog} = \text{Titer}$$

Experimental Infection

For the experimental infection, Gs and G1 were challenged at 7 months of age. Experimental chicks were randomly sampled from 563 chicks, including 202 naked necks (T1) and 261 normal feathered chicks (T2). Four families of one cock and five hens were formed for each treatment, and 150 chicks were sampled per treatment, including 100 chicks of both sexes for T1 and 50 chicks for the first control (T1C) and the same for T2 and T2C. Experimental birds were reared identically and vaccinated against other common viral diseases, including Gumboro and infectious bronchitis. They were then challenged against the NDV, and records on survival were taken 14 weeks after infection. All the birds that remained alive after the 14 days of the trial were considered to have survived, and the survival rate (SR) was estimated. The experimental birds were then reared for 14 more days. Then, hens were mated with unchallenged healthy and fertile cocks and cocks with unchallenged healthy and fertile hens. After 2 months, only hens that could lay and hatch fertile eggs were considered genetically viable, and the effective survival rate (ESR) of challenged birds was estimated.

The APMV1 Strain

Isolates of Avian Paramyxovirus type 1 (APMV1) were harvested from postmortem diagnostics of sick chickens following the ND challenge. The NDV was cultured in fresh eggs and tested for pathogenicity index according to the standard operating manual in use. Randomly selected chicks were infected, and blood samples were collected before the infection and 14 days after for antibody analysis. Then, viable birds capable of reproduction were presumed resistant and conserved for further studies.

Statistical Analysis

Three methods, as detailed in **Appendix 1–4**, were used to estimate heritability, including the breeder's equation method (Kelly, 2011), the graphical method adapted from Boyer (1958) and Verrier et al. (2009), and finally, the nested full-sib/half-sib analysis (Becker, 1975; Lynch and Walsh, 1998) and heritability method for discrete parameters (Razungles, 1977).

RESULTS

Improvement for Antibody Responsiveness to Vaccination

The effect of selection for antibody responsiveness to vaccination on the antibody mean titer of the offspring according to genetic type and sex of the experimental birds (**Tables 1** and **2**) shows that all the birds tested negative for NDV antibody at 29 weeks. For the same age, there was no significant difference between the genetic types and sex. The 1% selection intensity of cocks and 3% selection intensity

TABLE 1 | Antibody mean titer of parents, selected parents, and their offspring according to genetic types, sex, and age.

Genetic type	Sex (number of birds)	Age in Weeks	Antibody mean titrer (Unit/ml)			Sign
			Go	Gs	G1	
Naked neck	Cock (522)	29	314	311	300	ns
		31	6083[c] (300)	10,083[a] (3)	7052[b] (219)	**
	Hen (818)	29	371	306	263	ns
		31	6005[c] (570)	10,205[a] (17)	7089[b] (231)	**
Normal feathered	Cock (528)	29	298	315	304	ns
		31	6109[c] (302)	10,282[a] (3)	6617[b] (223)	**
	Hen (855)	29	217	308	291	ns
		31	5895[c] (610)	10,129[a] (18)	7055[b] (227)	**
Overall mean	(2723)	29	302	311	315	
		31	6036[c] (1782)	10,181[a] (41)	7058[b] (900)	

[a,b,c] *for the same age, between columns and for the same column between genetic types, variables bearing the same letters are statistically comparable ($p > 0.01$).*
ns, no significant difference observed; G0, parental population; Gs, selected parents; G1, offspring from selected parents.

TABLE 2 | Parameters of the distribution.

Parameters	Population mean (μ_p)	Selected parents mean (μ_s)	Mean offspring ((μ_{p1})
Mean (μ)	6036.01	10,181.95	7048.48
Standard Deviation (σ)	2391.04	1308.11	1029.94
Variance (σ^2)	5,717,105.40	1,711,155.38	1,128,366.22
Coefficient of variation	39.61	12.85	14.61

of hens resulted in a significant ($p < 0.01$) NDV antibody response of 1012.47 units/ml corresponding to 16.77% ABR of the parental population from an average mean titer of 6,023 to 7,058 units/ml for the offspring of selected parents as expected. However, the overall mean of antibody response to vaccination of selected parents (10,181 units/ml; **Table 2**) was higher than that of their progeny at the same age (7,058 units/ml).

Heritability Estimate of the Antibody Responsiveness to Vaccination

The analysis of the parameters of distributions (**Table 3**) shows that the 1% sire and 3% dam selection decreased the SD with a coefficient of variation of 13 and 15%, respectively, for selected parents and their offspring as compared with 39.61% for the population. h^2 was moderate irrespective of the method with estimates of 0.2227, 0.2442, and 0.2839 for the breeder's equation and the graphical methods (**Table 4**) and the full-sib/half-sib nested design (**Table 5**), respectively.

Heritability Estimate of Survival of NDV Infection

The estimate of the heritability of survival of NDV infection (**Table 5**) is very low = 0.0891, showing that only 9% of the superiority of selected resistant parents is transferred to their immediate generation.

Survival of NDV Infection

For the effect of the selection for antibody responsiveness to ND vaccination on the survival of ND infection of the progeny of local chickens (**Table 6**), the selection intensity of 1% cocks and 3% hens has led to an improvement of the survival rate from 6 to 21%, equivalent to a mortality rate of 94 to 79% and of the ESR from 2.25 to 14%. The improvement of the survival against ND infection is observed for all experimental groups irrespective of the genetic type or sex.

TABLE 3 | Selection parameters and heritability estimate from the direct method.

Selection parameters	Formula	Estimate
Δ_{ps}	$\mu_s - \mu_p$	4145.94
Δ_{p1}	$\mu_{p1} - \mu_P$	1012.47
h_1^2	$\Delta_{p1} / \Delta_{ps}$	0.2227
h_2^2	$\frac{P_{offspring} - \bar{P}_{population}}{\left(P_{selected\ parents} - \bar{P}_{population}\right)}$	0.2442

$\Delta_{ps} = selection\ differential.$

$\Delta_{p1} = G1 superiority = genetic\ gain.$

$h_1^2 = heritability\ from\ the\ graphical\ method.$

$h_2^2 = heritability\ from\ the\ breeder's equation.$

DISCUSSION

The progeny of immunized parents were all NDV-ab negative at 29 weeks. It was previously confirmed that parental NDV-ab of immunized birds is present in the blood and effective to protect their chicks only 10 days post-hatchery (Hako Touko et al., 2015). Similar findings of Shahid and Liaquat (2017) reported that maternal NDV-ab protects their progeny during the 1st week of life. The selection for high parental NDV-ab response improved NDV-ab response of their sibs 16.77% and survival of infection 15% SR and 11.75% ESR, respectively. ESR is more appropriate to conclude about the effectiveness of the selection as only birds that survive the infection and are capable of reproducing are considered genetically viable.

The significant increase ($p < 0.01$) of antibody titer (1012.47 units/ml) of the offspring as compared with the population mean induced a correlated response of 11.75% ESR of offspring of selected parents. This is evidence that the selection for antibody responsiveness to vaccination against NDV improves natural immunity against NDV and can, therefore, be recommended for selection to improve disease tolerance or resistance. Antibody response plays an important role in host resistance to ND, and selection for antibody response can effectively improve disease resistance in chickens (Luo et al., 2013). Moreover, chickens with high NDV-ab titers were associated with a higher frequency of MHC-B and QTL alleles of favorable effect on disease resistance in chickens (Hako Touko et al., 2015). Despite the improvement of the immune response in the progeny, the mortality rate observed in this study remains high (86%). A lower rate (83.72%) was earlier reported in the field (Hako Touko et al., 2015) although higher mortalities (90–100%) were recorded by several authors (Nanthakumar et al., 2000; Orsi et al., 2010; Spickler, 2016). This justifies why NDV is included in "listed" agents or reportable diseases by the Office International des Epizooties (OIE; Aldous and Alexander, 2001; Boynukara et al., 2013) and is considered to be the number one disease constraint of economical importance in poultry production around the globe with more impact in low- and middle-income countries (Liu et al., 2014; Deist et al., 2017; Rowland et al., 2018). The high mortality rate from experimental infection of unvaccinated birds is a call for concern, and the method of selection from experimental infections should be restricted.

TABLE 4 | Heritability estimate from variance component.

Factor	df	MS	E(MS)	σ^2	h^2
Sires	5	13,318,469.9	$\sigma_e^2 + 10\,\sigma_d^2 + 60\sigma_s^2$	76,057.5362	0.2839
Dam/Sire	30	3,839,841.5	$\sigma_e^2 + 10\,\sigma_d^2$	−75,038.6533	
Sibb/Dam	324	996,554.475	σ_e^2	1,070,373.33	

TABLE 5 | ANOVA of the survival to Newcastle disease virus (NDV) infection.

Factor	df	SS	MS	E(MS)	σ	h^2
Sires	3	0,36	0,1200	$\sigma_w^2 + 50\sigma_s^2$	$\sigma_s^2 = -0,0077$	$\frac{4\sigma_s^2}{\sigma_w^2 + \sigma_d^2 + \sigma_s^2}$ =0.0891
Dam/ Sire	196	−36,04	−0,1839	$\sigma_w^2 + 0,8163\sigma_d^2$	$\sigma_d^2 = -0,8459$	
Sibb/ Dam	150	76	0,5066	$\sigma_w^2 = 0,5066$	$\sigma_w^2 = 0,5066$	
Total	199	40,32	0,2026			

TABLE 6 | Effect of the selection of parent for antibody responsiveness to vaccination on the survival of local chicken according to genetic group and sex.

Genetic type	Sex	Parents			Offspring		
		Number of birds	Survival (%)	Effective survival (%)	Number of birds	Survival (%)	Effective survival (%)
Naked Neck	Cock	50	8	2	50	21	14
	Hen	50	4	3	50	17	14
Normal feathered	Cock	50	2	0	50	16	11
	Hen	50	6	4	50	21	17
Overall		200	5	2.25	200	18.75	14

The heritability of NDV-ab response to vaccination was estimated from the Breeder's equation (Kelly, 2011), graphical method (Verrier et al., 2009), and ANOVA full-sib/half-sib nested design (Lynch and Walsh, 1998) and were all low (0.2227, 0.2442, and 0.2839). It is known that low heritability is an indication that the genetic gain from the mass selection is very low. Consequently, the improvement of the antibody responsiveness to vaccination against NDV through selection is time-consuming. In this situation, an improvement of environmental factors and management practices, including feeding, housing, and vaccination, is likely to minimize the incidence of NDV infection and improve poultry productivity in low- and middle-income countries. All three estimates are less than the heritability estimates of 0.29, 0.31, and 0.48 earlier reported in low-input production systems (Peleg et al., 1976; Liu et al., 2014; Lwelamira, 2012). Statistically, the range of the heritability estimate is [0,1].

The heritability obtained for the survival to infection was lower (0.0891) compared with that of antibody response to the vaccination (0.2442). Considering the selection, lower values of heritability would be expected from the unselected population of indigenous chickens of Cameroon. These observations expose that parental improvement of low heritability traits such as survival of infection due to the selection for high antibody response is low. This may be explained by the fact that survival of infection as disease resistance is a complex trait under multiple genes' influence with each gene inducing only part of the overall effect (Fulton et al., 2006). In this context, bigger population size may better reflect the polygenetic mode of action of expected genes. The low ESR, despite the moderate antibody response to vaccination, may also be correlated with the velogenic nature of the NDV (Shabbir et al., 2012).

In this study, the higher survival or lower mortality rate and higher antibody titer are indicators of improved resistance to NDV infection. Previous studies define resistance as the ability of the host to interfere with the pathogen life cycle in various ways, including lower pathogen load, higher antibody, and less morbidity and/mortality (Rauw, 2012; Bishop, 2014; Deist et al., 2017). We used two variables as indicators of survival, including SR and ESR. The second that is either equal to or lower than the first gives the percentage of individuals that survive the infection and are capable of reproduction. Therefore, ESR is more reliable to conclude about the disease-resistance status of a breed as well as for the estimation of genetic parameters. Despite the selection for high antibody response, the level of resistance improvement was still low as evidenced by the relatively low SR obtained. It is known that the immune response to viruses is very complex, and antibody response to NDV is a quantitative trait under polygenic control (Yonash et al., 2001; Biscarini et al., 2010). According to Liu et al. (2014), many significant markers influence the innate and adaptative immune response of chickens, but none of them can explain more than 5% of the phenotypic variance.

CONCLUSION

The 1% selection intensity of cocks and 3% selection intensity of hens resulted in a significant increase ($p < 0.01$) of the NDV antibody response of 1012.47 units/ml corresponding to 16.77% ABR of the parental population from an average mean title of 6,023 to 7,058 units/ml for the offspring of selected parents. Low values of heritability for antibody responsiveness and survival to experimental infection confirm that mass selection is not effective for the improvement of natural resistance against NDV. Disease resistance being a quantitative trait under polygenic control, it is, therefore, suggested that a genome-wide associated study be conducted in view of identifying more genes involved and proposing an efficient selection strategy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics and Scientific Committee of University of Dschang.

AUTHOR CONTRIBUTIONS

BH designed the study, performed statistical analysis of the data, field supervision of the project, and drafting of the manuscript, and wrote the manuscript. AK managed experimental birds and collected data. TT carried out the laboratory test. JA-N provided the general supervision of the work. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by IFS/COMSTECH Research grant Upgrading natural disease resistance against Avian paramyxovirus 1 (Newcastle disease) and *in situ in vivo* conservation of selected local chickens Grant B/5115-2.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the International Foundation for science (IFS) and its partner the Organization of Islamic Cooperation's Standing Committee on Scientific and Technological Cooperation (COMSTECH). We highly appreciate support from Clean and Sustainable Agriculture Venture Alliance (CASAVA) for facilitating the process within the community.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.666947/full#supplementary-material

REFERENCES

Aldous, E. W., and Alexander, D. J. (2001). Detection and differentiation of Newcastle disease virus (avian paramyxovirus type 1). *J. Avian Pathol.* 30, 117–128. doi: 10.1080/03079450120044515

Becker, W. (1975). *Manual of Quantitative Genetics. 3rd Edn*. Washington DC: State University Press.

Biscarini, F., Bovenhuis, H., van Arendonk, J. A. M., Parmentier, H., Jungfrau's, A., and van der Poel, J. (2010). Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet.* 41, 26–38. doi: 10.1111/j.1365-2052.2009.01960.x

Bishop, S. (2014). "Disease genetics: successes, challenges and lessons learnt." in *10th World Congress of Genetics Applied to Livestock Production*, August 17–22, 2014. Vancouver, BC.

Boyer, J. (1958). Théorie et calcul des index de Sélection. *Ann. Zootech.* 7, 193–242. doi: 10.1051/animres:19580303

Boynukara, B., Gulhan, T., Coven, F., Kiziroglu, I., and Durmus, A. (2013). Determination of Newcastle disease virus among wild bird populations in Lake Van basin. *Turkish J. Vet. Anim. Sci.* 37, 86–93. doi: 10.3906/vet-1111-26

Buza, J., and Mwamhehe, H. (2001). "Country report: Tanzania." in *Proceedings of SADC Planning Workshop on Newcastle Disease Control in Village Chickens: Maputo, Mozambique. Vol. 103.* eds. R. G. Alder and P. B. Spradbrow; March 6–9, 2000, 38–42.

Deist, M., Gallardo, R., Bunn, D., Dekkersm, J., Zhou, H., and Lamont, S. (2017). Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics* 18:989. doi: 10.1186/s12864-017-4380-4

Fotsa, J., Poné, D., Bordas, A., Tixier-Boichard, M., and Rognon, X. (2011). Assessment of the genetic diversity of Cameroon indigenous chickens by the use of microsatellites. *Livest. Res. Rural. Dev.* 23:118.

Fulton, T., Juul-Madsen, H., Ashwell, C., McCarron, X., Arthur, D., O'Sullivan, M., et al. (2006). Molecular genotype identification of the Gallus gallus major histocompatibility complex. *Immunogenetics* 58, 407–421. doi: 10.1371/journal.pone.0116352

Hako Touko, B., Keambou, T., Han, J., Bembidé, C., Skilton, R., Ogugo, M., et al. (2015). Molecular typing of the major histocompatibility complex B microsatellite haplotypes in Cameroon chicken. *Anim. Genet. Res.* 56, 47–54. doi: 10.1017/S2078633614000538

Kelly, J. K. (2011). The breeder's equation. *Nat. Educ. Knowledge* 4:5.

Liu, T., Qu, H., Luo, C., Li, X., Shu, D., Lund, M., et al. (2014). Genomic selection for the improvement of antibody response to Newcastle disease and avian influenza virus in chickens. *PLoS One* 9:e112685. doi: 10.1371/journal.pone.0116352

Luo, C., Qu, H., Ma, J., Wang, J., Li, C., Yang, C., et al. (2013). Genome-wide association study of antibody response to Newcastle disease virus in chicken. *BMC Genet.* 14:42. doi: 10.1186/1471-2156-14-42

Lwelamira, J., Kifaro, G., and Gwakisa, P. (2009). Genetic parameters for body weights, egg traits and antibody response against Newcastle disease virus (NDV) vaccine among two Tanzania chicken ecotypes. *Trop. Anim. Health Prod.* 41, 51–59. doi: 10.1007/s11250-008-9153-2

Lwalamira, J. (2012). Phenotypic and genetic parameters for body weights and antibody response against Newcastle disease virus (NDV) vaccine for Kuchi chicken ecotype of Tanzania under extensive management. *Trop Anim Health Prod.* 44, 1529–1534. doi: 10.1007/s11250-012-0099-z

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, United States: Sinauer Associates.

Mtileni, B., Muchadeyi, F., Maiwashe, A., Chimonyo, M., and Dzama, K. (2012). Conservation and utilization of indigenous chicken in southern Africa. *Worlds Poult. Sci. J.* 68, 727–747. doi: 10.1017/S0043933912000852

Nanthakumar, T., Kataria, R., Tiwari, A., Butchaiah, G., and Kataria, J. (2000). Pathotyping of Newcastle disease viruses by RT-PCR and restriction enzyme analysis. *J. Vet. Res. Commun.* 24, 275–286. doi: 10.1023/A:1006403017578

Orsi, M., Doretto, L., Camillo, S., Reischak, D., Ribeiro, S., Ramazzoti, A., et al. (2010). Prevalence of Newcastle disease virus in broiler chickens (Gallus gallus) in Brazil. *Braz. J. Microbiol.* 41, 349–357. doi: 10.1590/S1517-83822010000200014

Peleg, B., Soller, M., Ron, N., Hornstein, K., Brody, T., and Kalmar, E. (1976). Familial differences in antibody response of broiler chickens to vaccination with attenuated and inactivated Newcastle disease virus vaccine. *Avian Dis.* 20, 661–668. doi: 10.2307/1589446

Rauw, W. (2012). Immune response from a resource allocation perspective. *Front. Genet.* 3:267. doi: 10.3389/fgene.2012.00267

Razungles, J. (1977). Héritabilité des caractères discrets. Etude bibliographique critique. *Ann. Genet. Sel. Anim.* 9, 43–61. doi: 10.1051/1297-9686-9-1-43

Rowland, K., Saelao, P., Wang, Y., Fulton, J., Liebe, G., McCarron, A., et al. (2018). Association of candidate genes with response to heat and Newcastle disease virus. *Gene* 9:560. doi: 10.3390/genes9110560

Shabbir, M., Goraya, M., Abbas, M., Yaqub, T., Shabbir, M., Ahmad, A., et al. (2012). Complete genome sequencing of a Velogenic Viscerotropic avian paramyxovirus 1 isolated from pheasants (Pucrasia macrolopha) in Lahore. *Pak. J. Virol.* 86, 13828–13829. doi: 10.1128/JVI.02626-12

Shahid, Y., and Liaquat, A. (2017). Effect Newcastle disease on poultry due to high morbidity and mortality rates. *Glob. J. Poult. Farm. Vaccinat.* 5, 362–367.

Spickler, A. (2016). Newcastle Disease. Available at: http://www.cfsph.iastate.edu/DiseaseInfo/factsheets.php (Accessed: September 15, 2021)

Verrier, E., Rognon, X., Leroy, G., and Heams, T. (2009). Amélioration génétique des animaux Polycopié AgroParisTech, 125.

Yonash, N., Cheng, H., Hillel, J., and Cahaner, A. (2001). DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chickens. *Poult. Sci.* 80, 22–28. doi: 10.1093/ps/80.1.22

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

REVIEW
published: 21 October 2021
doi: 10.3389/fgene.2021.723181

Aspects of Molecular Genetics in Dromedary Camel

Mohammed Piro *

Veterinary Genetics Laboratory (LAGEV), Hassan II Agronomic and Veterinary Institute, Rabat, Morocco

Dromedary camels are unique in their morphological and physiological characteristics and are capable of providing milk and meat even under extreme environmental conditions. Like other species, the dromedary camel has also benefitted from the development of the molecular genetics to increase the knowledge about different aspect in camel genetics (genetic variation, molecular marker, parentage control, gene of interest, whole genome, dating. . .etc.). In this paper we review the different molecular genetic technics used in this particular species and future prospects. Dromedary genetic studies started in the end of the 1980s with phenotypic evaluation and the attempts to highlight the protein and biochemical diversity. In the 2000s, with the development of molecular markers such as microsatellites, genetic diversity of different types in several countries were estimated and microsatellites were also used for parentage control. In terms of genetic characterization, microsatellites revealed a defined global structure, differentiating East African and South Arabian dromedaries from North African, North Arabian, and South Asian individuals, respectively. Also, mitochondrialDNA sequence analysis of ancient DNA proved to be crucial in resolving domestication processes in dromedaries. Ancient and modern DNA revealed dynamics of domestication and cross-continental dispersion of the dromedary. Nuclear SNPs, single nucleotide polymorphisms changes that occur approximately each 1000 bps in the mammalian genome were also applied in some studies in dromedary. These markers are a very useful alternative to microsatellites and have been employed in some studies on genetic diversity and relevant phenotypic traits in livestock. Finally, thanks to the use of Next Generation Sequencing (NGS) the whole-genome assemblies of the dromedary (*Camelus dromedarius*) and a work to establish the organization of the dromedary genome at chromosome level were recently published.

Keywords: molecular genetics, microsatellite, SNP, mtDNA, dromedary camel

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri,
Sultan Qaboos University, Oman

Reviewed by:
Pamela Burger,
University of Veterinary Medicine Vienna, Austria
Muneeb M. Musthafa,
South Eastern University of Sri Lanka, Sri Lanka

***Correspondence:**
Mohammed Piro
vetpiro@yahoo.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *10 June 2021*
Accepted: *24 September 2021*
Published: *21 October 2021*

Citation:
Piro M (2021) Aspects of Molecular Genetics in Dromedary Camel. Front. Genet. 12:723181. doi: 10.3389/fgene.2021.723181

INTRODUCTION

Genetic characterization and assessment of genetic diversity is the primary step in the conservation and the management of genetic resources (Rout et al., 2008). Genetic characterization and diversity can be assessed within and between populations by different methods including biochemical and molecular techniques.

Like other species, the dromedary camel has also benefited from the development of the molecular genetics to improve our knowledge about different aspects in camel genetics (genetic variation, molecular markers, parentage control, gene of interest, whole genome, dating. . ..). Old World camelids have a specific morphological and physiological characteristic and can provide milk and meat even under extreme harsh conditions. Dromedary camels first appeared was estimated in the

middle Eocene (around 40 million years ago) and the first ancestors of the camelid's family were found in North America (Balmus et al., 2007). Afterward, they split into Old and New world camelids. The old-world camelids migrated to the eastern hemisphere and differentiated into two species: *Camelus bactrianus* and *Camelus dromedarius*. The split between Old and New world camelids was estimated by using molecular studies to have happened 11–16 million years ago (Kadwell et al., 2001) or 25 million years ago (Ji et al., 2009). The divergence between one-humped (*Camelus dromedarius*) and two-humped (*Camelus bactrianus*) camel species was estimated between 4.4 and 8 million years ago (Ji et al., 2009; Wu et al., 2014). The wild two humped camels (*Camelus ferus*) has been recently recognized as a separate species, based on molecular genetic data especially by mitochondrial DNA and nuclear markers and the time of separation was estimated around 0.6 and 1.8 million years ago (Mohandesan et al., 2017).

Investigations on Dromedary genetics started in the end of the 1980s with phenotypic evaluation and the attempt to highlight the protein and biochemical diversities. In the 2000s, with the development of molecular markers such as microsatellites, genetic diversity of different types or breeds of dromedary camels was estimated in several countries (Mburu et al., 2003) and microsatellites were also used for parentage control (Mariasegaram et al., 2002). Also, comprehension of domestication process was resolved using mtDNA sequence analysis of ancient DNA (Almathen et al., 2016). Ancient and modern DNA revealed dynamics of domestication and cross-continental dispersion of the dromedary camel (Almathen et al., 2016; Ciani, 2018). Nuclear SNPs, single base pair changes that occur approximately each 1,000 bps in the mammalian genome are also applied in some studies in the dromedary camel (Sushma et al., 2014; Abd El-Aziem et al., 2015; Ruiz et al., 2015; Lado et al., 2020a). These markers can be used as an alternative to microsatellites especially in genetic diversity studies and detection of relevant phenotypic traits in livestock.

With the implementation of the Next Generation Sequencing (NGS) technique, two whole-genome assemblies for the camel (*Camelus dromedarius*) were published (GenBank assembly accession: GCA_000803125.1) (Fitak et al., 2015) with the reference genome for the wild camel (*Camelus ferus*) (GenBank assembly accession: GCA_000311805.2) (Wu et al., 2014) and two whole-genome assemblies for the Bactrian camel (*Camelus Bactrianus*) (GenBank assembly accession: GCA_000604445.1 (Burger and Palmeri, 2014) and GCA_000767855.1 (Wu et al., 2014); while, another independent whole genome assembly was established in 2016. Recently, studies established the organization of the dromedary genome at the chromosome level (Fitak et al., 2015; Ruvinskiy et al., 2019; Lado et al., 2020a; Ming et al., 2020).

DROMEDARY CAMEL'S GENETIC MARKERS

Several genetic markers such as microsatellites, mtDNA and SNP can be used for genetic characterization of different species or breeds, parentage control and/or determination of traits of economic interest by genome wild associations studies (GWAS), an essential step for their use for selection assisted by markers or introgression.

Microsatellites

Microsatellites are used in studies aimed at characterizing genetic diversity studies because of their simple use and also the very low genetic variation in protein polymorphism (Guerouali and Acharbane, 2004). A large number of studies has been conducted, all around the world, on the genetic diversity of livestock species based on microsatellite loci. The number of loci to be genotyped or the size of the samples per breed is necessary to correctly analyzing the results. Studies in livestock describe a minimum of 8–10 loci (Cornuet et al., 1999; Arthofer et al., 2018) and more than 25 to 30 individuals per population for population genetic studies based on microsatellite allele frequencies. (Cornuet et al., 1999; Hale et al., 2012).

A set of Camelidae microsatellites was generated from published data on New World camelids: Alpacas and Ilamas (Obreque et al., 1998; Penedo et al., 1999), and several research studies have successfully assessed genetic variability in dromedary camels using these microsatellites (Jianlin et al., 2000; Sasse et al., 2000). This set comprises six-teen primers with highest polymorphism (VOLP03, VOLP08, VOLP10, VOLP32, VOLP67, YWLL02, YWLL08, YWLL09, YWLL38, YWLL44, LCA33, LCA37, LCA56, LCA63 LCA66, LCA77) (Nolte et al., 2005).

In 2002, Mariasegaram and collaborators were able to determine eight camel-specific microsatellites: CVRL01, CVRL02, CVRL03, CVRL04, CVRL05, CVRL06, CVRL07 and CVRL08.

The international panel on Animal Genetic Diversity (ISAG-FAO, 2004) recommended a list of 25 microsatellites markers for the evaluation of genetic diversity in camelids (**Table 1**), namely CMS09, CMS13, CMS15, CMS17, CMS18, CMS25, CMS32, CMS50, CMS121, CVRL01, CVRL02, CVRL05, CVRL06, CVRL07, LCA66, VOLP03, VOLP08, VOLP10, VOLP32, VOLP67, YWLL08, YWLL09, YWLL38, YWLL44 and YWLL59. Sixteen out of these 25 microsatellites are considered to be the most polymorphic and are thus highly recommended for dromedary camel genetic characterization. These are: YWLL08, YWLL09, YWLL38, YWLL44, YWLL59, VOLP03, VOLP08, VOLP10, VOLP32, VOLP67, LCA66, CVRL01, CVRL05, CVRL06, CVRL07 and CMS50 (Nolte et al., 2005).

Furthermore, the different studies on dromedary genetic diversity were not limited to these 16 microsatellites, but used instead a range of different microsatellite markers whose the number and name of which differ from one study to another. Markers such as YWLL02, YWLL29, YWLL36, YWLL40, LCA18, LCA33 and CMS58, show a low amount of variability; whereas YWLL08, YWLL09, YWLL38, YWLL44, YWLL59, VOLP03, VOLP08, VOLP10, VOLP32, VOLP67, LCA66, CVRL01, CVRL05, CVRL06, CVRL07 and CMS50 show much more alleles per locus and also high PIC values. These markers are useful in describing heterozygosity levels and are more informative (Muneeb, 2014).

There is only one study, carried out by Sadder et al. (2015), that dealt with the development of simple sequence repeat (SSR)

TABLE 1 | Primers, allele range and reference of the 25 microsatellites recommended for Dromedary camel by the International Panel on Animal Genetic Diversity.

Locus	Primers	Allele range	References
VOLP03	F-GCCAAAATAGGCTTACCCTTG R-CCCGCTTCATCTATTGGAAA	144–176	Obreque et al. (1998)
VOLP08	F-CCATTCACCCCATCTCTC TCGCCAGTGACCTTATTTAGA	142–180	Obreque et al. (1998)
VOLP10	F-CTTCTCCTTCCTCCCTACT R-CGTCCACTCCTTCATTC	249–267	Obreque et al. (1998)
VOLP32	F-GTGATCGGAATGGCTTGAAA R-CAGCGAGCACCTGAAAGAA	192–262	Obreque et al. (1998)
VOLP67	F-TTAGAGGGTCTATCCAGTTTC R-TGGACCTAAAAGAGTGGAG	149–203	Obreque et al. (1998)
YWLL02	F-GTGCACTCAGATACCTTCACA R-TACATCTGCAATGATCGACCC	290–304	Lang et al. (1996)
YWLL08	F-ATCAAGTTTGAGGTGCTTTCC R-CCATGGCATTGTGTTGAAGAC	134–172	Lang et al. (1996)
YWLL09	F-AAGTCTAGGAACCGGAATGC R-AGTCAATCTACACTCCTTGC	158–162	Lang et al. (1996)
YWLL38	F-GGCCTAAATCCTACTAGAC R-CCTCTCACTCTTGTTCTCCTC	174–192	Lang et al. (1996)
YWLL44	F-CTCAACAATGCTAGACCTTGG R-GAGAACACAGGCTGGTGAATA	090–114	Lang et al. (1996)
YWLL59	F-TGTGCAGGAGTTAGGTGTA R-CCATGTCTCTGAAGCTCTGGA	96–136	Lang et al. (1996)
LCA66	F-GTGCAGCGTCCAAATAGTCA R-CCAGCATCGTCCAGTATTCA	212–262	Penedo et al. (1999)
CMS9	F-TGCTTTAGACGACTTTTACTTTAC R-ATTTCACTTTCTTCATACTTGTGAT	227–256	Evdotchenko et al. (2003)
CMS13	F-TAGCCTGACTCTATCCATTTCTC R-ATTATTTGGAATTCAACTGTAAGG	238–265	Evdotchenko et al. (2003)
CMS15	F-AAAACTAAAGCCAGAAAGGCAAA R-TTTTTCCAGATCTTGCACCAC	81–121	Evdotchenko et al. (2003)
CMS17	F-TATAAAGGATCACTGCCTTC R-AAAATGAACCTCCATAAAGTTAG	149–167	Evdotchenko et al. (2003)
CMS18	F-GAACGACCCTTGAAGACGAA R-AGCAGCTGGTTTTAGGTCCA	144–166	Evdotchenko et al. (2003)
CMS25	F-GATCCTCCTGCGTTCTTATT R-CTAGCCTTTGATTGGAGCAT	93–102	Evdotchenko et al. (2003)
CMS32	F-ACGGACAAGAACTGCTCATA R-ACAACCAATAAATCCCCATT	198–209	Evdotchenko et al. (2003)
CMS50	F-TTTATAGTCAGAGAGAGTGCTG R-TGTAGGGTTCATTGTAACA	149–191	Evdotchenko et al. (2003)
CMS121	F-CAAGAGAACTGGTGAG GATTTTC R-AGTTGATAAAAATACAGCTGGAAAG	147–167	Evdotchenko et al. (2003)
CVRL01	F-GAAGAGGTTGGGGCACTAC R-CAGGCAGATATCCATTGAA	188–253	Mariasegaram et al. (2002)
CVRL02	F-TGTCACAAATGGCAAGAT R-AGTGTACGTAGCAGCATTATTT	205–215	Mariasegaram et al. (2002)
CVRL05	F-CCTTGGACCTCCTTGCTCTG R-GCCACTGGTCCCTGTCATT	148–174	Mariasegaram et al. (2002)
CVRL06	F-TTTTAAAAATTCTGACCAGGAGTCTG R-CATAATAGCCAAAACATGGAAACAAC	185–205	Mariasegaram et al. (2002)
CVRL07	F-AATACCCTAGTTGAAGCTCTGTCCT R-GAGTGCCTTTATAAATATGGGTCTG	279–299	Mariasegaram et al. (2002)

markers in four dromedary breeds using genome sequencing. These SSR markers can be used also in genetic studies of camels. Accordingly, the partial genome revealed 613 SSR loci with a minimum number of 5 repeat units. The SSR abundance was one in every 84.3 kb of contigs and the SSR loci comprised di- (80.8%), tri- (10.8%), tetra- (7.6%), and pentamer (0.8%) motifs (TA)n and (AC)n were the most abundant dimers (58.6%). Thirty SSR loci were experimentally characterized for dromedary and Bactrian camels.

Mitochondrial DNA

The use of mitochondrial DNA (mtDNA) reflects the maternal inheritance and is useful for the establishment of genetic variation between species. For example, 1.9% nucleotide difference in the

mitochondrial control regions (CR) was determined between Mongolian domestic and wild Bactrian camels (Ji et al., 2009; Silbermayr et al., 2010). On another hand, mtDNA sequence analysis showed the domestication processes in dromedaries and revealed the dynamics of domestication and cross-continental movement of the dromedary camel (Almathen et al., 2016). In this study, 1083 DNA samples from modern dromedaries were used. These samples were issued from 21 countries (Eastern Africa, Western and Northern Africa, North Arabian Peninsula, South Arabian Peninsula, and Southern Asia including samples from Australia). Seven DNA samples of early-domesticated dromedary specimens from Syria, Turkey, Jordan and Austria and eight wild dromedary specimens originating from the United Arabian Emirates excavated from archaeological sites (Al-Buhais, Umm an-Nar, Al-Sufouh, and Tell Abraq) were used to study the genetic profile of the mtDNA. The sequencing of mtDNA showed an unstructured camel population in North Africa and Asia. These findings indicated the absence of phylogeographic patterns reflecting the movements and the trading in the selected countries. They supposed that the coastal southeast Arabian Peninsula was a possible place of domestication and suggested a single domestication scenario with recurrent introgression from wild into the early domestic dromedaries (Almathen et al., 2016; Burger, 2016). In another study, two main maternal lineages related to the Middle-East and Eastern Africa were found in Tunisia after the comparison of the Tunisian camel mtDNA haplotypes and those available in DNA database (Nouaïria et al., 2018). Also, establishment of sub-groups was not possible because of a non-sufficient genetic structuring and nucleotide divergence.

Single Nucleotide Polymorphism

SNPs are modifications of a single base pair that occur approximately every 1,000 bps in the mammalian genome (Riva and Kone, 2002). They are present on nuclear DNA and on mtDNA (Petkovski et al., 2005). This type of polymorphism is very abundant and uniformly distributed in the genome and constitutes the largest source of genetic polymorphism. These variations have been identified during different programs of genome sequencing known as "Expressed Sequence Tag programs" (Lee et al., 2006). Once obtained, these markers can be very helpful because of their easy to use and their reproducibility. They have been used in many studies on genetic diversity and relevant phenotypic traits for other species but have not yet been widely used in dromedaries. Recent studies are starting to use them to evaluate gene diversity at the level of specific genes (Lado et al., 2020a).

The development of SNP markers that cover the coding part of genome are also needed to understand the relationship between genetic and phenotypic variations in camels or other species. In these sites, SNPs are divided in synonymous, without any effect in the amino acid sequence and nonsynonymous affecting the amino acid sequence of protein (Sushma et al., 2014). In this study, the authors used direct ultra-high-throughput sequencing (RNA-Seq approach) for mapping and quantifying transcripts developed to analyze global gene expression in heart and kidney of *C. dromedarius* and C. *bactrianus* for the identification of gene and identifying polymorphism in camel at nucleotide level (SNPs). They identified 24 and 10 nonsynonymous SNPs in *C. dromedarius* in the heart and the kidney, respectively. In another study, Abd El-Aziem et al. (2015) detected one SNP (C/T) in Growth Hormone gene among five Egyptian camel breeds. They have concluded that this SNP can be used as a marker for the genetic diversity between the camel breeds. The allele C is related to higher growth rate and can be used in marker assisted selection for the enhancement of growth rate should it be needed in camel breeding program.

In another context, Ruiz et al., (2015) developed a diagnostic panel of SNPs to identify the hybridization patterns in camels with uncertain origins, to support hybrid breeding management and to detect potential rare dromedary introgression in the remaining wild Bactrian camels in Mongolia and China. Recently, in a genome-wide association study, Bitaraf Sani et al. (2021), using genotyping by sequencing, were able to identify 99 SNP markers that could be associated with important traits to improve camel breeding, namely birth, weight, daily gain and body weight.

Lado et al. (2020b), sequenced 22,721 SNP markers in order to understand how the dromedary populations history and the environmental adaptation could be influenced by human-induced migration patterns and historic demographic changes. The use of these molecular markers helped to understand the routes of domestication and how genetic diversity was built through the centuries. A genetic mixture within continental populations between Asia and Africa was detected.

In terms of genetic studies in livestock, SNPs have three main advantages over microsatellites: a more precise estimates of population-level diversity, a higher power to identify groups in clustering methods, and an ability to consider local adaptation (Zimmerman et al., 2020). These advantages offer better opportunities of using SNP in parentage control, GWAS for traits of economic interest and the diagnosis of genetic diseases by developing specific ships for dromedary camel.

DROMEDARY CAMEL PARENTAGE CONTROL

In the biannual conference of the International Society for Animal Genetics, the Camel Comparison Test (CCT) for the laboratory developing parentage test for dromedary was discussed in a workshop named "Applied Genetics and Genomics in Other Species of Economic Interest" since 2016. For the parentage control, a minimum of eight microsatellites markers are needed (LCA19, LCA37, LCA56, LCA65, LCA66, LCA8, YWLL29, YWLL44) and are named Core panel in the CCT. For this test in 2016, twenty-four samples and one reference sample were submitted to all the laboratories participating in the CCT. A deadline was given and the participants were asked to send their results for the core panel which would be considered in the ranking system for an accreditation. It was possible to add nine other microsatellite markers in the back-up Panel (LCA24, LCA77, LCA99, LGU49, VOLP3, VOLP32, VOLP59, YWLL08,

TABLE 2 | Main studies using microsatellites realized in dromedary camels for the evaluation of genetic diversity.

Species studied	Regions or countries	Number of microsatellites used	References
Dromedary camel and Bactrian camels	Kenya and China	20	Jianlin et al. (2000)
Dromedary camel and Bactrian camels	Kenya, Pakistan, United Arab Emirates, Saudi Arabia and China	14	Mburu et al. (2003)
Dromedary camel	India	6	Gautam et al. (2004)
Alpaca and Dromedary camel	South Africa, Namibia and Botswana	16	Nolte et al. (2005)
Dromedary camel	India	16	Mehta and Sahani, (2007)
Dromedary camel	India	23	Vijh et al. (2007)
Dromedary camel	Australia and africa	28	Spencer and Woolnough (2010)
Dromedary camel	Canarias, Algeria, Kenya, Saudi Arabia, UAE and Pakistan	13	Schulz et al. (2010)
Dromedary camel	Tunisia	6	Ould Ahmed et al. (2010)
Dromedary camel	Egypte	3	Mahrous et al. (2011)
Dromedary camel	Morocco	7	Piro et al. (2011)
Dromedary camel	Saudi Arabia	15	Mahmoud et al. (2012)
Dromedary camel	Libya	6	Bakory (2012)
Dromedary camel	India	12	Banerjee et al. (2012)
Dromedary camel	Sudan, Qatar, Somalia and Chad	25	Hashim et al. (2014)
Dromedary camel	India	40	Mehta (2014)
Dromedary camel	India	29	Sushma et al, (2014)
Dromedary camel	Tunisia	7	Nouaïria et al. (2015)
Dromedary camel	Sudan	12	Eltanany et al. (2015)
Dromedary camel	India	25	Sushma et al. (2015)
Dromedary camel	Saudi Arabia	18	Almathen et al. (2016)
Dromedary camel	Algeria and Egypt	20	Cherifi et al. (2017)
Dromedary camel	India	11	Tyagi et al. (2017)
Dromedary camel	Morocco	19	Piro et al. (2018)
Dromedary camel	Ethiopia	6	Legesse et al. (2018)
Dromedary camel	Pakistan	12	Tanveer et al. (2021)

YWLL36) and 30 other microsatellites can be used also to compare results between participating laboratories. In the last CCT, the overall marker concordance among laboratories was good (>95%) for six of the eight markers in the Core Panel. Only two markers showed a lower concordance: LCA37 (86%) because of missed alleles and wrong allele binning and YWL44 (87%) because of a missed allele.

Different studies tried to use different panels of microsatellites to evaluate the exclusion power for their dromedary populations. In 2005, Nolte et al. used 16 microsatellites for the evaluation of genetic diversity and for parentage analysis in cattle. Five calves were compared to their known parents and an unrelated male. It was successful in all cases. Only one mismatch at the VOLP08 locus was observed in one calf. They conclude that this mismatch is possibly due to a mutation or a laboratory typing error. Mariasegaram et al. (2002) identified eight dromedary specific microsatellites as described previously and used them in racing dromedary camels for the evaluation of the probability of exclusion for a future use in parentage control. The total PE was high and estimated at 0.992. In Australia, Spencer et al. (2010) conducted a study to evaluate and apply microsatellite multiplexes to develop a parentage control for racing dromedaries. They randomly sampled dromedaries in three geographically separated regions. 17 loci from 700 unrelated dromedary samples of both sexes, including individuals from Australia (n = 620), United Arab Emirates (n = 53) and Africa (n = 16) were used to build a database of unrelated adults. The 17 microsatellites were separated in three multiplex reactions and a high probability of parentage exclusion (PE = 0.9999) was found. PE is an index allowing quantification of the percentage of incorrect detected affiliations. They conclude that this multiplex system clearly demonstrates the importance of DNA testing to ensure accurate identification and allocation of parentage in reproductive centers.

In Morocco, PE and the probability of identity (PI: probability to take hazardly two individuals with the same genotype) were determinated for five camel populations by using seven microsatellites markers (VOLP03 YWLL44, YWLL59, CVRL01, CVRL05, CVRL06 and CVRL07. PE varied from 95 to 97% with one parent and was higher than 99.99% with two parents for the five populations. The PI of individual camels varied between $1/8.10^{-6}$ and $1/55.10^{-6}$ (Piro et al., 2011). They noticed that, among the five studied populations, the CVRL1 followed by VOLP3 loci were the most effective loci to exclude the false parents, while YWLL44 and YWLL59 loci present the lowest PE and PI.

In Tunisia, Nouaïria et al. (2015) evaluated twenty microsatellites to develop a microsatellite panel for parentage control in 130 dromedary camels. They concluded that a minimum of 12 microsatellites are needed to obtain a PE higher than 99% when one parent is unknown parent and 98% when both parents are unknown in dromedary camel.

With the actual development of new molecular biology techniques and with regard to several other species, it will probably be obvious that the establishment of parentage control using SNPs will soon be proposed by ISAG. Therefore, studies to choose the most informative SNPs on this subject should be carried out.

GENETIC STRUCTURE OF DIFFERENT BREEDS AROUND THE WORLD

Studies were conducted around the world on the genetic characterization and diversity of dromedary camels using microsatellites or SNPs (Lado et al., 2020b). Each study used a set of microsatellite markers to identify genetic difference between dromedary camels from different regions, in a specific country or through comparing breeds from different countries or continents. These different studies showed that it was quite difficult to establish a significant genetic differentiation between different types or breeds living in different regions and even with a different phenotype and most of the results suggested the probable existence of relatively less genetic variation in dromedary camels and that morphologic and regional distribution criterions are not enough for the dromedary camel classification or characterization (Jianlin et al., 2000; Mburu et al., 2003; Nolte et al., 2005; Vijh et al., 2007; Schulz et al., 2010; Ould Ahmed et al., 2010; Spencer and Woolnough, 2010; Piro et al., 2011; Mahrous et al., 2011; Mahmoud et al., 2012; Sushma et al., 2014; Sushma et al.,2015; Eltanany et al., 2015; Cherifi et al., 2017; Piro et al., 2018; Legesse et al., 2018; Tanveer et al., 2021). Not all the microsatellites markers used were polymorphic and the following markers showed a high level of polymorphism: CVRL01, CVRL07, CMS16, CMS50, CMS121, LCA56, LCA63, LCA65, LCA66, LCA70, VOLP03, VOLP10, VOLP32, VOLP50, VOLP55, VOLP67, YWLL08, YWLL09 (Jianlin et al., 2000; Nolte et al., 2005; Tyagi et al., 2017).

A higher level of cross-breeding or a gain of genetic variation following genetic drift subsequent to migration from one area to another, and traditional herding practices and dromedary camel particular history of domestication likely occurring from a bottlenecked and geographically confined wild progenitor can explain the poor genetic variation (Mahmoud et al., 2012; Cherifi et al., 2017; Piro et al., 2018). Hashim et al., (2014) indicated also that there was a relationship between the genetic makeup and geographical distributions and also between the genetic makeup and the phenotypic characteristic. Moreover, it is worth mentioning that most camelid populations are named as breeds after their tribal affiliation or the geographic location where they are found (Rosati et al., 2005). Only one study reported the existence of genetic variation amongst the four camel breeds studied in Libya (Bakory, 2012) and Banerjee et al. (2012) compared six Indian camel populations and showed a genetic differentiation among themselves due to selection pressure and breeding for specific economic traits and that the camel populations of India and South Africa are very well differentiated. Mehta, (2014) used more microsatellites (40) to analyze the genetic and demographic bottleneck of Indian camel breeds (Bikaneri, Jaisalmeri, Kachchhi and Mewari camel breeds) and they indicated that allelic polymorphism was observed only in 20 loci and a higher genetic variation was observed in most numerous Bikaneri breed. In comparison with the other three Indian dromedary breeds, the Mewari camels had relatively higher genetic distance from other breeds. These authors concluded that their results showed a demographic bottleneck in the four Indian dromedary populations and an appropriate conservation and improvement program is necessary.

The main studies using microsatellites and undertaken in several countries are reported in **table 2**, with emphasis on the number of markers used, the species and the regions studied. Recently, Lado et al., (2020b) used 22K SNPS markers for the evaluation of genetic diversity of African and Asian dromedary camels and showed also the existence of moderate genome-wide diversity and a low population structure.

CONCLUSION

The dromedary camel is a species of great zootechnical interest in several countries. Generally, camel genetic characterization studies have mainly been completed using microsatellite markers. However, to better characterize certain production interesting traits, other techniques using whole genome sequencing and SNPs are necessary. Currently, these techniques are under investigation to assess relation with some phenotypic trait. These studies will allow better understanding of this magnificent exotic species, better conservation of the different breeds and types of dromedary and would certainly provide better selection alternatives for the breeders.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Almathen, F., Charruau, P., Mohandesan, E., Mwacharo, J. M., Orozco-terWengel, P., Pitt, D., et al. (2016). Ancient and Modern DNA Reveal Dynamics of Domestication and Cross-Continental Dispersal of the Dromedary. *Proc. Natl. Acad. Sci. USA*. 113 (24), 6707–6712. doi:10.1073/pnas.1519508113

Arthofer, W., Heussler, C., Krapf, P., Schlick-Steiner, B. C., and Steiner, F. M. (2018). Identifying the Minimum Number of Microsatellite Loci Needed to Assess Population Genetic Structure: A Case Study in Fly Culturing. *Fly.* 12 (1), 13–22. doi:10.1080/19336934.2017.1396400

Bakory, M. A. (2012). Preliminary Genetic Study of Libyan Camels Using DNA Based RAPD and SSR Markers. *Anim. Sci. J.* 3 (1), 06–10.

Balmus, G., Trifonov, V. A., Biltueva, L. S., O'Brien, P. C. M., Alkalaeva, E. S., Fu, B., et al. (2007). Cross-Species Chromosome Painting Among Camel, Cattle, Pig and Human: Further Insights into the Putative Cetartiodactyla Ancestral Karyotype. *Chromosome Res.* 15, 499–514. doi:10.1007/s10577-007-1154-x

Banerjee, P., Joshi, J., Sharma, U., and Vijh, R. K. (2012). Population Differentiation in Dromedarian Camel: A Comparative Study of Camel Inhabiting Extremes of Geographical Distribution. *Int. J. Anim. Vet. Adv.* 4 (2), 84–92. Maxwell Scientific Organization.

Bitaraf Sani, M., Zare Harofte, J., Banabazi, M. H., Esmaeilkhanian, S., Shafei Naderi, A., Salim, N., et al. (2021). Genomic Prediction for Growth Using a Low-Density SNP Panel in Dromedary Camels. *Sci. Rep.* 11, 7675. doi:10.1038/s41598-021-87296-7

Burger, P. A., and Palmieri, N. (2014). Estimating the Population Mutation Rate From Ade novoAssembled Bactrian Camel Genome and Cross-Species Comparison With Dromedary ESTs. *Jhered.* 105 (6), 933–940. doi:10.1093/jhered/est005

Burger, P. A. (2016). The History of Old World Camelids in the Light of Molecular Genetics. *Trop. Anim. Health Prod.* 48, 905–913. doi:10.1007/s11250-016-1032-7

Cherifi, Y. A., Gaouar, S. B. S., Guastamacchia, R., El-Bahrawy, K. A., Abushady, A. M. A., Sharaf, A. A., et al. (2017). Weak Genetic Structure in Northern African Dromedary Camels Reflects Their Unique Evolutionary History. *Plos One.* 12 (1), e0168672. doi:10.1371/journal.pone.0168672

Ciani, E. (2018). "From Knowledge to Practice: Ongoing and Future Steps in Camel Genetic Management," in Proceedings of the 5th Conference of the International Society of Camelid Research and Development ISOCARD, "Recent Advances in camelids biology, health and production".

Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A., and Solignac, M. (1999). New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals. *Genetics.* 153, 1989. doi:10.1093/genetics/153.4.1989

Eltanany, M., Elfaroug Sidahmed, O., and Distl, O. (2015). Assessment of Genetic Diversity and Differentiation of Two Major Camel Ecotypes (*Camelus dromedarius*) in Sudan Using Microsatellite Markers. *Arch. Anim. Breed.* 58 (2), 269–275. doi:10.5194/aab-58-269-2015

Evdotchenko, D., Han, Y., Bartenschlager, H., Preuss, S., and Geldermann, H. (2003). New Polymorphic Microsatellite Loci for Different Camel Species. *Mol. Ecol. Notes.* 3, 431–434. doi:10.1046/j.1471-8286.2003.00477.x

Fitak, R. R., Mohandesan, E., Corander, J., and Burger, P. A. (2015). The De Novo Genome Assembly and Annotation of a Female Domestic Dromedary of North African Origin. *Mol. Ecol. Resour.* 16, 314–324. doi:10.1111/1755-0998.12443

Gautam, L., Mehta, S. C., Gahlot, R. S., and Gautam, K. (2004). Genetic Chracterisation of Jaisalmeri Camel Using Microsatellite Markers. *Indian J. Biotech.* 3, 457–459.

Guerouali, A., and Acharbane, R. (2004). "Camel Genetic Resources in Morocco," in FAO-ICAR. Seminar on Camelidis, Sousse, Tunisia, May 30.

Hale, M. L., Burg, T. M., and Steeves, T. E. (2012). Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals Per Population Is Enough to Accurately Estimate Allele Frequencies. *Plos One.* 7, e45170. doi:10.1371/journal.pone.0045170

Hashim, W. M., Ahmed, S. E. S., Makkaw, A. A., Yousif, G. M., and Mehta, S. C. (2014). Twenty-Five Polymorphic Microsatellite Loci Used to Genotype Some Camel Types & Subtypes From Sudan, Qatar, Chad and Somalia. *J. Camel Pract. Res.* 21 (2), 127–135. doi:10.5958/2277-8934.2014.00024.1

ISAG-FAO (2004). *Measurement of Domestic Animal Diversity (MoDAD): Recommended Microsatellite Markers. New Microsatellite Marker Sets – Recommendations of Joint ISAG/FAO Standing Committee*. Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Available at: http://www.fao.org/3/aq569e/aq569e.pdf/

Ji, R., Cui, P., Ding, F., Geng, J., Gao, H., Zhang, H., et al. (2009). Monophyletic Origin of Domestic Bactrian Camel (*Camelus bactrianus*) and its Evolutionary Relationship With the Extant Wild Camel (*Camelus bactrianus* Ferus). *Anim. Gen.* 40, 377–382. doi:10.1111/j.1365-2052.2008.01848.x

Jianlin, H., Mburu, D., Ochieng, J., Kaufmann, B., Rege, J. E. O., and Hanotte, O. (2000). Application of New World Camelidae Microsatellite Primers for Amplification of Polymorphic Loci in Old World Camelids. *Anim. Genet.* 31 (6), 404–406. doi:10.1111/j.1365-2052.2000.00683.pp.x

Kadwell, M., Fernandez, M., Stanley, H. F., Baldi, R., Baldi, J. C., Rosadio, R., et al. (2001). Genetic Analysis Reveals the Wild Ancestors of the Llama and the Alpaca. *Proc. R. Soc. Lond. B.* 268, 2575–2584. doi:10.1098/rspb.2001.1774

Lado, S., Elbers, J. P., Rogers, M. F., Melo-Ferreira, J., Yadamsuren, A., Corander, J., et al. (2020a). Nucleotide Diversity of Functionally Different Groups of Immune Response Genes in Old World Camels Based on Newly Annotated and Reference-Guided Assemblies. *BMC Genomics.* 21, 606. doi:10.1186/s12864-020-06990-4

Lado, S., Elbers, J. P., Doskocil, A., Scaglione, D., Trucchi, E., Banabazi, M. H., et al. (2020b). Genome-Wide Diversity and Global Migration Patterns in Dromedaries Follow Ancient Caravan Routes. *Commun. Biol.* 3, 387. doi:10.1038/s42003-020-1098-7

Lang, K. D., Wang, Y., and Plante, Y. (1996). Fifteen Polymorphic Dinucleotide Microsatellites in Llamas and Alpacas. *Anim. Genet.* 27, 293. doi:10.1111/j.1365-2052.1996.tb00502.x

Lee, M. A., Keane, O. M., Glass, B. C., Manley, T. R., Cullen, N. G., Dodds, K. G., et al. (2006). Establishment of a Pipeline to Analyse Non-Synonymous SNPs in *Bos taurus*. *BMC. Genomics.* 7, 298. doi:10.1186/1471-2164-7-298

Legesse, Y. W., Dunn, C. D., Mauldin, M. R., Garza, N. O., Rowden, G. R., Mekasha, Y., et al. (2018). Morphometric and Genetic Variation in Eight Breeds of Ethiopian Camels (*Camelus dromedarius*). *J. Anim. Sc.* 96 (12), 4925–4934. doi:10.1093/JAS/SKY351

Mahmoud, A. H., Alshaikh, M. A., Aljumaah, R. S., and Mohammed, O. B. (2012). Genetic Variability of Camel (*Camelus dromedarius*) Populations in Saudi Arabia Based on Microsatellites Analysis. *Afr. J. Biotechnol.* 11 (51), 11173–11180. doi:10.5897/ajb12.1081

Mahrous, K. F., Ramadan, H. A., Abdel-Aziem, S. H., and Abd-El MordyandHemdan, M. D. M. (2011). Genetic Variations Between Camel Breeds Using Microsatellite Markers and RAPD Techniques. *J. Appl. Biosci.* 39, 2626–2634. Available at: http://m.elewa.org/JABS/2011/39/5.pdf.

Mariasegaram, M., Pullenayegum, S., Jahabar Ali, M., AliShah, R. S., WerneryShah, U., et al. (2002). Isolation and Characterization of Eight Microsatellite Markers in *Camelus dromedarius* and Cross-Species Amplification in *C. bactrianus* and *Lama pacos*. *Anim. Genet.* 33, 385–387. doi:10.1046/j.1365-2052.2002.00896_6.x

Mburu, D. N., Ochieng, J. W., Kuria, S. G., Jianlin, H., Kaufmann, B., Rege, J. E. O., et al. (2003). Genetic Diversity and Relationships of Indigenous Kenyan Camel (*Camelus dromedarius*) Populations: Implications for Their Classification. *Anim. Genet.* 34, 26–32. doi:10.1046/j.1365-2052.2003.00937.x

Mehta, S. C. (2014). Genetic and Demographic Bottleneck Analysis of Indian Camel Breeds by Microsatellite Markers. *Trop. Anim. Health Prod.* 46 (8), 1397–1406. doi:10.1007/s11250-014-0653-y

Mehta, S. C., and Sahani, M. S. (2007). Microsatellite Markers for Genetic. Characterization of Bikaneri Camel. *Indian J. Anim. Sci.* 77 (6), 509–512. doi:10.5958/2277-8934.2017.00037.6

Ming, L., Yuan, L., Yi, L., Ding, G., Hasi, S., Chen, G., et al. (2020). Whole-Genome Sequencing of 128 Camels Across Asia Reveals Origin and Migration of Domestic Bactrian Camels. *Commun. Biol.* 3, 1–11. doi:10.1038/s42003-019-0734-6

Mohandesan, E., Fitak, R. R., Corander, J., Yadamsuren, A., Chuluunbat, B., Abdelhadi, O., et al. (2017). Mitogenome Sequencing in the Genus Camelus Reveals Evidence for Purifying Selection and Long-Term Divergence between Wild and Domestic Bactrian Camels. *Sci. Rep.* 7, 9970. doi:10.1038/s41598-017-08995-8

Muneeb, M. M. (2014). A Review of Microsatellite Marker Usage in the Assessment of Genetic Diversity of camelus. *Iranian J. Appl. Anim. Sci.* 5 (1), 1–4. doi:10.1016/j.jevs.2020.103358

Nolte, M., Kotze, A., Van der Bank, F., and Grobler, J. (2005). Microsatellite Markers Reveal Low Genetic Differentiation Among Southern African *Camelus Dromedarius* Populations. *South Afr. J. Anim. Sci.* 35, 152–161.

Nouaïria, G., Cassandro, M., Hammadi, M., Korchani, T., and Yahyaoui, M. H. (2018). "Mitochondrial DNA Analysis of Dromedary Camel (Camelus Dromedarius) Population in Tunisia," in Proceedings of the 5th Conference of the International Society of Camelid Research and Development ISOCARD, "Recent Advances in camelids biology, health and production", 124–125.

Nouaïria, G., Kdidi, S., Salah, R., Hammadi, M., Khorchani, T., and Yahyaoui, M. (2015). Assessing Genetic Diversity of Three Tunisian Dromedary Camel (*Camelus dromedarius*) Sub Populations Using Microsatellite Markers. *Emir. J. Food Agric.* 27 (4), 362–366. doi:10.9755/ejfa.v27i4.19258

Obreque, V., Coogle, L., Henney, P. J., Bailey, E., Mancilla, R., García-Huidobro, J., et al. (1998). Characterization of 10 Polymorphic Alpaca Dinucleotide Microsatellites. *Anim. Genet.* 29, 461–462.

Ould Ahmed, M., Ben Salem, F., Bedhiaf, S., Rekik, B., and Djemali, M. (2010). Genetic Diversity in Tunisian Dromedary (*Camelus dromedarius*) Populations Using Microsatellite Markers. *Livestock Sci.* 132, 182–185. doi:10.1016/j.livsci.2010.05.016

Penedo, M. C., Caetano, A. R., and Cordova, K. (1999). Eight Microsatellite Markers for South American Camelids. *Anim. Genet.* 30, 166–167. doi:10.1046/j.1365-2052.1999.00382-8.x

Petkovski, E., Keyser-Tracqui, C., Hienne, R., and Ludes, B. (2005). SNPs and MALDI-TOF MS: Tools for DNA Typing in Forensic Paternity Testing and Anthropology. *J. Forensic Sci.* 50 (3), 1–7. doi:10.1520/jfs2004245

Piro, M., Bouazzati, O., Bengoumi, M., El Allali, K., and Achaaban, M. R. (2011). Genetic Characterisation of Moroccan Camel Populations Using Microsatellites Markers. *J. Camel Pract. Res.* 18, 167–172.

Piro, M., Khattaby, N., Laghouaouta, H., Badaoui, B., and Boujenane, I. (2018). "Genetic Variability of Three Camel (*Camelus Dromedarius*) Populations in the South of Morocco Based on Microsatellites Analysis," in Proceedings of the 5th Conference of the International Society of Camelid Research and Development ISOCARD, "Recent Advances in camelids biology, health and production".

Riva, A., and Kohane, I. S. (2002). SNPper: Retrieval and Analysis of Human SNPs. *Bioinformatics.* 18 (12), 1681–1685. doi:10.1093/bioinformatics/18.12.1681

Rosati, A., Tewolde, A., and Mosconi, C. (2005). *A Review on Developments and Research in Livestock Systems. Animal Production and Animal Science Worldwide*. Wageningen Academic Publishers. doi:10.3920/978-90-8686-564-2

Rout, P. K., Joshi, M. B., Mandal, A., Laloe, D., Singh, L., and Thangaraj, K. (2008). Microsatellite-Based Phylogeny of Indian Domestic Goats. *BMC Genet.* 9, 11. doi:10.1186/1471-2156-9-11

Ruiz, E., Mohandesan, E., Fitak, R. R., and Burger, P. A. (2015). Diagnostic Single Nucleotide Polymorphism Markers to Identify Hybridization Between Dromedary and Bactrian Camels. *Conservation Genet. Resour.* 7, 329–332. doi:10.1007/s12686-015-0420-z

Ruvinskiy, D., Larkin, D. M., and Farré, M. (2019). A Near Chromosome Assembly of the Dromedary Camel Genome. *Front. Genet.* 10, 32. doi:10.3389/fgene.2019.00032

Sadder, M., MigdadiAl-Haidary, H. A., AL-Haidary, A., and Okab, A. (2015). Identification of Simple Sequence Repeat Markers in the dromedary(*Camelus dromedarius*) Genome by Next-Generation Sequencing. *Turk J. Vet. Anim. Sci.* 39, 218–228. doi:10.3906/vet-1402-72

Sasse, J., Mariasegaram, M., Babu, R., Kinne, J., and Wernery, U. (2000). South American Camelid Microsatellite Amplification in *Camelus Dromedarius*. *Anim. Genet.* 31, 75–76. doi:10.1111/j.1365-2052.2000.579-10.x

Schulz, U., Tupac-Yupanqui, I., Martínez, A., Méndez, S., Delgado, J. V., Gómez, M., et al. (2010). The Canarian Camel: a Traditional Dromedary Population. *Diversity.* 2 (4), 561–571. doi:10.3390/d2040561

Sekena, H. A. E.-A., Heba, A. M. A. E.-K., Sally, S. A., and Othman, E. O. (2015). Detection of MspI Polymorphism and the Single Nucleotide Polymorphism (SNP) of GH Gene in Camel Breeds Reared in Egypt. *Afr. J. Biotechnol.* 14 (9), 752–757. doi:10.5897/AJB2014.14374

Silbermayr, K., Orozco-terWengel, P., Charruau, P., Enkhbileg, D., Walzer, C., Vogl, C., et al. (2010). High Mitochondrial Differentiation Levels Between Wild and Domestic Bactrian Camels: a Basis for Rapid Detection of Maternal Hybridization. *Anim. Genet.* 41, 315–318. doi:10.1111/j.1365-2052.2009.01993.x

Spencer, P. B. S., Wilson, K. J., and Tinson, A. (2010). Parentage Testing of Racing Camels (*Camelus Dromedarius*) Using Microsatellite DNA Typing. *Anim. Genet.* 41 (6), 662–665. doi:10.1111/j.1365-2052.2010.02044.x

Spencer, P., and Woolnough, A. (2010). Assessment and Genetic Characterisation of Australian Camels Using Microsatellite Polymorphisms. *Livestock Sci.* 129, 241–245. doi:10.1016/j.livsci.2010.01.006

Sushma, P., Sharique, A. A., Banerjee, P., Joshi, J., Sharma, U., and Vijh, R. K. (2014). Genetic Characterization of Malvi Camel Using Microsatellite Markers. *IOSR J. Agric. Vet. Sci. (Iosr-javs) e-ISSN.* 7, 65–70.

Sushma, P., Sharique, A. A., Banerjee, P., Joshi, J., Sharma, U., and Vijh, R. K. (2015). Population Genetic Structure of the Camel, *Camelus Dromedarius* Based on Microsatellite Loci: Knock-On Effect for Conservation. *Biosci. Biotech. Res. Comm.* 8 (2), 153–160.

Tanveer, H., Fiaz, H., Abdul, W., Masroor, E. B., Muneeb, M. M., and Faiz, M. M. T. M. (2021). Microsatellite Markerbased Genetic Diversity in Mareecha and Barela Breeds of Dromedary Camel From Pakistan. *Veterinarska Stanica.* 52 (3). doi:10.46419/vs.52.3.6

Tyagi, K., Gautam, L., Waiz, H. A., Mehta, S. C., and Beniwal, B. K. (2017). Genetic Characterisation of Bikaneri Camel Using Microsatellite Markers. *J. Camel Pract. Res.* 24 (2), 1–4. doi:10.5958/2277-8934.2017.00037.6

Vijh, R., Tantia, M., Mishra, B., and Bharani Kumar, S. (2007). Genetic Diversity and Differentiation of Dromedarian Camel of India. *Anim. Biotechnol.* 18 (2), 81–90. doi:10.1080/10495390600648741

Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., et al. (2014). Camelid Genomes Reveal Evolution and Adaptation to Desert Environments. *Nat. Commun.*, 5, 5188. doi:10.1038/ncomms6188

Zimmerman, S. J., Aldridge, C. L., and Oyler-McCance, S. J. (2020). An Empirical Comparison of Population Genetic Analyses Using Microsatellite and SNP Data for a Species of Conservation Concern. *BMC Genomics.* 21, 382. doi:10.1186/s12864-020-06783-9

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fgene.2021.724058

Estimation of Genetic Parameters of Type Traits in First Parity Cows of the Autochthonous Cika Cattle in Slovenia

Mojca Simčič[1], Barbara Luštrek[1], Miran Štepec[1], Betka Logar[2] and Klemen Potočnik[1]*

[1]*Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia,* [2]*Agricultural Institute of Slovenia, Ljubljana, Slovenia*

The aim of this study was to estimate genetic parameters of 26 individual and four composite type traits in first parity Cika cows. An analysis of variance was performed with the generalized linear model procedure of the SAS/STAT statistical package, where the fixed effects of year of recording, cow's age at recording and days after calving as a linear regression were included in the model. The variance components for the direct additive genetic effect and the herd effect in all type traits were estimated using the REML method in the VCE-6 software package. The estimated heritabilities ranged from 0.42 to 0.67 for the measured body frame traits, from 0.36 to 0.80 for the scored autochthonous traits, from 0.11 to 0.61 for the scored body frame traits, and from 0.20 to 0.47 for the scored udder traits. The estimated heritabilities for the composite traits called "autochthonous characteristics", "muscularity", "body frame" and "udder" were 0.55, 0.19, 0.19, and 0.26, respectively. The estimated genetic correlations among the measured body frame traits were positive and high, while the majority of them among the scored body frame traits were low to moderate. The estimated proportions of variance explained by the herd effect for the composite traits "autochthonous characteristics," "muscularity," "body frame" and "udder" were 0.09, 0.28, 0.14, and 0.10, respectively. The estimated heritabilities for the type traits of first parity Cika cows were similar to those reported for other breeds where breeding values have been routinely predicted for a long time. All estimated genetic parameters are already used for breeding value prediction in the Cika cattle population.

Keywords: Cika cattle, first parity cow, type traits, genetic parameters, variance components

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman

Reviewed by:
Hugo Oswaldo Toledo-Alvarado, National Autonomous University of Mexico, Mexico
Cristina Sartori, University of Padua, Italy

****Correspondence:***
Mojca Simčič
mojca.simcic@bf.uni-lj.si

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 June 2021*
Accepted: *27 October 2021*
Published: *22 November 2021*

Citation:
Simčič M, Luštrek B, Štepec M, Logar B and Potočnik K (2021) Estimation of Genetic Parameters of Type Traits in First Parity Cows of the Autochthonous Cika Cattle in Slovenia. Front. Genet. 12:724058. doi: 10.3389/fgene.2021.724058

INTRODUCTION

Cika cattle is a Slovenian autochthonous dual-purpose breed. The breeding goal is to preserve the original type traits of the breed and to prevent an increase of inbreeding. Cika cattle are widespread all over Slovenia, especially in their regions of origin (Bohinj, Kamnik, Tolmin). The coat color pattern is red pied sided, which is typical for this breed and very different from other cattle breeds in Slovenia. Some Cika animals have a coat color pattern similar to Pinzgauer cattle and some to Tux-Zillertaler cattle from Austria (Sambraus, 1999). Based on microsatellite genotyping, Cika cattle form an Eastern Alpine breed cluster with Pinzgauer and Pustertaler cattle. Cika cattle are considered as an authentic and valuable genetic resource (Felius et al., 2011; Simčič et al., 2013). The possibility of excluding admixed animals (sires and sire dams) from the Cika cattle breeding program using SNP haplotypes has been reported in detail elsewhere (Simčič et al., 2015a,b).

Cika cattle breeding is based on the recording of type traits of all first parity cows (Žan Lotrič et al., 2010). The age of first parity cows at the recording day is restricted to a minimum of 560 days and there are no restrictions on the maximum age. Recording time is adjusted to the rearing technology, so the animals are not recorded during the grazing season. According to the breeding program, the recording of first parity cows is planned to take place 15–120 days after calving, which is often not possible due to the grazing season in the higher mountain pastures.

The current population size of Cika cattle is 5,531 animals (www.fao.org/dad-is), which is lower than in all other breeds in Slovenia. Therefore, all breeding animals are scored by a single expert. Compared to the small population size, a large number of sires (98 in the year 2020) is used for natural service and artificial insemination in the population. The effective population size is estimated to be 117 animals. Of the animals born in 2016, 96.6% had complete pedigree data in the second generation, while 65.7% of the animals had a complete pedigree in the sixth generation. The inbreeding coefficient in the population is estimated to be 2.1%.

Genetic parameters for type traits in dairy cattle breeds have been reported in first parity cows of Holstein-Friesian (Van Dorp et al., 1998; Rupp and Boichard, 1999; Larroque and Ducrocq, 2001; Neuenschwander et al., 2005; de Haas et al., 2007; Němcová et al., 2011), Jersey (Thomas et al., 1985; Norman et al., 1988; Rogers et al., 1991), Brown Swiss (Norman et al., 1988; Dal Zotto et al., 2007; de Haas et al., 2007; Samore et al., 2010), Red Holstein (de Haas et al., 2007), Ayrshire, Guernsey and Shorthorn (Norman et al., 1988), where correlations among type traits, milk yield, fertility and longevity were estimated.

In first parity cows of beef breeds, genetic parameters for type traits were studied in Belgian blue cattle with double muscling by Gengler et al. (1995), and in Spanish Asturiana de los Valles cattle by Guitiérrez and Goyache (2002), and Guitiérrez et al. (2002). Forabosco et al. (2004) indirectly estimated longevity using correlated type traits. Genetic parameters for type traits in first parity cows of some autochthonous cattle breeds were reported as well, e.g. in the Italian autochthonous breeds Chianina (Forabosco et al., 2004), Piemontese (Mantovani et al., 2010), Rendena (Mazza and Mantovani, 2012; Mazza et al., 2014), Cabanina (Comunod et al., 2013), Valdostana cattle (Mazza et al., 2015), Alpine Grey cattle (Mancin et al., 2021) and Italian Simmental (Frigo et al., 2013).

The aim of this pilot study was to estimate genetic parameters for the measured and scored type traits included in the routine breeding values estimation of Cika cows according to the breeding program.

MATERIALS AND METHODS

Data

The data were obtained from the Central database for cattle kept by the Agricultural Institute in Slovenia, which included information from the type traits recording of Slovenian first parity Cika cows scored in the years from 2006 to 2019. We assumed that the type trait recording was intended for first parity cows, so we limited the age at recording to a maximum of 1,460 days or 4 years as recommended by de Haas et al. (2007) and a minimum of 607 days. On average, the animals were 1,030.7 ± 167.3 days old at recording. In addition, we excluded all cows younger than 4 years of age that were scored after the second parity. First parity cows that were not in the recommended stage of lactation on the day of recording according to the breeding program were not excluded from further analysis. Consequently, the recorded first parity cows were one to 446 days after calving, 227.5 ± 112.7 days on average. After applying these limitations, 1,815 first parity cows were included in the final analysis (**Table 1**).

All first parity cows (1,815) with type traits and their known ancestors were included in the additive relationship matrix among all animals (2,953 animals in total). Almost 98% of the first parity cows had a known sire and 96% had a known dam. In total, 95% of the first parity cows had both parents known.

Estimation of Variance Components

To investigate the non-genetic effects to be included in the model, an analysis was performed using the GLM (generalized linear model) procedure in the SAS/STAT statistical package (version 9.4, SAS Institute Inc, 2001) with the statistical model

$$y_{ijk} = \mu + Y_i + b_I\left(x_{ijk} - \bar{x}\right) + b_{II}\left(z_{ijk} - \bar{z}\right) + e_{ijk}$$

where y_{ijk} was the type trait, μ was the mean of the population, Y_i was the fixed effect of the recording year (i = 2006, . . . , 2019), b_I was the linear regression coefficient for the age at recording, x_{ijk} was the age at recording (days), $\bar{x}$ was the mean age at recording, b_{II} was the linear regression coefficient for days after calving, z_{ijk} was the number of days after calving, $\bar{z}$ was the mean number of days after calving and e_{ijk} was an error.

The matrix notation for the complete model including both non-genetic and genetic effects was expressed as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Wq} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where y is an N x 1 vector of observations, β is the vector of systematic fixed effects of order p, q is the vector of herd effect considered as random effect, u is the vector of animal effects with order m, and e is the vector of residual effects. Likewise, X, W and Z are the corresponding incidence matrices with the appropriate dimensions.

The variance components and heritability for each individual and composite trait were estimated with the REML method in the VCE-6 software package (Groeneveld et al., 2010).

The analyzed type traits were divided into six groups according to specific body regions (Simčič et al., 2016). The measured body frame traits (wither height, rump height, body length, chest girth) were assigned to Group 1. The scored autochthonous traits were divided into three groups: head characteristics (head length, head nobility, eyes, horn base circumference, horns length, horns direction) were assigned to Group 2, neck characteristics (neck, dewlap) were assigned to Group 3, and coat color traits (coat color, back stripe, rear legs stripe, front legs stripe) were assigned to Group 4. The scored body frame traits (top line, rump angle, rear leg set, hock quality,

TABLE 1 | Descriptive statistics of analyzed traits.

Traits	n	Mean ± SD	Min	Max	Descriptor[a]
Measured body frame traits					
Wither height (cm)	1,815	125.18 ± 5.28	108.0	142.0	—
Rump height (cm)	1,815	128.50 ± 5.37	113.0	146.0	—
Body length (cm)	1,815	125.66 ± 6.42	105.0	162.0	—
Chest girth (cm)	1,812	173.88 ± 9.64	150.0	208.0	—
Scored autochthonous traits					
Head length	1,815	5.45 ± 1.26	2	9	long - short
Head nobility	1,179	5.66 ± 1.15	2	9	heavy - fine
Eyes	1,815	5.71 ± 1.05	2	9	small – large
Horn base circumference	1,765	5.02 ± 1.35	1	9	thick – thin
Horns length	1,571	5.56 ± 1.33	2	9	long - short
Horns direction	1,765	5.37 ± 1.60	2	9	outward - forward
Neck	1,815	5.51 ± 1.28	2	9	heavy - fine
Dewlap	1,811	5.25 ± 1.39	2	9	heavy - fine
Coat color	1,815	5.14 ± 1.21	1	9	very dark - very bright
Back stripe	1,815	5.21 ± 1.42	1	9	wide - narrow
Rear legs stripe	1,811	5.40 ± 1.68	1	9	wide - narrow
Front legs stripe	1,812	6.25 ± 1.56	1	9	wide - narrow
Scored body frame traits					
Top line	1,815	4.76 ± 0.65	2	7	weak - straight
Rump angle	1,815	5.24 ± 0.84	2	8	high pins – extreme sloped
Rear leg set	1,815	5.55 ± 0.77	3	8	straight - sickled
Hock quality	1,815	5.74 ± 1.28	2	9	a lot of fluid – clean and dry
Foot angle	1,815	5.57 ± 1.05	2	8	low - steep
Heel height	1,815	5.49 ± 1.03	2	9	low - tall
Scored udder traits					
Fore udder attachment	1,815	4.95 ± 1.09	2	8	loose - strong
Udder depth	1,815	5.78 ± 1.10	2	8	deep - shallow
Teat thickness	1,815	4.91 ± 1.07	2	9	thin - thick
Front teat length	1,814	5.33 ± 1.16	2	9	short - long
Composite traits					
Autochthonous characteristics	1,815	5.44 ± 1.47	1	9	poor - excellent
Muscularity	1,815	5.36 ± 1.13	2	9	poor - excellent
Body frame	1,815	5.57 ± 1.14	2	8	fine - heavy
Udder	1,815	4.98 ± 1.14	1	8	poor - excellent

[a]*Minimum = 1, maximum = 9*

foot angle, heel height, body frame as a composite trait) were assigned to Group 5. Udder traits (fore udder attachment, udder depth, teat thickness, front teat length, udder as a composite trait) were assigned to Group 6. The variance components and genetic parameters for these anatomically similar traits divided into six groups were estimated separately for each group using multivariate mixed models. The two remaining composite traits (autochthonous characteristics, muscularity) were analyzed with a univariate mixed model.

RESULTS AND DISCUSSION

Variance Component Estimates

The heritabilities (h^2) for all included type traits in the first parity Cika cows ranged between 0.11 and 0.80, while the proportions of variance explained by the herd effect ranged between 0.01 and 0.28. Considering the six different trait groups, the heritabilities for the measured body frame traits ranged from 0.42 to 0.67, for the scored autochthonous traits from 0.36 to 0.80, for the scored body frame traits from 0.11 to 0.61, and for the scored udder traits from 0.20 to 0.47. The estimated heritabilities for the composite traits of autochthonous characteristics, muscularity, body frame, and udder were 0.55, 0.19, 0.19, and 0.26, respectively (**Table 2**).

The estimated h^2 for measured wither height and rump height in the first parity Cika cows were very similar, 0.66 ± 0.05 and 0.67 ± 0.05, respectively (**Table 2**). A lower h^2 for wither height was estimated in first parity Piemontese cows (0.31 ± 0.02; Mantovani et al., 2010). A similar h^2 for rump height was estimated in first parity Holstein cows (0.69 ± 0.03), American Brown Swiss cows (0.64 ± 0.02), and Red Holstein cows (0.74 ± 0.03) in Switzerland (de Haas et al., 2007), whereas the estimated h^2 was lower in first parity Brown Swiss cows in Slovenia (0.46; Špehar et al., 2012). A lower h^2 for rump height was estimated in first parity Rendena cows (0.52; Mazza et al., 2014), American Brown Swiss cows (0.32; Dal Zotto et al., 2007), and first parity

TABLE 2 | Estimated proportions of variance components ± standard errors for type traits in the first parity Cika cows.

Traits	h^2	c^2	e^2
Measured body frame traits			
Wither height	0.66 ± 0.05	0.10 ± 0.02	0.24 ± 0.04
Rump height	0.67 ± 0.05	0.11 ± 0.02	0.22 ± 0.04
Body length	0.54 ± 0.05	0.13 ± 0.02	0.33 ± 0.04
Chest girth	0.42 ± 0.05	0.25 ± 0.03	0.33 ± 0.04
Scored autochthonous traits			
Head length	0.48 ± 0.03	0.06 ± 0.01	0.46 ± 0.03
Head nobility	0.43 ± 0.04	0.08 ± 0.02	0.49 ± 0.04
Eyes	0.39 ± 0.03	0.06 ± 0.02	0.55 ± 0.03
Horn base circumference	0.52 ± 0.03	0.11 ± 0.02	0.37 ± 0.03
Horns length	0.53 ± 0.03	0.04 ± 0.01	0.43 ± 0.03
Horns direction	0.36 ± 0.03	0.05 ± 0.02	0.60 ± 0.03
Neck	0.47 ± 0.04	0.07 ± 0.02	0.47 ± 0.05
Dewlap	0.63 ± 0.04	0.05 ± 0.02	0.32 ± 0.04
Coat color	0.63 ± 0.04	0.05 ± 0.01	0.32 ± 0.04
Back stripe	0.80 ± 0.04	0.02 ± 0.01	0.18 ± 0.04
Rear legs stripe	0.78 ± 0.04	0.02 ± 0.009	0.20 ± 0.04
Front legs stripe	0.67 ± 0.04	0.01 ± 0.007	0.32 ± 0.04
Scored body frame traits			
Top line	0.19 ± 0.05	0.09 ± 0.02	0.72 ± 0.05
Rump angle	0.32 ± 0.03	0.01 ± 0.008	0.67 ± 0.04
Rear leg set	0.15 ± 0.03	0.11 ± 0.02	0.74 ± 0.03
Hock quality	0.61 ± 0.04	0.05 ± 0.01	0.34 ± 0.04
Foot angle	0.11 ± 0.02	0.16 ± 0.02	0.73 ± 0.03
Heel height	0.12 ± 0.02	0.23 ± 0.02	0.65 ± 0.03
Scored udder traits			
Fore udder attachment	0.20 ± 0.03	0.13 ± 0.02	0.66 ± 0.03
Udder depth	0.26 ± 0.03	0.13 ± 0.02	0.61 ± 0.04
Teat thickness	0.34 ± 0.03	0.10 ± 0.02	0.56 ± 0.03
Front teat length	0.47 ± 0.05	0.06 ± 0.01	0.47 ± 0.04
Composite traits			
Autochthonous characteristics	0.55 ± 0.05	0.09 ± 0.02	0.36 ± 0.04
Muscularity	0.19 ± 0.05	0.28 ± 0.03	0.53 ± 0.04
Body frame	0.19 ± 0.03	0.14 ± 0.02	0.67 ± 0.03
Udder	0.26 ± 0.03	0.10 ± 0.02	0.64 ± 0.03

h^2 = heritability, c^2 = herd effect, e^2 = residual.

TABLE 3 | Estimated genetic correlations ± standard errors (above diagonal) and phenotypic correlations (below diagonal) for measured body frame traits from Group 1.

Trait	Wither height	Rump height	Body length	Chest girth
Wither height	—	0.99 ± 0.002	0.98 ± 0.009	0.91 ± 0.02
Rump height	0.98	—	0.97 ± 0.01	0.90 ± 0.03
Body length	0.83	0.83	—	0.94 ± 0.02
Chest girth	0.70	0.70	0.75	—

Holstein-Friesian cows in the Czech Republic (0.45; Němcová et al., 2011). The estimated h^2 for body length in the first parity Cika cows was 0.54 ± 0.05 (**Table 2**). A lower h^2 for body length was estimated in Rendena cows (0.41; Mazza et al., 2014). The estimated h^2 for chest girth in first parity Cika cows was 0.42 ± 0.05, which was only slightly higher than in Holstein (0.38 ± 0.02), American Brown Swiss (0.35 ± 0.02) and Red Holstein cows (0.36 ± 0.02) in Switzerland (de Haas et al., 2007).

The estimated h^2 for head length in the first parity Cika cows was 0.48 ± 0.03 (**Table 2**). A lower h^2 for head length was estimated in the Spanish Asturiana de los Valles (0.25 ± 0.02; Gutiérrez and Goyache, 2002), and in Piemontese cows (0.15 ± 0.02; Mantovani et al., 2010). The estimated h^2 for the composite autochthonous trait in first parity Cika cows was 0.55 ± 0.05 (**Table 2**). In Asturiana de los Valles cows, the estimated h^2 for a similar composite trait called "breed characteristics" was 0.33 ± 0.02 (Gutiérrez and Goyache, 2002). The estimated h^2 for the coat color and coat pattern traits (coat color, back stripe, rear legs stripe, front legs stripe) in the first parity Cika cows were between 0.63 and 0.80.

The estimated h^2 for the top line in the first parity Cika cows was 0.19 ± 0.05 (**Table 2**). A lower h^2 for top line was estimated in first parity Asturiana de los Valles cows (0.11 ± 0.01; Gutiérrez and Goyache, 2002), Piemontese cows (0.07 ± 0.01; Mantovani et al., 2010), Brown Swiss cows in Slovenia (0.16; Špehar et al., 2012) and American Brown Swiss cows (0.10; Dal Zotto et al., 2007). Estimated h^2 for rump angle in first parity Cika cows was 0.32 ± 0.03 (**Table 2**). A similar h^2 for rump angle was estimated in first parity Rendena cows (0.36; Mazza et al., 2014) and Czech Holstein-Friesian cows (0.34; Němcová et al., 2011). The h^2 for rump angle in first parity Brown Swiss cows in Slovenia was 0.22 (Špehar et al., 2012), and 0.24 in American Brown Swiss cows (Dal Zotto et al., 2007). The estimated h^2 for rear leg set in the first parity Cika cows was 0.15 ± 0.03 (**Table 2**). Similar h^2 for rear legs set was found in the Rendena cows (0.21; Mazza et al., 2014), Piemontese (0.12 ± 0.02; Mantovani et al., 2010), Slovenian Brown Swiss (0.13; Špehar et al., 2012), American Brown Swiss cattle (0.14; Dal Zotto et al., 2007) and Czech Holstein-Friesian cows (0.16; Němcová et al., 2011). The estimated h^2 for hock quality in the first parity Cika cows was 0.61 ± 0.04 (**Table 2**). A lower h^2 for hock quality was estimated in Slovenian Brown Swiss (0.11; Špehar et al., 2012) and American Brown Swiss cows (0.08; Dal Zotto et al., 2007). The estimated h^2 for heel height in first parity Cika cows was 0.12 ± 0.02 (**Table 2**). An equal or very similar h^2 for heel height was estimated in Rendena cows (0.12; Mazza et al., 2014), Piemontese (0.09 ± 0.01; Mantovani et al., 2010), Slovenian Brown Swiss (0.04; Špehar et al., 2012), American Brown Swiss (0.09; Dal Zotto et al., 2007) and in Czech Holstein-Friesian cows (0.10; Němcová et al., 2011). Likewise, the estimated h^2 for the composite trait body frame in the first parity Cika cows was 0.19 ± 0.03, which was very similar to Rendena cows (0.18; Mazza et al., 2014), an autochthonous small frame cattle breed from north-eastern Italy.

The estimated h^2 for udder attachment in the first parity Cika cows was 0.20 ± 0.03 (**Table 2**), which was lower than in Rendena cows (0.32; Mazza et al., 2014), and similar to Czech Holstein-Friesian (0.24; Němcová et al., 2011) as well as French Holstein cows (0.18; Rupp and Boichard, 1999). The estimated h^2 for fore udder attachment was higher than in Slovenian Brown Swiss (0.14; Špehar et al., 2012) and American Brown Swiss cows (0.14; Dal Zotto et al., 2007). The estimated h^2 for udder depth in the first parity Cika cows was 0.26 ± 0.03, similar to Rendena cows (0.27; Mazza et al., 2014), Slovenian Brown Swiss (0.22; Špehar et al., 2012), American Brown Swiss (0.23; Dal Zotto et al., 2007),

TABLE 4 | Estimated genetic correlations ± standard errors (above diagonal) and phenotypic correlations (bellow diagonal) for scored autochthonous traits from Group 2.

Trait	Head length	Head nobility	Eyes	Horn base circumference	Horn length	Horns direction
Head length	—	0.59 ± 0.04	0.76 ± 0.04	0.38 ± 0.05	–0.05 ± 0.06	0.10 ± 0.07
Head nobility	0.54	—	0.74 ± 0.04	0.67 ± 0.04	0.31 ± 0.06	0.45 ± 0.07
Eyes	0.42	0.47		0.29 ± 0.04	0.18 ± 0.06	0.17 ± 0.07
Horn base circumference	0.20	0.41	0.23	—	0.60 ± 0.04	0.42 ± 0.05
Horn length	0.09	0.22	0.09	0.42	—	0.44 ± 0.05
Horns direction	0.14	0.32	0.23	0.32	0.37	—

Czech Holstein-Friesian (0.32; Němcová et al., 2011) and French Holstein cows (0.29; Rupp and Boichard, 1999). The estimated h^2 for teat thickness in first parity Cika cows (0.34 ± 0.03; **Table 2**) was similar to Rendena (0.34; Mazza et al., 2014), Slovenian Brown Swiss (0.33; Špehar et al., 2012), American Brown Swiss (0.32; Dal Zotto et al., 2007), Czech Holstein-Friesian (0.28; Němcová et al., 2011) and French Holstein cows (0.30; Rupp and Boichard, 1999). Taking into account estimated composite trait "udder" in the first parity Cika cows (h^2 = 0.26 ± 0.03; **Table 2**), the results were lower than in Rendena cows (0.37; Mazza et al., 2014), and higher than in Slovenian Brown Swiss cows (0.16; Špehar et al., 2012).

The estimated h^2 for muscularity in the first parity Cika cows was 0.19 ± 0.05 (**Table 3**), which was similar to Asturiana de los Valles cows (0.22 ± 0.01; Gutiérrez and Goyache, 2002) and Brown Swiss cows (0.16; Špehar et al., 2012). On the other hand, h^2 for muscularity in the first parity Cika cows was lower than in Rendena (0.31; Mazza et al., 2014), American Brown Swiss (0.42 ± 0.02) and Red Holstein cows (0.59 ± 0.03) (de Haas et al., 2007). The huge variability in the heritabilities for the trait muscularity could be due to the recording procedure, the degree of harmonization among the experts, the statistical models, and data quality.

The proportion of variance in the type traits of the first parity Cika cows explained by the effect of the herd (c^2) ranged from 0.01 to 0.28 (**Table 3**). Interestingly, c^2 was on average higher for measured body frame traits (0.10–0.25) compared to the scored body frame traits (0.01–0.23). Likewise, the scored autochthonous and udder traits had a low c2 (0.01–0.11 and 0.06 to 0.13, respectively). The estimated c^2 for the composite traits (autochthonous characteristics, muscularity, body frame, udder) were 0.09, 0.28, 0.14, and 0.10, respectively. This might reflect the effect of the huge variability of production systems on the farms that cause variability in the body condition of the animals in the herd. The proportion of residual variance ranged from 0.18 to 0.74. The standard errors of the variance components varied between 0.007 and 0.05.

Genetic and Phenotypic Correlations

The estimated genetic and phenotypic correlations between the measured body frame traits (**Table 3**) were positive and very high. Based on these high correlations, it would be recommended to reduce the number of measured traits in the scoring form, to make the whole procedure less time consuming. The ICAR guidelines for conformation recording (ICAR, 2018) recommend only rump height (called "stature") to be scored.

TABLE 5 | Estimated genetic correlation ±standard error (above diagonal) and phenotypic correlation (bellow diagonal) for scored autochthonous traits from Group 3.

Trait	Neck	Dewlap
Neck	—	0.77 ± 0.05
Dewlap	0.53	—

Likewise, Mazza et al. (2014) found high genetic (0.79) and phenotypic (above 0.53) correlations between body frame traits in first parity Rendena cows. On the other hand, lower positive genetic correlations between rump height and chest girth were estimated in Holstein (0.45), American Brown Swiss (0.34) and Red Holstein cows (0.54) (de Haas et al., 2007).

The estimated genetic correlations between the traits describing parts of the head (**Table 4**) were low to moderate, and mostly positive. The only slightly negative genetic correlation was found between head and horn length (−0.05), where SE was 0.06 and, therefore, the genetic correlation was not different from zero. The highest genetically correlated traits were head length and eyes (0.76). The phenotypic correlations were lower than the genetic ones, ranging from 0.09 (horn length - head length, horn length - eyes) to 0.54 (head nobility - head length).

The estimated phenotypic and genetic correlation between the neck and dewlap (**Table 5**) was positive and moderate (0.53). Animals with thin skin on the neck had a less expressed dewlap and vice versa, which was expected.

The estimated genetic and phenotypic correlations between the scored traits describing the coat color and white stripes patterns (**Table 6**) were positive, very low one the hand or very high on the other. The coat color was weakly correlated with all traits describing white stripe patterns, whereas the back stripe, the rear legs stripe, and the front legs stripe were highly correlated with each other. Likewise, the genetic correlation between the coat color and the front legs stripe (0.02) had a SE of 0.03 and, therefore, was not different from zero. First parity Cika cows with a wider white back stripe had wider white strips on the rear and front legs. Cows with a wider white stripe on the rear legs had a wider white stripe on the front legs as well.

The majority of the estimated correlations among the body frame traits (**Table 7**) were low to moderate. The lowest genetic correlation was found between rump angle and hock quality (−0.03), while the highest was between heel height and foot angle (0.89). However, the genetic correlation between rump angle and hock quality (−0.03) had a SE of 0.06 and was not different from

TABLE 6 | Estimated genetic correlations ± standard errors (above diagonal) and phenotypic correlations (bellow diagonal) for scored autochthonous traits from Group 4.

Trait	Coat color	Back stripe	Rear legs stripe	Front legs stripe
Coat color	—	0.10 ± 0.02	0.06 ± 0.02	0.02 ± 0.03
Back stripe	0.07	—	0.97 ± 0.008	0.91 ± 0.02
Rear legs stripe	0.06	0.82	—	0.95 ± 0.01
Front legs stripe	0.01	0.75	0.80	—

TABLE 7 | Estimated genetic correlations ± standard errors (above diagonal) and phenotypic correlations (bellow diagonal) for scored and composite body frame traits from Group 5.

Trait	Top line	Rump angle	Rear leg set	Hock quality	Foot angle	Heel height	Body frame – composite trait
Top line		0.49 ± 0.07	–0.30 ± 0.15	0.23 ± 0.08	0.10 ± 0.13	–0.06 ± 0.15	0.31 ± 0.10
Rump angle	0.25	—	0.12 ± 0.11	–0.03 ± 0.06	0.09 ± 0.09	–0.12 ± 0.10	–0.40 ± 0.07
Rear legs set	–0.06	–0.01		–0.10 ± 0.05	–0.42 ± 0.10	–0.24 ± 0.12	–0.59 ± 0.09
Hock quality	0.07	–0.05	0.05	—	–0.12 ± 0.10	–0.40 ± 0.10	0.27 ± 0.07
Foot angle	0.07	–0.03	–0.26	–0.04	—	0.89 ± 0.05	0.58 ± 0.10
Heel height	0.02	–0.04	–0.14	–0.12	0.62	—	0.58 ± 0.09
Body frame – composite trait	0.23	–0.13	–0.35	0.12	0.59	0.54	—

TABLE 8 | Estimated genetic correlations ± standard errors (above diagonal) and phenotypic correlations (bellow diagonal) for scored udder traits and the composite trait udder from Group 6.

Trait	Fore udder attachment	Udder depth	Teat thickness	Front teat length	Udder – composite trait
Fore udder attachment	—	0.78 ± 0.06	–0.52 ± 0.10	-0.45 ± 0.10	0.89 ± 0.04
Udder depth	0.42	—	–0.71 ± 0.07	–0.74 ± 0.05	0.94 ± 0.03
Teat thickness	–0.09	–0.36	—	0.78 ± 0.04	–0.62 ± 0.09
Front teat length	–0.19	–0.39	0.59	—	–0.64 ± 0.08
Udder – composite trait	0.79	0.56	–0.22	–0.35	—

zero. The highest phenotypic correlation (0.62) was also found between heel height and foot angle, whereas the lowest phenotypic correlation (–0.001) was between rear leg set and rump angle. Animals with a low foot angle had low heel height, while animals with a steep foot angle had high heel height. Moderate genetic correlations were estimated between the composite trait body frame and heel height as well as with foot angle (both 0.58). The composite trait body frame had the highest negative correlation with rear leg set (–0.59). Cows with high scores for the composite trait of body frame had relatively steep foot angles and high heel height, and a steep rear leg set. Špehar et al. (2012) estimated a similar genetic correlation between rear leg set and heel height (–0.33) in Slovenian Brown Swiss cows. Němcová et al. (2011) estimated lower genetic and the same phenotypic correlation between rump angle and heel height (–0.06; –0.04) in Czech Holstein-Friesian cows.

The estimated phenotypic correlations between udder traits (**Table 8**) were moderate and positive (up to 0.79 between fore udder attachment and the composite udder trait), and weak to moderate negative (up to –0.39 for front teat length and udder depth). Some estimated genetic correlations were high and positive, with the highest correlation between udder depth and the composite trait "udder" (0.94). The others were moderate and negative (e.g. –0.74 between udder depth and front teat length). The lowest genetic correlation was estimated between front teat length and fore udder attachment (–0.45). A functional udder should be extended under the abdomen, well attached, with thin and short teats.

Fore udder attachment and udder depth were relatively strongly correlated (0.78), as well as front teat length and teat thickness (0.78). Cows with a genetic predisposition for weakly attached udders usually have deeper udders as well. Němcová et al. (2011) estimated higher phenotypic and lower genetic correlations between fore udder attachment and udder depth (0.44; 0.75) in Czech Holstein-Friesian cows. Mazza et al. (2014) estimated lower phenotypic and genetic correlations between fore udder attachment and the composite trait "udder" (0.68; 0.78) in Rendena cows.

Genetic Trends

All 26 individual type traits as well as the four composite traits were already introduced in the genetic evaluation of first parity Cika cows in the year 2016. The composite traits, which include all individual traits, are widely used. There is a plan to decrease the number of traits in the future since some of the traits within each group are highly correlated. Nevertheless, **Figure 1** includes the genetic trends of all 26 individual type traits as well as the four

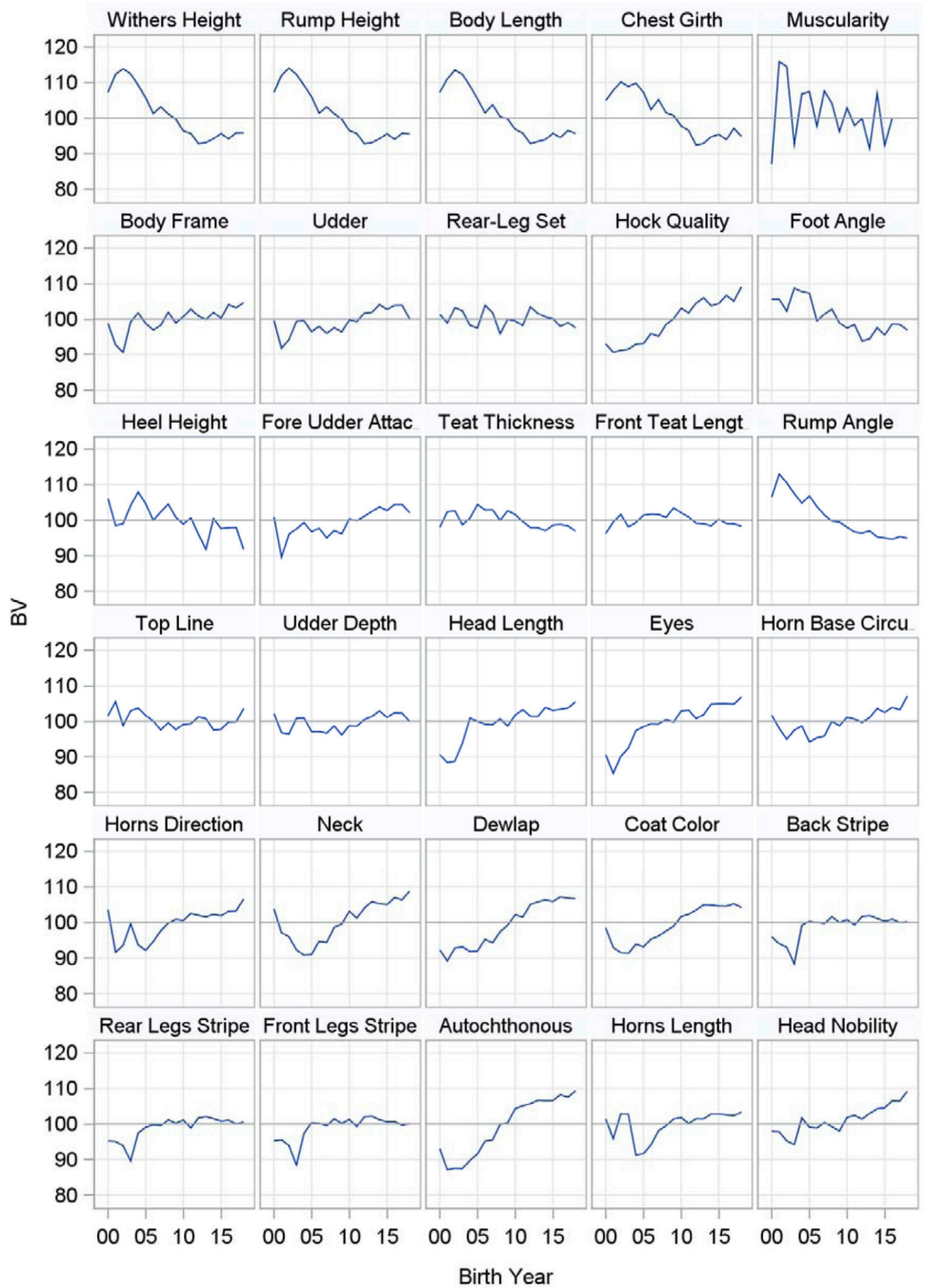

FIGURE 1 | Genetic trends for type traits in first parity Cika cows born in the years 2000–2019.

composite traits for first parity Cika cows born from 2000 to 2019. The genetic trends of all four measured body frame traits are negative, which is desirable and expected since the breeders prefer Cika animals with a smaller body frame that were not admixed with Pinzgauer in the past. Moreover, potential Cika sires are genotyped each year to select those from the group suitable for artificial insemination that do not have genes of Pinzgauer or other foreign breeds. The autochthonous characteristics of the cows are represented by 12 individual traits divided into traits of the head, neck and coat color as well as coat color patterns. The genetic trends for the traits of the head are all positive. First parity Cika cows born in the last years had shorter and finer heads with larger eyes and thinner, shorter horns with a more forward direction compared to Cika cows born before 2010, which is in accordance with the breeding goals of the breeding program. A very similar situation is reflected by the positive genetic trends of

the neck and dewlap, which became finer in younger cows. Likewise, heavier necks and dewlaps are typical for Cika cows that were admixed with Pinzgauer in the past. According to the genetic trends for coat color and its patterns, the coat color is becoming brighter, while white stripes on the back as well as on the rear and front legs have not changed very much in animals born during the last years. Among the scored body frame traits, only hock quality had a positive genetic trend, while rump angle, foot angle and heel height had negative genetic trends. Consequently, cows born in the last years have more sloping rumps, as well as cleaner and drier hocks with steeper angles of the foot and taller heels. In the past, a lot of cows had high pins and hocks with a lot of fluid, which was very undesirable for the breeders, i.e. the genetic trends are in accordance with the breeding goals now. Among the scored udder traits, only fore udder attachment and udder depth had slightly positive trends, which means that cows have more strongly attached and shallower udders than in the past. Finally, three (autochthonous characteristics, body frame, udder) of four composite traits had positive genetic trends. All in all, first parity Cika cows born in the last years showed more typical autochthonous type traits of purebred Cika cows, which distinguished them from Cika cows admixed with Pinzgauer in the past. Likewise, first parity Cika cows nowadays have more excellent body frames and udders.

CONCLUSION

Comparing the estimated heritabilities for the majority of included type traits with heritabilities from the literature, it was found that they were within the expected range and similar to those reported for other breeds. Among the measured body frame traits, the highest heritability was estimated for rump height. The heritabilites of scored body frame traits in Cika were similar to those of other breeds as well. In the scoring form for Cika, only four scored udder traits are included (fore udder attachment, udder depth, teat thickness, front teat length). Nevertheless, their heritabilities were very similar to those reported for cows of dairy breeds. The estimated heritabilities of the scored autochthonous traits, which describe the breed characteristics, were moderate to high. This suggests the possibility to breed first parity Cika cows with unique type traits known for autochthonous Cika cattle. Unfortunately, according to the best of our knowledge, no autochthonous type traits like those included in the analysis have been investigated before. Consequently, comparison with the literature was not possible.

All above estimated genetic parameters are already used for breeding value prediction in the Cika cattle population. According to the genetic trends, first parity Cika cows born in the last years showed more typical autochthonous type traits of purebred Cika cows, which distinguished them from Cika cows admixed with Pinzgauer in the past. Likewise, first parity Cika cows nowadays have more excellent body frames and udders.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MS wrote the manuscript and gave the main idea, BL wrote the manuscript and prepared a plan of analysis, MŠ prepared all analysis, BL prepared the data, KP checked the analysis and manuscript in detail.

REFERENCES

Comunod, S., Guida, S., Vigo, D., Beretti, V., Munari, E., Colombani, C., et al. (2013). Body Measures and Milk Production, Milk Fat Globules Granulometry and Milk Fatty Acid Content in Cabanina Cattle Breed. *Ital. J. Anim. Sci.* 12 (1), e18. doi:10.4081/ijas.2013.e18

Dal Zotto, R., De Marchi, M., Dalvit, C., Cassandro, M., Gallo, L., Carnier, P., et al. (2007). Heritabilities and Genetic Correlations of Body Condition Score and Calving Interval with Yield, Somatic Cell Score, and Linear Type Traits in Brown Swiss Cattle. *J. Dairy Sci.* 90, 5737–5743. doi:10.3168/jds.2007-0280

de Haas, Y., Janss, L. L. G., and Kadarmideen, H. N. (2007). Genetic and Phenotypic Parameters for Conformation and Yield Traits in Three Swiss Dairy Cattle Breeds. *J. Anim. Breed. Genet.* 124, 12–19. doi:10.1111/j.1439-0388.2007.00630.x

Felius, M., Koolmees, P. A., Theunissen, B., and Lenstra, J. A. (2011). On the Breeds of Cattle-Historic and Current Classifications. *Diversity* 3, 660–692. doi:10.3390/d3040660

Forabosco, F., Groen, A. F., Bozzi, R., Van Arendonk, J. A. M., Filippini, F., Boettcher, P., et al. (2004). Phenotypic Relationships between Longevity, Type Traits, and Production in Chianina Beef Cattle1. *J. Anim. Sci.* 82, 1572–1580. doi:10.2527/2004.8261572x

Frigo, E., Samorè, A. B., Vicario, D., Bagnato, A., and Pedron, O. (2013). Heritabilities and Genetic Correlations of Body Condition Score and Muscularity with Productive Traits and Their Trend Functions in Italian Simmental Cattle. *Ital. J. Anim. Sci.* 12, e40. doi:10.4081/ijas.2013.e40

Gengler, N., Seutin, C., Boonen, F., and Van Vleck, L. D. (1995). Estimation of Genetic Parameters for Growth, Feed Consumption, and Conformation Traits for Double-Muscled Belgian Blue Bulls Performance-Tested in Belgium. *J. Anim. Sci.* 73, 3269–3273. doi:10.2527/1995.73113269x

Groeneveld, E., Kovač, M., and Mielenz, N. (2010). VCE User's Guide and Reference Manual Version 6.0. Neustadt, Germany: Institute of Farm Animal Genetics. Available at: https://www.pdffiller.com/303612605-vce6-manual-31-A4pdf-VCE-bUser39s-Guideb-and-Reference-bManualb-Version-60-ftp-tzv-fal-.

Gutiérrez, J. P., and Goyache, F. (2002). Estimation of genetic parameters of type traits in Asturiana de los Valles beef cattle breed. *J. Anim. Breed. Genet.* 119, 93–100. doi:10.1046/j.1439-0388.2002.00324.x

ICAR (2018). *ICAR Guidelines for Conformation Recording of Dairy Cattle, Beef Cattle, Dual Purpose Cattle and Dairy Goats – Section 5*. New Delhi: ICAR – International Committee for Animal Recording, 76. Available at: https://www.icar.org/Guidelines/05-Conformation-Recording.pdf.

Larroque, H., and Ducrocq, V. (2001). Relationships between Type and Longevity in the Holstein Breed. *Genet. Selection Evol.* 33, 39–59. doi:10.1186/1297-9686-33-1-39

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and

Morphological Traits in the Alpine Grey Cattle Breed. *Animals* 11, 1340. doi:10.3390/ani11051340

Mantovani, R., Cassandro, M., Contiero, B., Albera, A., and Bittante, G. (2010). Genetic Evaluation of Type Traits in Hypertrophic Piemontese Cows. *J. Anim. Sci.* 88, 3504–3512. doi:10.2527/jas.2009-2667

Mazza, S., Guzzo, N., Sartori, C., Berry, D. P., and Mantovani, R. (2014). Genetic Parameters for Linear Type Traits in the Rendena Dual-Purpose Breed. *J. Anim. Breed. Genet.* 131 (1), 27–35. doi:10.1111/jbg.12049

Mazza, S., and Mantovani, R. (2012). Heritability of Linear Type Traits in the Autochthonous Rendena Dual Purpose Breed. *Acta Agric. Slov., Suppl.* 3, 161–165.

Mazza, S., Sartori, C., and Mantovani, R. (2015). Genetic Parameters of Type Traits in Two Strains of Dual Purpose Autochthonous Valdostana Cattle. *Livestock Sci.* 178, 35–42. doi:10.1016/j.livsci.2015.06.008

Němcová, E., Štrípková, M., and Zavadilová, L. (2011). Genetic Parameters for Linear Type Traits in Czech Holstein Cattle. *Czech J. Anim. Sci.* 56 (4), 157–162.

Neuenschwander, T., Kadarmideen, H. N., Wegmann, S., and de Haas, Y. (2005). Genetics of Parity-Dependant Production Increase and its Relationship with Health, Fertility, Longevity, and Conformation in Swiss Holsteins. *J. Dairy Sci.* 88, 1540–1551. doi:10.3168/jds.s0022-0302(05)72823-x

Norman, H. D., Powell, R. L., Wright, J. R., and Cassell, B. G. (1988). Phenotypic and Genetic Relationship between Linear Functional Type Traits and Milk Yield for Five Breeds. *J. Dairy Sci.* 71, 1880–1896. doi:10.3168/jds.s0022-0302(88)79758-1

Rogers, G. W., Hargrove, G. L., Cooper, J. B., and Barton, E. P. (1991). Relationships Among Survival and Linear Type Traits in Jerseys. *J. Dairy Sci.* 74, 286–291. doi:10.3168/jds.s0022-0302(91)78172-1

Rupp, R., and Boichard, D. (1999). Genetic Parameters for Clinical Mastitis, Somatic Cell Score, Production, Udder Type Traits, and Milking Ease in First Lactation Holsteins. *J. Dairy Sci.* 82, 2198–2204. doi:10.3168/jds.s0022-0302(99)75465-2

SAS Institute Inc (2001). *The SAS System for Windows, 8.02*. Cary, NC: SAS Institute Inc.

Sambraus, H. H. (1999). *Gefährdete Nutztierrassen Ihre Zuchtgeschichte, Nutzung und Bewahrung [Endangered breeds, their breeding history, use and preservation]*. Ulmer, Stuttgart.

Simčič, M., Lenstra, J. A., Baumung, R., Dovč, P., Čepon, M., and Kompan, D. (2013). On the Origin of the Slovenian Cika Cattle. *J. Anim. Breed. Genet.* 130, 487–495. doi:10.1111/jbg.12034

Simčič, M., Smetko, A., Sölkner, J., Seichter, D., Gorjanc, G., Kompan, D., et al. (2015a). Recovery of Native Genetic Background in Admixed Populations Using Haplotypes, Phenotypes, and Pedigree Information-Uusing Cika Cattle as a Case Breed. *Plos One* 10 (4), e0123253. doi:10.1371/journal.pone.0123253

Simčič, M., Sölkner, J., Kompan, D., and Medugorac, I. (2015b). "Exclusion of Admixed Animals from Cika Cattle *In Situ* Conservation Program Using SNP Haplotypes," in *Book of Abstracts of the 66th Annual Meeting of the European Federation of Animal Science* (Wageningen,Netherlands: Wageningen Academic Publishers), 101.

Simčič, M., Štepec, M., Logar, B., Potočnik, K., and Terčič, D. (2016). Analysis of Type Traits of Cika First Calving Cows. *Acta Agric. Slov.* 108 (2), 93–102. doi:10.14720/aas.2016.108.2.4

Špehar, M., Štepec, M., and Potočnik, K. (2012). Variance Components Estimation for Type Traits in Slovenian Brown Swiss Cattle. *Acta Agric. Slov.* 100 (2), 107–115.

Thomas, C. L., Vinson, W. E., Pearson, R. E., and Norman, H. D. (1985). Components of Genetic Variance and Covariance for Linear Type Traits in Jersey Cattle. *J. Dairy Sci.* 68, 2989–2994. doi:10.3168/jds.s0022-0302(85)81194-2

Van Dorp, T. E., Dekkers, J. C. M., Martin, S. W., and Noordhuizen, J. P. T. M. (1998). Genetic Parameters of Health Disorders, and Relationships with 305-day Milk Yield and Conformation Traits of Registered Holstein Cows. *J. Dairy Sci.* 81, 2264–2270. doi:10.3168/jds.s0022-0302(98)75806-0

Žan Lotrič, M., Čepon, M., Kompan, D., Klinkon, Z., Rokavec, R., Jeretina, J., et al. (2010). *Rejski Program Za Cikasto Govedo*. Medvode: Združenje Rejcev Avtohtonega Cikastega Goveda V Sloveniji, 83.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fgene.2021.723599

Genomic Uniqueness of Local Sheep Breeds From Morocco

Abdessamad Ouhrouch[1,2], Simon Boitard[3], Frédéric Boyer[4], Bertrand Servin[5], Anne Da Silva[6], François Pompanon[4], Abdelmajid Haddioui[2] and Badr Benjelloun[1]**

[1]Livestock Genomics Laboratory, Regional Center of Agricultural Research Tadla, National Institute of Agricultural Research INRA, Rabat, Morocco, [2]Biotechnologies and Valorization of Plant-Genetic Resources Laboratory, Sultan Moulay Slimane University, Beni Mellal, Morocco, [3]CBGP, Université de Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France, [4]Université Grenoble Alpes, Université Savoie MT-Blanc, CNRS, LECA, Grenoble, France, [5]GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, Castanet-Tolosan, France, [6]PEREINE/E2LIM, Faculty of Science and Technics, Limoges, France

Sheep farming is a major source of meat in Morocco and plays a key role in the country's agriculture. This study aims at characterizing the whole-genome diversity and demographic history of the main Moroccan sheep breeds, as well as to identify selection signatures within and between breeds. Whole genome data from 87 individuals representing the five predominant local breeds were used to estimate their level of neutral genetic diversity and to infer the variation of their effective population size over time. In addition, we used two methods to detect selection signatures: either for detecting selective sweeps within each breed separately or by detecting differentially selected regions by contrasting different breeds. We identified hundreds of genomic regions putatively under selection, which related to several biological terms involved in local adaptation or the expression of zootechnical performances such as Growth, UV protection, Cell maturation or Feeding behavior. The results of this study revealed selection signatures in genes that have an important role in traits of interest and increased our understanding of how genetic diversity is distributed in these local breeds. Thus, Moroccan local sheep breeds exhibit both a high genetic diversity and a large set of adaptive variations, and therefore, represent a valuable genetic resource for the conservation of sheep in the context of climate change.

Keywords: sheep, whole genome sequences, local breeds, demography, selection signatures

OPEN ACCESS

Edited by:
Olivier Hubert Hanotte, University of Nottingham, United Kingdom

Reviewed by:
Lubos Vostry, Czech University of Life Sciences Prague, Czechia
Gaspar Manuel Parra-Bracamonte, Instituto Politécnico Nacional (IPN), Mexico

****Correspondence:***
Abdessamad Ouhrouch
ouhrouch1991@gmail.com
Badr Benjelloun
badr.benjelloun@inra.ma

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *11 June 2021*
Accepted: *09 November 2021*
Published: *02 December 2021*

Citation:
Ouhrouch A, Boitard S, Boyer F, Servin B, Da Silva A, Pompanon F, Haddioui A and Benjelloun B (2021) Genomic Uniqueness of Local Sheep Breeds From Morocco. Front. Genet. 12:723599. doi: 10.3389/fgene.2021.723599

1 INTRODUCTION

Sheep were among the first domesticated animals about 11,000 years B.P (Vigne, 2011; Demirci et al., 2013). They are one of the main sources of meat and milk around the world (Ritchie and Roser, 2017). According to the FAO, the worldwide stocks of sheep reached 1,238 M heads in 2019 with proportions of 42.6% in Asia, 32.9% in Africa, 10.3% in Europe, 7.5% in Oceania and 6.7% in the Americas (FAOSTAT, 2020). The world production increase of sheep products was about 13.7% for meat and about 9.9% for milk between 2009 and 2019 (FAOSTAT, 2020). In the context of environmental changes, the improvement and conservation of this species are major challenges to sustainably meet the growing needs of human populations for meat and dairy products, both at the national and international levels (Bruford et al., 2015). Furthermore, in order to achieve the United Nations SDG#2 (Sustainable Development Goals), the target #2.5 is directly concerned by maintaining genetic diversity in domestic species, and two associated indicators are related to the proportion of endangered breeds (#2.5.1 & #2.5.2). Thus, local traditionally-managed sheep

breeds represent an officially recognized valuable genetic resource for the conservation of this species on a global scale.

Technological advances during the last decade have made it possible to produce and process Whole Genome data for hundreds of individuals (Kulski, 2016). Similarly, bioinformatic improvement has incredibly advanced (Mangul et al., 2019). The DNA sequence of an individual is the most comprehensive collection of its genetic variation, and today's sequencing technology that is much more increasingly efficient, faster and cheaper than ever (Kulski, 2016) allows access to this genome-wide variation for many individuals from the same population or breed. This allows a precise characterization of genetic resources, by characterizing properly their demographic dynamics and geographic structuration, as well as their adaptive diversity (Benjelloun et al., 2019).

In Morocco, sheep are marked by a high global genetic diversity indicating a high adaptive potential (Benjelloun, 2015; Benjelloun et al., 2021). The main local breeds currently farmed have significant ability to adapt to their breeding system and environment. However, the genetic diversity within each breed and the demographic and adaptive history that shaped this diversity are not yet very well known. Only a few hypotheses have been emitted to describe their origin (Boujenane, 1999).

Sheep farming in Morocco plays an important economic and sociological role. It is practiced all over the country where it is often one of the main sources of farmers' income. Thus, sheep are bred under various environmental conditions and anthropogenic pressures. One of the main threats to these breeds is the unsupervised crossing practiced by farmers under increasing economic pressure as recently demonstrated by Belabdi et al. (2019). These practices, evaluated by the FAO to establish risk status of breeds, have led to genetic dilution (FAO, 2007), to which is added the risk of replacement of local breeds by cosmopolitan breeds (Taberlet et al., 2008).

The Moroccan sheep population is made up of about 1% of exogenous breeds and 99% of local breeds (Boujenane, 2006), among which the most important ones are Sardi, Dman, Timahdite, Beni Guil and Boujaad. Since 1980, five main local sheep breeds have been standardized and officially recognized by a large management program named the National Sheep Plan (MARA, 1980) which was based on assigning theses breeds to their exclusive specific habitat or area named "cradle of the breed". Thus, Beni Guil is bred in the Eastern plateaus in large herds using pastoralism as main feeding sources. The Dman breed has been bred in the palm groves of the pre-Saharan regions of the South-Eastern Morocco for a long time and is mainly located in the oases. Dman is considered as the most isolated and phenotypically distinguished breed in Morocco. The Sardi breed belongs to the sheep population of the western plateaus in Morocco. It is considered as a large-body sheep and very appreciated in religious events. Timahdite is a rustic breed well adapted to mountainous areas with low-input systems. Finally, the Ouled Jellal breed, which is not considered indigenous in Morocco, was shown to be among the main breeds farmed in the Eastern region of the country together with Beni Guil since a long time (Belabdi et al., 2019), and is considered as a true breed of the steppe, well adapted to nomadism.

Many characteristics that condition the fitness of livestock are associated with the production performances. Thus, the identification of signatures of selection is a valuable approach to identify the genes and polymorphisms underlying the phenotypic variation of these traits that are subjected to both anthropogenic selection (Rochus et al., 2018) and natural selection related to e.g., climatic or ecological constraints. While previous studies using WGS have identified many genome-environment associations in Moroccan sheep (Benjelloun, 2015), the genomic bases of traits specific to local breeds are still undetermined.

In this context, this study aims at characterizing the main five Moroccan breeds by describing their genomic diversity, inferring their demographic history and identifying shared and breed-specific selective sweeps using whole genome data. The data obtained enabled us to distinguish genomic regions putatively involved in local adaptation from those more directly related to anthropogenic pressures and associated to breed-specific zootechnical performances.

2 METHODS

2.1 Samples and Breeds

We sampled 87 unrelated sheep representatives of the geographic distribution of five local Moroccan breeds (Sardi, Dman, Timahdite, Beni Guil and Ouled Jellal; **Supplementary Table S1**) across the Northern half of Morocco (North of latitude 28°) between January 2008 and March 2012 in accordance with ethical regulations of the European Union Directive 86/609/EEC, as described in Benjelloun et al. (2021) and Benjelloun (2015). Tissue samples were taken from the distal part of the ear and then placed in alcohol for 1 day, after which they were transferred to a silica gel tube pending the extraction of DNA.

Additionally, a worldwide breed panel consisting of 20 sheep representing 20 different worldwide breeds was provided by the International Sheep Genome Consortium. The panel represents sheep from Asia, Africa Australia, America and Europe (**Supplementary Table S1**). Similarly, 13 wild Asiatic mouflons (*O. orientalis*) were collected either from captive or recently hunted animals, and from frozen samples available at the Iranian Department of Environment. These worldwide sheep (n = 20) and Asiatic mouflons (n = 13) were previously used in Alberto et al. (2018) and their whole genome sequence data were included here for comparisons and for increasing power when identifying selection signatures in Moroccan sheep (see **section 2.3** and **section 2.4** in the Methods).

2.2 Data Processing

As described by Benjelloun et al. (2021), Alberto et al., 2018 and Benjelloun et al. (2019), Illumina reads were aligned to the sheep reference genome (OAR v3.1, GenBank assembly GCA_000317765.1 (Jiang et al., 2014), and variant discovery was performed using three different algorithms: Samtools Mpileup (Li et al., 2009), GATK UnifiedGenotyper (McKenna

et al., 2010) and Freebayes (Garrison and Marth, 2012). Variants were called using a larger dataset than that used in this study (i.e., 160 sheep). Then two successive rounds of filtering variant sites were run. Filtering stage 1 merged together calls from the three algorithms, whilst filtering out the lowest-confidence calls. A variant site passed if it was called by at least two different calling algorithms with variant phred-scaled quality>30. An alternate allele at a site passed if it was called by any one of the calling algorithms, and the genotype count >0. Filtering stage 2 used Variant Quality Score Recalibration by GATK. First, a training set was generated of the highest-confidence variant sites where 1) the site was called by all three variant callers with variant phredscaled quality >100; 2) the site was biallelic, and 3) the minor allele count was at least three, counting only samples with genotype phredscaled quality >30. The training set was used to build a Gaussian model using the tool GATK VariantRecalibrator using the following variant annotations from UnifiedGenotyper: QD, HaplotypeScore, MQRankSum, ReadPosRankSum, FS, DP, Inbreeding Coefficient.

The Gaussian model was applied to the full data set, generating a VQSLOD (log odds ratio of being a true variant). Sites were filtered out if VQSLOD < cutoff value. The cutoff value was set for each population by the following: Minimum VQSLOD = {the median value of VQSLOD for training set variants}—3 * {the median absolute deviation VQSLOD of training set variants}. Measures of the transition/transversion ratio of SNPs suggest that this chosen cutoff criterion gives the best balance between selectivity and sensitivity.

Genotypes were improved and phased by Beagle 4 (Browning and Browning, 2013), and then filtered out where the genotype probability calculated by Beagle was <0.95. The resulting dataset was constituted of 47,622,950 variants.

2.3 Genetic Diversity and Demography

We used vcftools (Danecek et al., 2011) to estimate the heterozygosity (Ho), inbreeding coefficient (F) using polymorphic diploid bi-allelic SNPs and nucleotide diversity (π) for all diploid SNPs without missingness. We also calculated the F_{ST} index (Weir and Cockerham, 1984) using the same program (VCFtools) for each variant and their average over the whole genome between each two of the five Moroccan breeds, and between the 87 Moroccan individuals and the 22 sheep representing 12 cosmopolitan breeds.

In order to determine the demographic history of the studied breeds, we used the approximate Bayesian computation (ABC) approach implemented in PopSizeABC (Boitard et al., 2016a) to estimate the effective population size (Ne) through time from 130 K years to the present. PopSizeABC determines how Ne changes through time, by estimating empirical summary statistics from our VCF files and matching them to the simulated summary statistics obtained from simulated genomic data. The simulations explore a large set of Ne possible values for each of 21 pre-defined time windows, together with several values of the per generation per bp recombination rate (r), while assuming a fixed and pre-defined per generation per site mutation rate (μ). In order to compare the demographic history Moroccan sheep with the main worldwide breeds/populations, we included, in the analysis, genomic data of the worldwide sheep and Asiatic mouflons. The latters represent the closest wild descents of the species from which the current domestic sheep (*O. aries*) diverged since their domestication. The worldwide breeds were subjected to the improvement of their production performances by intensive selection (Alberto et al., 2018).

We estimated the recombination rate by using a uniform prior interval from $8e^{-9}$ to $13e^{-9}$ with an approximation of $\mathbf{r} = 10.65e^{-9}$, and then empirical data for each population. The minor allele count threshold (mac) for the Allele Frequency Spectrum (AFS) and Identity By State (IBS) statistics computation was set for each population to about **5% × N** (where N is the number of samples) and the minor allele count threshold for LD statistics computation (mac_ld) was set to about **20% × N** as recommended by Boitard et al., (2016a). The number of simulated datasets was nb_rep = 10,000, the number of independent segments in each dataset was nb_seg = 50 and the size of each segment in bp was L = 2,000,000, which overall gives a total of 1 Terabp of simulated data.

2.4 Selection Signatures

To identify selective sweeps, we have defined an integrated framework based on two complementary approaches as proposed in Boitard et al., (2016b). The first one (freqHMM) detects selection occurring within a single population while the second (FLK & hapFLK) is for detecting selection events differentiating populations.

A- A genome scan using the freqHMM program (Boitard et al., 2012) allowed identifying independently the putative regions under selection within each of the five studied breeds. This method is based on contrasting the local and genome-wide distributions of allele frequencies using data from a single population by assuming each SNP to have a hidden state which can take three different values: 1) "3 = Selection"; attributed to SNPs that are located in a selective sweep, 2) "intermediate", for SNPs not selected but located close to a selective sweep and 3) "1 = neutral", for SNPs that are far from any selective sweep. It aims at identifying ancient selection signatures that arose from new variants (*hard sweeps*) as described by Boitard et al. (2012). Ancestral alleles at each marker were defined using the homozygotes of the reference goat genome (CHIR_1.0; Dong et al., 2013) for the corresponding loci, by assuming that these are the ancestral alleles that mutated in *Ovis* to produce SNPs. Those ancestral alleles were added to the input VCF files using an in-house shell script. Afterwards, five different datasets representing whole genome variants with a total of polymorphic 31,442,046 SNPs (for which ancestral alleles were identified) of each of the five breeds were extracted using vcftools (Danecek et al., 2011) and data related to each chromosome was put in a separate file for each breed. Then, we calculated the number of derived and ancestral alleles for each SNP using vcftools. Parameter K of freqHMM was calibrated for each breed by simulating data using ms (Hudson, 2002) under neutrality, accounting for the demographic history previously estimated for the breed by PopSizeABC, and running freqHMM on these neutral data, as previously described in Boitard et al., (2016b). The value of K obtained by this approach was then used

TABLE 1 | Number of the specific variants and neutral genetic diversity parameters, for five Moroccan breeds.

Breed parameter	Beni Guil	Dman	Ouled Jellal	Sardi	Timahdite
# Exclusive variants	341,296	1,783,651	460,461	1,621,728	934,212
Ho	0.345796	0.222807	0.313409	0.242957	0.273747
F	0.000713	0.095303	0.02129	0.026200	0.007275
Pi	0.346784	0.247173	0.321013	0.250383	0.274969

to run the freqHMM analysis on the empirical data for each one of the five breeds.

B- The second approach to detect selective sweeps was a combination of the FLK (Bonhomme et al., 2010) and hapFLK (Fariello et al., 2013) methods implemented in the hapFLK (1.4 version) program (https://forge-dga.jouy.inra.fr/projects/hapflk/files). It was used to detect selection signatures by differentiating haplotypes among hierarchically structured populations, as described by Fariello et al. (2013). This has been done using all the genomic autosomes with no missingness allowed. FLK is a test for inbreeding coefficient heterogeneity that uses phylogenetically estimated relationships between populations. The allele frequencies are rescaled using a population kinship matrix which is estimated from the genomic data, measuring the amount of genetic drift expected under neutral evolution. HapFLK uses the differences in the frequencies of allele haplotypes between populations and the hierarchical structure of subpopulations.

The shared variants between the five Moroccan breeds and the 19 *Ovis Orientalis* (included as outgroup) were combined using the option "-T CombineVariants" of GATK (McKenna et al., 2010). The resulting dataset consisting of 31,721,507 polymorphic SNPs was used for FLK analysis to calculate Reynolds distances and the resulting Kinship Matrix. Then, this dataset was subdivided into several files each of which has a window of 10 Mbp with an overlap of 1 Mbp between each two successive windows. Subsequently, we launched FLK/hapFLK on all files one by one with a specified number of clusters set to K = 25. The analysis was launched with the option of keeping the Outgroup when computing FLK/hapFLK scores. We applied the approach of Storey and Tibshirani (2003), implemented in the q-value R package (Storey et al., 2015) to control the false discovery rate (FDR) based on FLK/hapFLK *p*-values.

Lists of genes that include or less than 5 kb away from the identified candidate SNPs (Downstream 5′-end and upstream 3′-end) were established and used for the Gene Ontology (GO) enrichment analyses. Similarly, outlier genomic regions were constituted of 50 kb segments surrounding outlier SNPs. GO enrichment analyses were performed using GOwinda (Kofler and Schlötterer, 2012) in order to explore the biological processes in which the identified genes under selection are involved. Bos Taurus was used as the reference species for that analysis. GOwinda effectively corrects for the gene length bias while identifying clearly over-represented GO categories, considering only SNPs being located in an exon are associated with the corresponding gene (using option: -gene-definition exon). A 5% FDR threshold was applied on GOwinda outputs. The identified GO terms in each population were clustered into homogenous groups using REVIGO (Supek et al., 2011).

3 RESULTS

3.1 Whole Genome Diversity

The individual observed heterozygosity (*Ho*) in the five breeds was 0.28 on average, varying from 0.22 in Dman to 0.35 in Beni Guil (**Table 1**). The inbreeding coefficient varied thus from 0.0007 in Beni Guil to 0.095 in Dman which was the most inbred of the five Moroccan breeds.

The whole genome nucleotide diversity (π), calculated over the 87 Moroccan sheep was 0.29. The Dman and Sardi breeds showed close π values of 0.247 and 0.25, respectively. Timahdite, Ouled Jellal and Beni Guil showed slightly higher π values of 0.274, 0.32 and 0.346, respectively (**Table 1**).

The number of breed-specific SNPs strongly varied among breeds, from 341 k and 460 k in Beni Guil and Ouled Jellal, respectively, to 1.6 and 1.78 M in Sardi and Dman, respectively (**Table 1**).

Regarding genetic distances, the pairwise F_{ST} was low, from 0 between Beni Guil and both Dman and Ouled Jellal to 0.004 between Sardi and both Dman and Ouled Jellal (**Table 2**). Furthermore, the pairwise F_{ST} between these five breeds and the cosmopolitan sheep was 0.027.

According to FLK tree (Figure S1), Timahdite is closely grouped with Sardi and Dman, Beni Guil is the most divergent, with rather long branches, among the Moroccan breeds. As expected, wilds were distant from all domestic breeds.

3.2 Demography of Moroccan Sheep

The PopsizeABC analysis highlighted that before 10,000 years ago, the estimated effective population size (Ne) was similar in all populations (**Figure 1**). Then, the variations of Ne differed in domestic and wild animals. For domestic populations, there was a strong bottleneck probably matching the time of domestication, which was not observed for the wilds. For Moroccan sheep, the low Ne persisted until 1 K years before present, and then increased. All Moroccan breeds had a very similar trajectory, with the exception of Timahdite which showed a slightly higher *Ne* in the recent past. Conversely, in cosmopolitan sheep, the bottleneck was first less pronounced but during the last 1 K years, the Ne regularly decreased until about 300 at present. However, we do not exclude that using a pool of individuals representing several breeds may impact the accuracy of the estimates.

TABLE 2 | Pairwise F_{ST} values between breeds.

Population	Beni Guil	Sardi	Ouled Jellal	Dman	Timahdite	Cosmopolitans	Wilds
Beni Guil	-					-	-
Sardi	0.002	-				-	-
Ouled Jellal	0.000	0.004	-			-	-
Dman	0.000	0.004	0.001	-		-	-
Timahdite	0.001	0.002	0.003	0.002	-	-	-
All five Moroccan breeds (87 individuals)	-	-	-	-	-	0.027	0.101

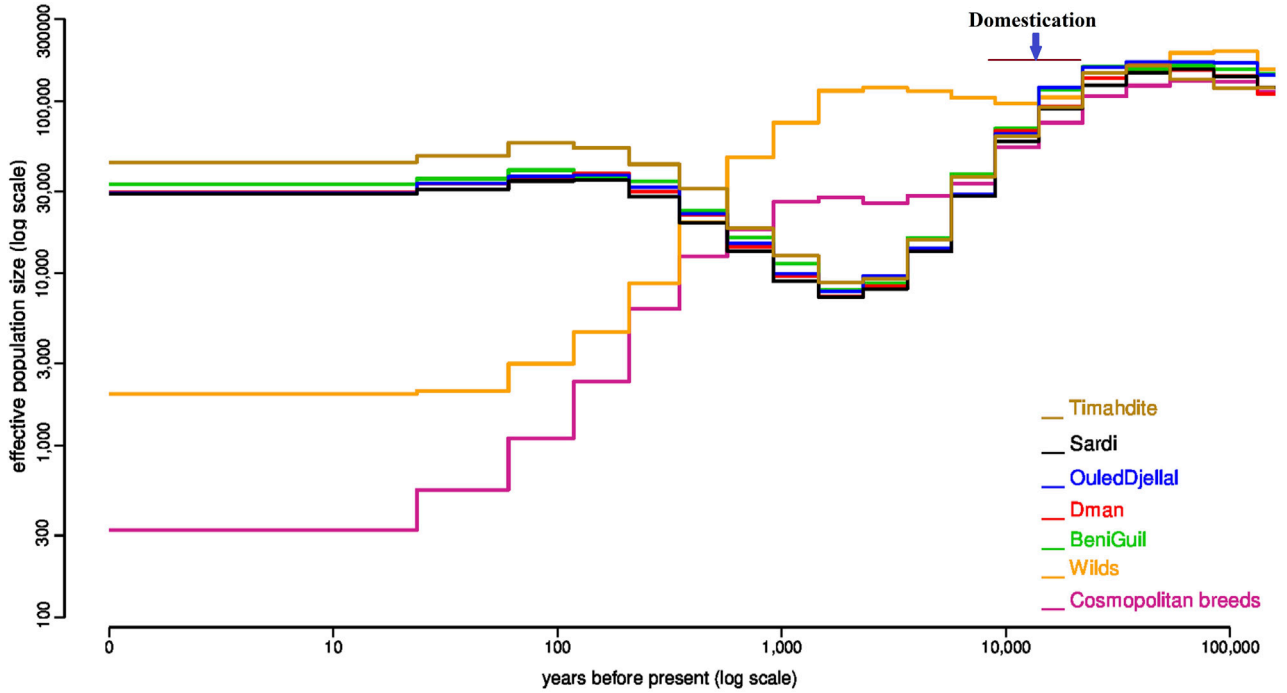

FIGURE 1 | Variation of the effective population size in five Moroccan breeds in comparison with the wild Asiatic mouflon and a multi-breed group of cosmopolitan sheep. Estimates were obtained independently for each group by the popsizeABC algorithm.

TABLE 3 | Number of candidate regions and SNPs under selection, with the corresponding genes, within each Moroccan sheep breed (intra-breed selection, using freqHMM).

Breed	Beni Guil	Dman	Ouled Jellal	Sardi	Timahdite
Number of regions	95	203	109	96	59
Number of SNPs	54,845	56,218	53,346	43,616	34,909
Number of genes	183	202	187	146	131

3.3 Selection Signatures

The analysis of the whole genome variants of the 87 Moroccan sheep belonging to five breeds, detected selection signatures both within (FreqHMM, Boitard et al., 2012) and between (hapFLK, Fariello et al., 2013) breeds.

3.3.1 Intra-Breed Selection Signatures

We identified 182,337 SNPs in 364 genomic regions (**Supplementary Table S3**) under selection. Comparing the studied breeds, we found that Dman displayed the higher number of regions under selection (203) followed by Ouled Jellal (109), Sardi (96), Beni Guil (95), and Timahdite in which we identified 59 genomic regions only (**Table 3**).

Regarding the shared selection signatures, we noticed that Dman share more SNPs/genes under selection with Sardi and Timahdite (respectively 14 K SNPs/70 genes and 9 K SNPs/60 genes; **Table 4**). A total of 219 SNPs were identified under selection in all five breeds, among which 203 were intergenic

TABLE 4 | Number of selected SNPs (above the diagonal) and genes (below the diagonal) in common between two Moroccan sheep breeds (intra-breed selection, using freqHMM).

Breed	Beni Guil	Dman	Ouled Jellal	Sardi	Timahdite
Beni Guil		8,180	5,307	6,311	5,880
Dman	45		5,795	14,568	9,436
Ouled Jellal	36	44		8,238	6,846
Sardi	40	82	46		9,708
Timahdite	34	78	47	58	

and 16 associated with three genes: *HMGA2* (3 SNPs), *RCOR1* (9 SNPs) and *SBF2* (4 SNPs).

3.3.2 Inter-Breeds Selection Signatures

An FDR and local FDR framework was applied to the hapFLK results in order to determine a reliable selection threshold, which was set to 0.1% FDR (q-values <0.001) in order to be sufficiently conservative. The whole process identified 8,887 SNPs in 27 different regions, of which 4,131 were associated with 20 genes and 4,756 were intergenic. Among the all outliers we identified 3,908 SNPs as top candidates with *p*-values lower than 10^{-12} (**Figure 2**). They were distributed across five regions associated with 9 genes (of which 7 genes have 31 missense variants). Among these regions, the greatest selective sweep was located on **Chr10: 29,363,691-29,806,294**. It was related to a high differentiation between Dman and the other breeds as shown in the cluster-plot (**Figure 2B**) where Dman had a different haplotype from that of the other breeds, and in the local tree (**Figure 2C**) where Dman displayed a stronger *p*-value in both SNP local tree (FLK scores) and haplotype local tree (hapFLK scores). This selective sweep was associated with the genes *RXFP2* and *ENSOARG00000011616*. The other strongest selective sweeps were related to 1) the region located on **Chr14: 13,329,709-14,250,423** differentiating Sardi and associated to 8 genes; 2) the region located on **Chr19: 2,143,797-2,261,064** differentiating Sardi and Beni Guil from the other breeds and associated with intergenic SNPs; 3) the region located on **Chr16: 34,551,942-34,646,856** differentiating Timahdite and associated with intergenic SNPs; 4) the region located on Chr2: 84,619,111-84,759,276 differentiating Timahdite and associated to the *BNC2* gene and some intergenic SNPs (**Figure 2A**).

3.3.3 Overall Selection Signatures

When combining both approaches, from a total of 31,721,507 analyzed SNPs, we identified as putatively under selection 46,799 variants associated with 155 genes in Sardi, 37,511 SNPs associated with 138 genes in Timahdite, 59,351 variants associated with 206 genes in Dman, 56,995 associated with 186 genes in the Beni Guil breed and 53,990 variants associated with 189 genes in Ouled Jellal. The Venn diagrams in **Figure 3** illustrate the number of regions, SNPs and associated genes that are specific or common to the studied breeds. We identified 219 SNPs within 7 genomic regions under selection in all 5 breeds, which are associated with 3 genes: *HMGA2*, *RCOR1* and *SBF2*.

3.3.3.1 Biological Processes Targeted by Selection

The selection signatures detected by both methods (i.e., FreqHMM & HapFLK) allowed identifying a total of 7,735 GO categories enriched in genes under selection (**Table 5**), based on a threshold of FDR<5% (**Supplementary Tables S4–S7**). The lowest number of enriched GO categories per breed was for Timahdite, where 1,465 categories clustered in 223 homogenous groups of biological processes. The Sardi breed exhibited the highest number of enriched GO term, with 1,611 categories clustering in 229 groups of processes. In regard to these high numbers of enriched biological processes, we'll limit our discussion to those which roles are the most straightforward given the related phenotypic traits and husbandry practices.

4 DISCUSSION

4.1 Whole Genomic Diversity

The heterozygosity measured in Moroccan breeds was generally moderate (mean of 0.28) and globally lower than the values reported for Iranian sheep breeds (Eydivandi et al., 2021) and Welsh breeds (Beynon et al. (2015). Dman and Sardi breeds have even much lower heterozygosity values. Our results showed low inbreeding in Moroccan sheep, comparably to those reported by

TABLE 5 | Examples of biological processes enriched in candidate genes in the five sheep breeds.

GO term	Biological process	Candidate genes associated	*p*-value	Associated breed
GO:0070344	Regulation of fat cell proliferation	PID1	0.0000688058	Beni Guil
GO:0098743	Cell aggregation	COL11A1	0.0000688058	Beni Guil
GO:0008544	Epidermis development	BNC1	0.0000741624	Dman
GO:0007292	Female gamete generation	FSHR BNC1	0.0000741624	Dman
GO:0048469	Cell maturation	RXFP2	0.0000741624	Dman
GO:0009650	UV protection	SDF4	0.0000667761	Ouled Jellal
GO:0043112	Receptor metabolic process	SH3GLB1 LMBRD1	0.0000667761	Ouled Jellal
GO:1901568	Fatty acid derivative metabolic process	FADS1 PLA2G10	0.0000679420	Sardi
GO:0043586	Tongue development	BNC2	0.0000692810	Timahdite
GO:0040007	Growth	BNC2 ESR1	0.0000692810	Timahdite
GO:0060749	Mammary gland alveolus development	ESR1	0.0000692810	Timahdite
GO:0007631	Feeding behavior	MRAP2	0.0000692810	Timahdite

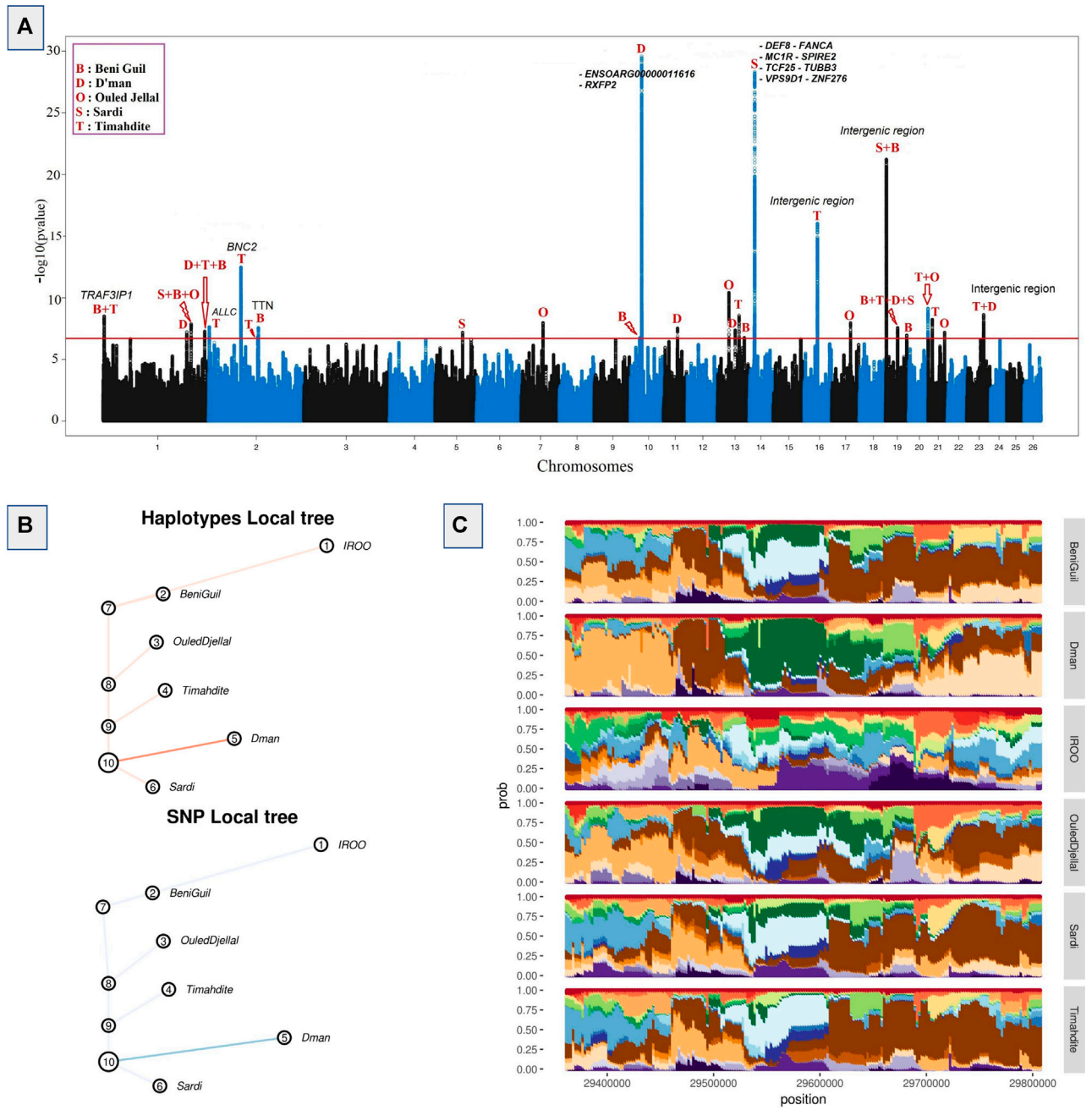

FIGURE 2 | Selective sweeps detected by hapFLK and plots for an example candidate region. **(A)** Manhattan plot depicting hapFLK scores in five Moroccan sheep breeds along autosomal chromosomes. Each dot represents a SNP. The horizontal red line represents the 0.1% FDR threshold of significance. **(B)** Local tree for the example region on Chromosome 10 in Dman. **(C)** Haplotype cluster plot of the same example region on chromosome 10 in Dman (Cluster frequencies plot).

Eydivandi et al. (2021) except for Dman which was the most inbred (F = 0.1). F_{ST} values showed no clear differentiation between Moroccan breeds with an average of ~0.002. These F_{ST} values are lower than those found in Welsh local sheep (Beynon et al., 2015) but still comparable to that of Russian local sheep (Deniskova et al., 2018), which were both estimated from WGS data.

Based on nucleotide diversity (π), we found that Dman was the most diversified of the Moroccan breeds. This could be explained by a lower intensity of selection, a higher founder population and wider breeding area for this breed. The slightly higher inbreeding in Dman would be explained by its sub-structuration in isolated Oases that leads to the mating of closely related individuals and by the longstanding use

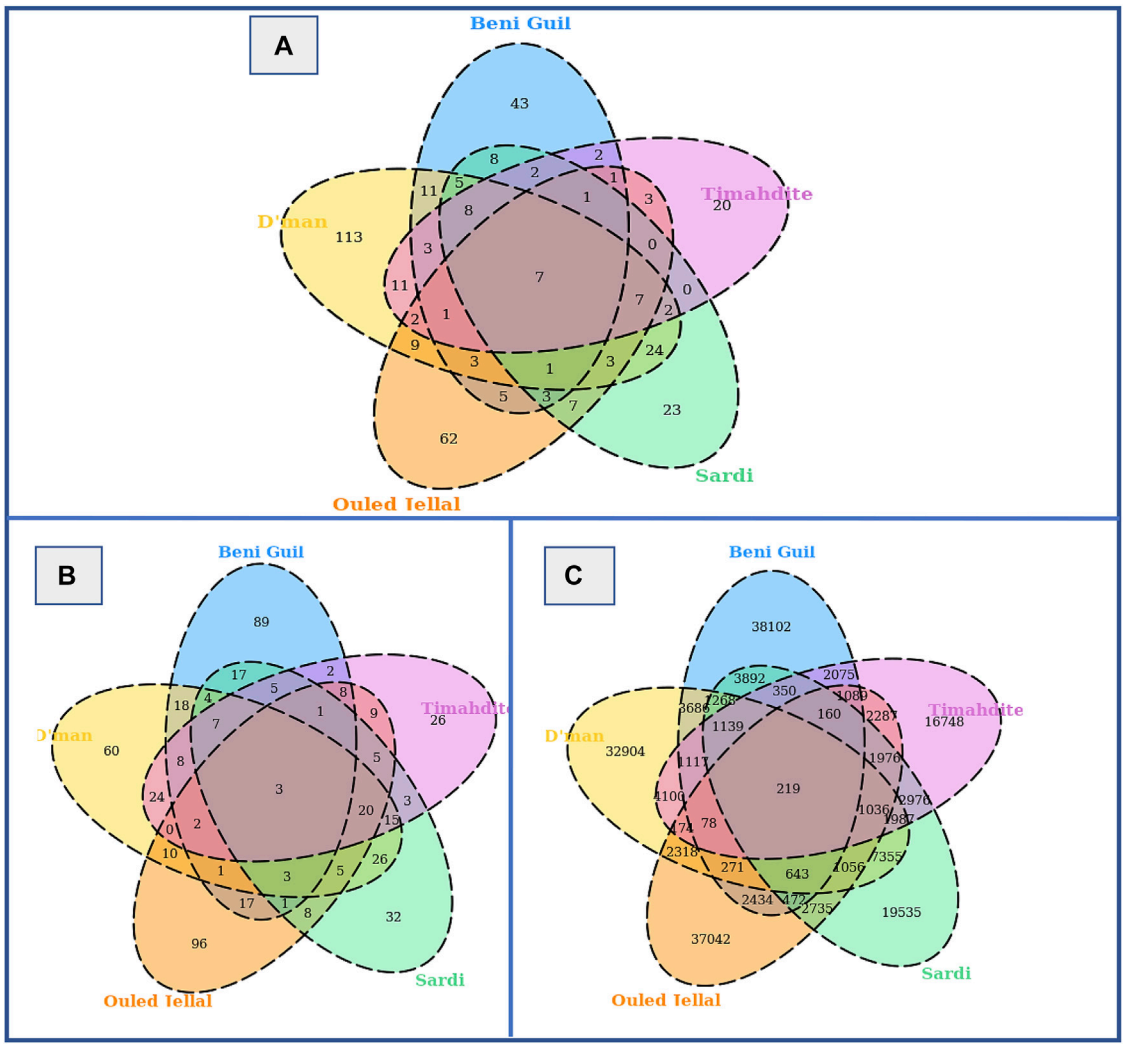

FIGURE 3 | Selective sweeps, SNPs and genes identified by both freqHMM and hapFLK methods. **(A)** Venn diagram of candidate regions under selection. **(B)** Venn diagram of genes under selection. **(C)** Venn diagram of SNPs under selection.

reproductive rams (Bouix and Kadiri, 1974; Darfaoui, 2000). However, these levels are still much lower than those measured by Beynon et al. (2015) on Welsh sheep breeds.

Furthermore, the past demographic dynamics (**Figure 1**), inferred with PopsizeABC and associating a pool of wild animals, was consistent with the occurrence of a domestication event 10,500 years ago in the Fertile Crescent (Vigne, 2011; Demirci et al., 2013). Indeed, earlier than 10.5 K years ago, effective sizes of all wild and domestic populations were similar, which is consistent with a common origin from the same ancestral species of *Ovis Orientalis* (Vigne, 2011). Around 10 K years B.P, the strong bottleneck linked to domestication is observed in all domestic breeds, Moroccan and cosmopolitan (**Figure 1**). The lasting decline in Moroccan breeds sizes between around 10 K and 2 K years would correspond to a long-term migration process consisting in colonizing gradually new constraining environments. Indeed, the colonization process lasted for millennia (Taberlet et al., 2008) and Morocco, which is located at the end of several migration routes, was marked by two main arrivals of domestic sheep; the first one occurred around 8,600 years B.P and would be associated with the first Berbers who settled in Morocco, while the second originated from Iberia around 7,100 years B.P (kandoussi et al., 2020).

Between 2,000- and 1,000-years B.P, all Moroccan breeds gradually increased in size which corroborates their common history and the absence of isolation obetween breeds at that time. The Timahdite breed shows a higher increase in effective population size for about five centuries in comparison with the other breeds. Inversely, the industrial breeds show a continuous decline during the last millennium which has been accentuated during the last century. Similarly, the effective size of

the wild mouflon gradually declined for about a millennium, which is in line with the degradation of their habitat by humans and the intensification of hunting activity (Rezaei, et al., 2010). More recently, over the last 700 years, the *Ne* of Moroccan and cosmopolitan breeds evolves in different ways with a huge drop of the cosmopolitan populations, which is consistent with the evolution of breeding practices which can go as far as the intensive use of artificial insemination. Finally, if we except Timahdite which behaved differently for the last 500 years, the other Moroccan breeds remain very close, which suggests their recent foundation.

Our estimates of the current *Ne* (**Supplementary Table S1**) and diversity parameters (**Table 1**) would illustrate the uniqueness of Moroccan local sheep breeds. The current *Ne* values for domestics (highest for Timahdite with Ne ~ 44 K) are very high when compared to what is reported in the literature for sheep (e.g., Maiwashe and Blackburn, 2004; Tapio et al., 2005; Beynon et al., 2015), while the estimates for mouflon populations (Ne ~ 2 K) are comparable to those reported by Beynon et al. (2015) in Welsh domestic sheep. We should however consider that estimates for recent time using popsize ABC can sometimes reach 5 to 10 times those based on pedigree or molecular data. This has been observed in a cattle breed (Boitard et al., 2016b). These results confirm actual measures of diversity using metrics such as nucleotide diversity. In any case, these results would also show the strong adaptive potential of Moroccan sheep breeds and the opportunity of implementing efficient programs for their breeding while maintaining this richness. Indeed, a high effective population size would be associated to a high intra-breed genomic diversity that include both adaptive and standing genetic variation. Furthermore, the latter is known to be a main driver of new adaptive traits that can be needed/useful in the context of new environmental pressures (Savolainen et al., 2013). Inversely, the very low effective size (Ne = 317) estimated from a mix of cosmopolitan breeds, despite their presence in very large numbers worldwide, show the huge threat they represent if their use to replace local breeds continues (Taberlet et al., 2008).

4.2 Selection Signatures in Moroccan Sheep

The selection signatures identified would shed light on the biological processes underlying both adaptive and zootechnical traits selected in each breed. Most of the regions identified under selection were intergenic (64%). This would illustrate the important role of regulation in the realization of biological processes and the expression of traits as reported by ENCODE Project Consortium et al. (2007). However, this could sometimes be due to hitch-hiking mechanisms and also to some limitations in the functional annotation of genomes (Benjelloun et al., 2015).

From the literature, the three genes identified as under selection in all five studied breeds (i.e., *HMGA2*, *RCOR1* and *SBF2*) are generally involved in cellular functioning. The expression of *HMGA2* is strongly associated with body size and growth in mice, humans and dogs (Webster et al., 2015; Vignali and Marracci, 2020; Yang et al., 2010). Furthermore, the inactivation of *HMGA2* in pigs resulted in a huge body-size reduction (Chung et al., 2018). The *RCOR1* gene has a role in transcriptional regulation, and is involved in repressing neuronal gene expression in non-neuronal cells (Coulson, 2005). Mutations in the *SBF2* gene were associated with autosomal recessive Charcot-Marie-Tooth Disease type 4B2 in humans (Senderek et al., 2003), and has been associated with growing traits in cattle (Jahuey-Martínez et al., 2016) and horse (Al Abri et al., 2018).

Besides, we identified many selection signatures (589 candidate regions) specific to one or a few breeds. Several GO Terms were enriched in genes putatively selected (**Supplementary Tables S4–S8**); they were associated with large categories such as: organ development, pigmentation pathways, proliferation and lipid metabolism. We discuss here the genes which role appears to be quite straightforward, which includes participation to functional processes related to morphology, pigmentation or skin coloration, zootechnical performance and prolificacy.

4.3 Sardi Breed

The *MC1R* gene, which is candidate in the Sardi breed, has been associated with a large panel of skin or coat colors (Sturm and Duffy, 2012). It may be involved in the particular coloration pattern specific to this breed and preferred by consumers (**Supplementary Figure S1C**) in relation to the religious ceremony of Aid Al Adha. This breed is characterized by a white head devoid of wool with black spots around the eyes, muzzle, paws (feet) and at the tips of the ears (Boujenane 1999, http://www.anoc.ma/les-races/races-ovines/sardi/, January 2021). Another candidate specific of Sardi is *SLC9A3*. The deficiency of this gene causes severe obstructive azoospermia and infertility in male mice (Wang et al., 2017).

4.4 Timahdite Breed

The *BNC2* and *EDN3* genes, identified in Timahdite, have been reported as potentially associated with skin pigmentation (Hider et al., 2013; Fariello et al., 2014), and may be involved in the specific skin coloration of this breed (**Supplementary Figure S1D**): a brown head, without spots neither black nor yellow, always very clear, sometimes, and extending behind the ears and into the trough area (Boujenane, 1999).

We also identified *ESR1* as a candidate, which is a major mediator of estrogen action and is strongly linked to bone mass and osteoporosis in mice (Nakamura et al., 2007; Börjesson et al., 2010). Some alleles were significantly associated with adult human height (Wood et al., 2014). This gene would play a role in growth, as well as another candidate, *MRAP2,* which modulates melanocortin receptor signaling and have been associated with severe obesity in human (Asai et al., 2013; Schonnop et al., 2016).

4.5 Dman Breed

RXFP2 has been associated with the horned/polled phenotype in many sheep breeds (Johnston et al., 2011; Dominik et al., 2012; Kijas et al., 2012; Wang et al., 2014). This gene was identified under selection only in Dman which is the only Moroccan breed

(**Supplementary Figure S1E**) with both polled males and females (Boujenane 2006; Boujenane et al., 2013). Also, the seasonal expression of this gene was correlated with the differentiation of Leydig cells in the testis (Hombach-Klonisch et al., 2004; Serranito et al., 2021), which would suggest a possible involvement in fertility. Similarly, *BNC1* and *FSHR* were identified, while Dman is known for its exceptional prolificacy in comparison with the other Moroccan breeds. It has an ovulation rate of 2.8 versus a maximum of 1.3 in the other breeds, with the ability to breed all year long with 190–350 days of interval between two lambings (Boujenane, 2006) depending on husbandry practices. It is also characterized by its early puberty (210 days in average, Boujenane 2006). The *FSHR* gene would be related to fertility, as the protein it encodes for is located in the testis and granulosa cells of the ovaries (Kwok et al., 2005), and is involved in follicle maturation and proliferation of granulosa cells (Sudo et al., 2002). Also, mutations in this gene were associated with the Polycystic Ovary Syndrome in humans which is characterized by obesity and anovulatory infertility (Gu et al., 2010; Laven, 2019). Besides, *BNC1* is a candidate which encodes for a zinc finger protein present, among other places, in the basal cell layer of the epidermis and in hair follicles and which is also expressed in the germ cells of testis and ovary (Luchi and Green, 1999). This protein is thought to play a regulatory role in keratinocyte proliferation and has been shown to be involved in premature ovarian failure and testicular premature aging (Zhang et al., 2018; Li et al., 2020). It is a crucial transcription factor for spermatogenesis and male fertility (Li et al., 2020). Thus, we could hypothesize thus that these last two genes play a role in the reproductive performance of the Dman breed.

4.6 Beni Guil Breed

The *TTN* gene was identified under selection (with 5 missense variants) exclusively in the Beni Guil breed. It has been associated with meat and carcass traits in pigs (Braglia et al., 2013), meat colour, pH and conductivity in loin 24 h postmortem (Wimmers et al., 2007). Similarly, *PID1* is identified as a candidate. This gene modulates insulin signaling and mitochondrial function in adipocytes and muscle cells, and was reported as a candidate gene for fat deposition, in humans, based on its high expression in adipose tissue of obese subjects in comparison with normal subjects (Wang et al., 2006). These two genes may be responsible for meat quality and the important fat percentage known of Beni Guil. Indeed, this breed is one of the best meat breeds in Morocco. Its carcass scores a high quality and a high fatness state with a white and firm cover fat, what makes it well appreciated by professionals and consumers (Belhaj et al., 2020). Since 2011, it has been certified by the PGI (Protected Geographical Indication) label (Belhaj et al., 2018), which represents the excellence of European agricultural food production.

4.7 Ouled Jellal Breed

The *SDF4* (also called *CAB45*) found in Ouled Jellal was involved in protecting against UV-induced damages (Zhu et al., 2008) and was revealed as a modulator of cell proliferation and tumor growth (Chen et al., 2016). Mutations in this gene may have occurred, as an adaptation, in response to the effect of sunlight (i.e., exposure to Ultraviolets) on the entirely white skin color of the Ouled Jellal breed (Chellig, 1992). in fact, fair/clear types of skin color are found, in humans, to be significantly more sensitive to UV rays than darker skin types (Halder and Bridgeman-Shah 1995; Hemminki et al., 2002).

4.8 Common Selection Signatures

Besides specific-breed candidates, the Microphthalmia-associated transcription factor (*MITF*) is a selected gene in both Beni Guil, Dman, Sardi and Timahdite. It is a basic transcription factor which regulates the differentiation and development of melanocytes and pigment cell-specific transcription of the melanogenesis enzyme genes (Saravanaperumal et al., 2014). This gene was also associated with eye and coat spotting color in some dog breeds (Rothschild et al., 2006; Stritzel et al., 2009). We could hypothesize that this gene may interact with other genes to produce different coat color patterns.

5 CONCLUSION

We characterized the neutral diversity, demographic history and selection signatures using whole genome variants in the five main sheep breeds from Morocco. Globally, these five breeds are not very genetically differentiated, but they show a high number of specific variants and very high effective population sizes unlike cosmopolitan breeds. This illustrates that Moroccan indigenous breeds are highly diversified and have thus the potential to develop key adaptive characteristics to face upcoming climate changes. This makes them valuable resources for conservation and future preservation of sheep even at the worldwide scale.

Investigations on selection signatures provided valuable insights about genes and biological processes targeted by selection, which are essentially involved in pigmentation, zootechnical performance, adaptation and reproduction traits. Many genomic regions highlighted here need further investigations to decipher their specific roles in the expression of phenotypes specific to the studied breeds. However, the outlier genomic variants identified here represent valuable candidates and should be conserved in priority when conceiving the upcoming genomic breeding programs. Those programs have to consider improving production as well as adaptation traits while maintaining the diversity of these local sheep breeds.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: ftp://ftp.ebi.ac.uk/pub/databases/nextgen/ovis/variants/genus_snps/.

ETHICS STATEMENT

The animal study was reviewed and approved by Data analyzed here have been produced previously within the EU FP7 NextGen project in respect of all ethical requirements.

AUTHOR CONTRIBUTIONS

BB, FP, AH, BS, and ADS designed and supervised the study. AO analyzed the data in the light of discussions with SB, BB and FB. AO together with BB wrote the manuscript. FP, SB, and ADS revised the manuscript. All authors contributed to discussions about the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.723599/full#supplementary-material

REFERENCES

Al Abri, M. A., Posbergh, C., Palermo, K., Sutter, N. B., Eberth, J., Hoffman, G. E., et al. (2018). Genome-Wide Scans Reveal a Quantitative Trait Locus for Withers Height in Horses Near the ANKRD1 Gene. *J. Equine Vet. Sci.* 60, 67–73.e1. doi:10.1016/j.jevs.2017.05.008

Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., et al. (2018). Convergent Genomic Signatures of Domestication in Sheep and Goats. *Nat. Commun.* 9, 813. doi:10.1038/s41467-018-03206-y

Asai, M., Ramachandrappa, S., Joachim, M., Shen, Y., Zhang, R., Nuthalapati, N., et al. (2013). Loss of Function of the Melanocortin 2 Receptor Accessory Protein 2 Is Associated with Mammalian Obesity. *Science* 341 (6143), 275–278. doi:10.1126/science.1233000

Association Nationale des Ovins et Caprins (2021). Sardi. Availale at: http://www.anoc.ma/les-races/races-ovines/sardi/ (Accessed January 2021).

Baek, B., Park, J., and Baek, K. (2010). Genetic Variations of Follicle Stimulating Hormone Receptor Are Associated with Polycystic Ovary Syndrome. *Int. J. Mol. Med.* 26, 107–112. doi:10.3892/ijmm_00000441

Belabdi, I., Ouhrouch, A., Lafri, M., Gaouar, S. B. S., Ciani, E., Benali, A. R., et al. (2019). Genetic Homogenization of Indigenous Sheep Breeds in Northwest Africa. *Sci. Rep.* 9, 7920. doi:10.1038/s41598-019-44137-y

Belhaj, K., Mansouri, F., Ben Moumen, A., Fauconnier, M.-L., Boukharta, M., Caid, H. S., et al. (2018). Physicochemical and Nutritional Characteristics of Béni Guil Lamb Meat Raised in Eastern Morocco. *Mnm* 11, 175–185. doi:10.3233/MNM-17195

Belhaj, K., Mansouri, F., Tikent, A., Ouchatbi, A., Boukharta, M., Hana Serghini, C., et al. (2020). Quality Characteristics of the Carcass of Beni-Guil Sheep, a Protected Geographical Indication Certified Product of Eastern Morocco: Preliminary Study. *Rev. Elev. Med. Vet. Pays Trop.* 73 (1), 21–26. doi:10.19182/remvt.31843

Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F. d. r., Coissac, E., Stucki, S., et al. (2015). Characterizing Neutral Genomic Diversity and Selection Signatures in Indigenous Populations of Moroccan Goats (*Capra hircus*) Using WGS Data. *Front. Genet.* 6, 107. doi:10.3389/fgene.2015.00107

Benjelloun, B., Boyer, F., Streeter, I., Zamani, W., Engelen, S., Alberti, A., et al. (2019). An Evaluation of Sequencing Coverage and Genotyping Strategies to Assess Neutral and Adaptive Diversity. *Mol. Ecol. Resour.* 19, 1497–1515.13070. doi:10.1111/1755-0998.13070

Benjelloun, B., Leempoel, K., Boyer, F., Stucki, S., Streeter, I., Orozco ter Wengel, P., et al. (2021). Multiple Adaptive Solutions to Face Climatic Constraints: Novel Insights in the Debate Over the Role of Convergence in Local Adaptation. *bioRxiv* [Preprint]. doi:10.1101/2021.11.18.469099

Benjelloun, B. (2015). *Diversité des génomes et adaptation locale des petits ruminants d'un pays méditerranéen: le Maroc. Thèse de doctorat. Biodiversité, Ecologie, Environnement*. Grenoble, France: Université Grenoble Alpes.

Beynon, S. E., Slavov, G. T., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., et al. (2015). Population Structure and History of the Welsh Sheep Breeds Determined by Whole Genome Genotyping. *BMC Genet.* 16, 65. doi:10.1186/s12863-015-0216-x

ENCODE Project ConsortiumBirney, E., Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., et al. (2007). Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project. *Nature* 447 (7146), 799–816. doi:10.1038/nature05874

Boitard, S., Boussaha, M., Capitan, A., Rocha, D., and Servin, B. (2016b). Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics* 203 (1), 433–450. doi:10.1534/genetics.115.181594

Boitard, S., Rodríguez, W., Jay, F., Mona, S., and Austerlitz, F. (2016a). Inferring Population Size History from Large Samples of Genome-wide Molecular Data - an Approximate Bayesian Computation Approach. *Plos Genet.* 12 (3), e1005877. doi:10.1371/journal.pgen.1005877

Boitard, S., Schlotterer, C., Nolte, V., Pandey, R. V., and Futschik, A. (2012). Detecting Selective Sweeps from Pooled Next-Generation Sequencing Samples. *Mol. Biol. Evol.* 29, 2177–2186. doi:10.1093/molbev/mss090

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics* 186 (1), 241–262. doi:10.1534/genetics.110.117275

Börjesson, A. E., Lagerquist, M. K., Liu, C., Shao, R., Windahl, S. H., Karlsson, C., et al. (2010). The Role of Estrogen Receptor α in Growth Plate Cartilage for Longitudinal Bone Growth. *J. Bone Miner Res.* 25 (12), 2690–2700. doi:10.1002/jbmr.156

Bouix, J., and Kadiri, M. (1975). Un des éléments majeurs de la mise en valeur des palmeraies: la race ovine DMan (One of the important factors in development of oases : the DMan breed of sheep). *Options Mdditerr* 26, 87–93.

Boujenane, I., Chikhi, A., Lakcher, O., and Ibnelbachyr, M. (2013). Genetic and Environmental Factors Affecting Perinatal and Preweaning Survival of D'man Lambs. *Trop. Anim. Health Prod.* 45, 1391–1397. doi:10.1007/s11250-013-0376-5

Boujenane, I. (1999). *Les ressources génétiques ovines au Maroc*. First Edition. Publisher: Actes Editions, Rabat, Morocco. 9981-801-41-0.

Boujenane, I. (2006). Reproduction and Production Performance of Moroccan Sheep Breeds. *CAB Rev.* 1, 014. doi:10.1079/PAVSNNR20061014

Braglia, S., Davoli, R., Zappavigna, A., Zambonelli, P., Buttazzoni, L., Gallo, M., et al. (2013). SNPs of MYPN and TTN Genes Are Associated to Meat and Carcass Traits in Italian Large White and Italian Duroc Pigs. *Mol. Biol. Rep.* 40, 6927–6933. doi:10.1007/s11033-013-2812-z

Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F. J., et al. (2015). Prospects and Challenges for the Conservation of Farm Animal Genomic Resources, 2015-2025. *Front. Genet.* 6, 314. doi:10.3389/fgene.2015.00314

Chellig, R. (1992). *Les Races Ovines Algériennes*. Ben-AknounAlger: Office des Publications Universitaires.

Chen, L., Xu, S., Xu, Y., Lu, W., Liu, L., Yue, D., et al. (2016). Cab45S Promotes Cell Proliferation through SERCA2b Inhibition and Ca2+ Signaling. *Oncogene* 35, 35–46. doi:10.1038/onc.2015.56

Chung, J., Zhang, X., Collins, B., Sper, R. B., Gleason, K., Simpson, S., et al. (2018). High Mobility Group A2 (HMGA2) Deficiency in Pigs Leads to Dwarfism, Abnormal Fetal Resource Allocation, and Cryptorchidism. *PNAS* 115 (21), 5420–5425. doi:10.1073/pnas.1721630115

Coulson, J. M. (2005). Transcriptional Regulation: Cancer, Neurons and the REST. *Curr. Biol.* 15, R665–R668. doi:10.1016/j.cub.2005.08.032

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Darfaoui, E. M. (2000). D'man Sheep Breeding Programme in Morocco. *ICAR Technical Series 3*. Rome, Italy: International Committee for Animal Recording, 319–330.

Demirci, S., Koban Baştanlar, E., Dağtaş, N. D., Pişkin, E., Engin, A., Özer, F., et al. (2013). Mitochondrial DNA Diversity of Modern, Ancient and Wild Sheep *(Ovis Gmelinii Anatolica)* from Turkey: New Insights on the Evolutionary History of Sheep. *PLoS ONE* 8 (12), e81952. doi:10.1371/journal.pone.0081952

Deniskova, T. E., Dotsev, A. V., Selionova, M. I., Kunz, E., Medugorac, I., Reyer, H., et al. (2018). Population Structure and Genetic Diversity of 25 Russian Sheep Breeds Based on Whole-Genome Genotyping. *Genet. Sel. Evol.* 50, 29. doi:10.1186/s12711-018-0399-5

Dominik, S., Henshall, J. M., and Hayes, B. J. (2012). A Single Nucleotide Polymorphism on Chromosome 10 Is Highly Predictive for the Polled Phenotype in Australian Merino Sheep. *Anim. Genet.* 43 (4), 468–470. doi:10.1111/j.1365-2052.2011.02271.x

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and Automated Whole-Genome Optical Mapping of the Genome of a Domestic Goat (*Capra hircus*). *Nat. Biotechnol.* 31 (2), 135–141. doi:10.1038/nbt.2478

Eydivandi, S., Roudbar, M. A., Karimi, M. O., and Sahana, G. (2021). Genomic Scans for Selective Sweeps through Haplotype Homozygosity and Allelic Fixation in 14 Indigenous Sheep Breeds from Middle East and South Asia. *Sci. Rep.* 11, 2834. doi:10.1038/s41598-021-82625-2

Fariello, M.-I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., Cristobal, M. S., et al. (2014). Selection Signatures in Worldwide Sheep Populations. *PLOS ONE* 9 (8), e103813. doi:10.1371/journal.pone.0103813

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting Signatures of Selection through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics* 193 (3), 929–941. doi:10.1534/genetics.112.147231

Food and Agriculture Organization of the United Nations (2020). Compare Data. Available at: http://www.fao.org/faostat/en/#compare/(Accessed in December, 2020).

Food and agriculture organization of the United Nations (2007). *Commission on Genetic Resources for Food and Agriculture*. Rome.The State of the World's Animal Genetic Resources for Food and Agriculture – in Brief

Garrison, E., and Marth, G. (2012). Haplotype-based Variant Detection from Short-Read Sequencing. arXiv:1207.3907.

Halder, R. M., and Bridgeman-Shah, S. (1995). Skin Cancer in African Americans. *Cancer* 75, 667–673. doi:10.1002/1097-0142(19950115)75:2+<667:aid-cncr2820751409>3.0.co;2-i

Ritchie, Hannah., and Roser, Max. (2017). "Meat and Dairy Production". Published online at OurWorldInData.org

Hemminki, K., Xu, G., Kause, L., Koulu, L. M., Zhao, C., and Jansen, C. T. (2002). Demonstration of UV-Dimers in Human Skin DNA *In Situ* 3 Weeks after Exposure. *Carcinogenesis* 23 (4), 605–609. doi:10.1093/carcin/23.4.605

Hider, J. L., Gittelman, R. M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J. M., et al. (2013). Exploring Signatures of Positive Selection in Pigmentation Candidate Genes in Populations of East Asian Ancestry. *BMC Evol. Biol.* 13, 150. doi:10.1186/1471-2148-13-150

Hombach-Klonisch, S., Schön, J., Kehlen, A., Blottner, S., and Klonisch, T. (2004). Seasonal Expression of INSL3 and Lgr8/Insl3 Receptor Transcripts Indicates Variable Differentiation of Leydig Cells in the Roe Deer Testis. *Biol. Reprod.* 71 (4), 1079–1087. doi:10.1095/biolreprod.103.024752

Hudson, R. R. (2002). Generating Samples under a Wright-Fisher Neutral Model of Genetic Variation, *Bioinformatics*, vol. 18 (pg. 337–338). doi:10.1093/bioinformatics/18.2.337

Iuchi, S., and Green, H. (1999). Basonuclin, a Zinc finger Protein of Keratinocytes and Reproductive Germ Cells, Binds to the rRNA Gene Promoter. *Proc. Natl. Acad. Sci.* 96 (17), 9628–9632. doi:10.1073/pnas.96.17.9628

Jahuey-Martínez, F. J., Parra-Bracamonte, G. M., Sifuentes-Rincón, A. M., Martínez-González, J. C., Gondro, C., García-Pérez, C. A., et al. (2016). Genomewide Association Analysis of Growth Traits in Charolais Beef Cattle1. *J. Anim. Sci.* 94 (11), 4570–4582. doi:10.2527/jas.2016-0359

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The Sheep Genome Illuminates Biology of the Rumen and Lipid Metabolism. *Science* 344, 1168–1173. doi:10.1126/science.1252806

Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., et al. (20112011). Genome-wide Association Mapping Identifies the Genetic Basis of Discrete and Quantitative Variation in Sexual Weaponry in a Wild Sheep Population. *Mol. Ecol.* 20, 2555–2566. doi:10.1111/j.1365-294x.2011.05076.x

Kandoussi, A., Boujenane, I., Auger, C., Serranito, B., Germot, A., Piro, M., et al. (2020). The Origin of Sheep Settlement in Western Mediterranean. *Sci. Rep.* 10, 10225. doi:10.1038/s41598-020-67246-5

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., et al. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *Plos Biol.* 10 (2), e1001258. doi:10.1371/journal.pbio.1001258

Kofler, R., and Schlotterer, C. (2012). Gowinda: Unbiased Analysis of Gene Set Enrichment for Genome-wide Association Studies. *Bioinformatics* 28, 2084–2085. doi:10.1093/bioinformatics/bts315

Kulski, J. K. (2016). Next-Generation Sequencing - an Overview of the History, Tools, and "Omic" Applications. *Next Generation Sequencing - Advances, Applications and Challenges*. doi:10.5772/61964

Kwok, H.-F., So, W.-K., Wang, Y., and Ge, W. (2005). Zebrafish Gonadotropins and Their Receptors: I. Cloning and Characterization of Zebrafish Follicle-Stimulating Hormone and Luteinizing Hormone Receptors- Evidence for Their Distinct Functions in Follicle Development1. *Biol. Reprod.* 72 (6), 1370–1381. doi:10.1095/biolreprod.104.038190

Laven, J. S. E. (2019). Follicle Stimulating Hormone Receptor (FSHR) Polymorphisms and Polycystic Ovary Syndrome (PCOS). *Front. Endocrinol.* 10, 23. doi:10.3389/fendo.2019.00023

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, J.-Y., Liu, Y.-F., Xu, H.-Y., Zhang, J.-Y., Lv, P.-P., Liu, M.-E., et al. (2020). Basonuclin 1 Deficiency Causes Testicular Premature Aging: BNC1 Cooperates with TAF7L to Regulate Spermatogenesis. *J. Mol. Cel Biol* 12 (1), 71–83. doi:10.1093/jmcb/mjz035

Li, J.-Y., Ying, Y.-Y., Qian, Y.-L., Chen, J.-P., Huang, Y., Liu, J., et al. (2021). BNC1 Promotes Spermatogenesis by Regulating Transcription of Ybx2 and Papolb via Direct Binding to Their Promotor Elements. *Reprod. Sci.* 28 (3), 785–793. doi:10.1007/s43032-020-00342-z

Maiwashe, A. N., and Blackburn, H. D. (2004). Genetic Diversity in and Conservation Strategy Considerations for Navajo Churro Sheep1. *J. Anim. Sci.* 82, 2900–2905. doi:10.2527/2004.82102900x

Mangul, S., Martin, L. S., Eskin, E., and Blekhman, R. (2019). Improving the Usability and Archival Stability of Bioinformatics Software. *Genome Biol.* 20, 47. doi:10.1186/s13059-019-1649-8

MARA (Ministère de l'Agriculture et de la Réforme Agraire) (1980). *Le Plan Moutonnier*. Rabat: Direction de l'Élevage.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Nakamura, T., Imai, Y., Matsumoto, T., Sato, S., Takeuchi, K., Igarashi, K., et al. (2007). Estrogen Prevents Bone Loss via Estrogen Receptor α and Induction of Fas Ligand in Osteoclasts. *Cell* 130, 811–823. doi:10.1016/j.cell.2007.07.025

Rezaei, H. R., Naderi, S., Chintauan-Marquier, I. C., Taberlet, P., Virk, A. T., Naghash, H. R., et al. (2010). Evolution and Taxonomy of the Wild Species of the Genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Mol. Phylogenet. Evol.* 54 (2), 315–326. doi:10.1016/j.ympev.2009.10.037

Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., et al. (2018). Revealing the Selection History of Adaptive Loci Using Genome-wide Scans for Selection: An Example from Domestic Sheep. *BMC Genomics* 19, 71. doi:10.1186/s12864-018-4447-x

Rothschild, M. F., Van Cleave, P. S., Glenn, K. L., Carlstrom, L. P., and Ellinwood, N. M. (2006). Association of MITF with white Spotting in Beagle Crosses and Newfoundland Dogs. *Anim. Genet.* 37, 606–607. doi:10.1111/j.1365-2052.2006.01534.x

Saravanaperumal, S. A., Pediconi, D., Renieri, C., and La Terza, A. (2014). Alternative Splicing of the Sheep MITF Gene: Novel Transcripts Detectable in Skin. *Gene* 552 (1), 165–175. 15. doi:10.1016/j.gene.2014.09.031

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological Genomics of Local Adaptation. *Nat. Rev. Genet.* 14, 807–820. doi:10.1038/nrg3522

Schonnop, L., Kleinau, G., Herrfurth, N., Volckmar, A.-L., Cetindag, C., Müller, A., et al. (2016). Decreased Melanocortin-4 Receptor Function Conferred by an Infrequent Variant at the Human Melanocortin Receptor Accessory Protein 2 Gene. *Obesity* 24 (9), 1976–1982. doi:10.1002/oby.21576

Senderek, J., Bergmann, C., Weber, S., Ketelsen, U. P., Schorle, H., Rudnik-Schöneborn, S., et al. (2003). Mutation of the SBF2 Gene, Encoding a Novel Member of the Myotubularin Family, in Charcot-Marie-Tooth Neuropathy Type 4B2/11p15. *Hum. Mol. Genet.* 12 (Issue 3,1), 349–356. doi:10.1093/hmg/ddg030

Serranito, B., Cavalazzi, M., Vidal, P., Taurisson-Mouret, D., Ciani, E., Bal, M., et al. (2021). Local Adaptations of Mediterranean Sheep and Goats through an Integrative Approach. *BioRxiv - the Preprint Server for Biology [Preprint]*. Available at: https://www.biorxiv.org/content/10.1101/2021.01.22.427461v1.full.pdf. doi:10.1038/s41598-021-00682-z

Storey, J. D., Bass, A. J., Dabney, A., and Robinson, D. (2015). Qvalue: Q-Value Estimation for False Discovery Rate Control. R package version 2.2.2. Available at: http://github.com/jdstorey/qvalue.

Storey, J. D., and Tibshirani, R. (2003). Statistical Significance for Genomewide Studies. *Proc. Natl. Acad. Sci.* 100, 9440–9445. doi:10.1073/pnas.1530509100

Stritzel, S., Wöhlke, A., and Distl, O. (2009). A Role of Themicrophthalmia-Associated Transcription Factorin Congenital Sensorineural Deafness and Eye Pigmentation in Dalmatian Dogs. *J. Anim. Breed. Genet.* 126, 59–62. doi:10.1111/j.1439-0388.2008.00761.x

Sturm, R. A., and Duffy, D. L. (2012). Human Pigmentation Genes under Environmental Selection. *Genome Biol.* 13 (139), 248. doi:10.1186/gb-2012-13-9-248

Sudo, S., Kudo, M., Wada, S., Sato, O., Hsueh, A. J., and Fujimoto, S. (2002). Genetic and Functional Analyses of Polymorphisms in the Human FSH Receptor Gene. *Mol. Hum. Reprod.* 8, 893–899. doi:10.1093/molehr/8.10.893

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6, e21800. doi:10.1371/journal.pone.0021800

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are Cattle, Sheep, and Goats Endangered Species? *Mol. Ecol.* 17 (1), 275–284. doi:10.1111/j.1365-294x.2007.03475.x

Tapio, M., Tapio, I., Grislis, Z., Holm, L.-E., Jeppsson, S., Kantanen, J., et al. (2005). Native Breeds Demonstrate High Contributions to the Molecular Variation in Northern European Sheep. *Mol. Ecol.* 14, 3951–3963. doi:10.1111/j.1365-294X.2005.02727.x

Vignali, R., and Marracci, S. (2020). HMGA Genes and Proteins in Development and Evolution. *Ijms* 1921 (2), 654. doi:10.3390/ijms21020654

Vigne, J.-D. (2011). The Origins of Animal Domestication and Husbandry: a Major Change in the History of Humanity and the Biosphere. *Comptes Rendus Biologies* 334, 171–181. doi:10.1016/j.crvi.2010.12.009

Wang, B., Zhang, M., Ni, Y.-h., Liu, F., Fan, H.-q., Fei, L., et al. (2006). Identification and Characterization of NYGGF4, a Novel Gene Containing a Phosphotyrosine-Binding (PTB) Domain that Stimulates 3T3-L1 Preadipocytes Proliferation. *Gene* 379, 132–140. doi:10.1016/j.gene.2006.05.008

Wang, X., Zhou, G., Li, Q., Zhao, D., and Chen, Y. (2014). Discovery of SNPs in RXFP2 Related to Horn Types in Sheep. *Small Ruminant Res.* 116, 133–136. doi:10.1016/j.smallrumres.2013.10.022

Wang, Y.-Y., Lin, Y.-H., Wu, Y.-N., Chen, Y.-L., Lin, Y.-C., Cheng, C.-Y., et al. (2017). Loss of SLC9A3 Decreases CFTR Protein and Causes Obstructed Azoospermia in Mice. *Plos Genet.* 13 (4), e1006715. doi:10.1371/journal.pgen.1006715

Webster, M. T., Kamgari, N., Perloski, M., Hoeppner, M. P., Axelsson, E., Hedhammar, Å., et al. (2015). Linked Genetic Variants on Chromosome 10 Control Ear Morphology and Body Mass Among Dog Breeds. *BMC Genomics* 16 (1), 474. doi:10.1186/s12864-015-1702-2

Weir, B. S., and Cockerham, C. C. (1984). Estimating F -Statistics for the Analysis of Population Structure. *Evolution* 38, 1358–1370. doi:10.2307/240864110.1111/j.1558-5646.1984.tb05657.x

Wimmers, K., Murani, E., Te Pas, M. F. W., Chang, K. C., Davoli, R., Merks, J. W. M., et al. (2007). Associations of Functional Candidate Genes Derived from Gene-Expression Profiles of Prenatal Porcine Muscle Tissue with Meat Quality and Muscle Deposition. *Int. Soc. Anim. Genet. Anim. Genet.* 38, 474–484. doi:10.1111/j.1365-2052.2007.01639.x

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height. *Nat. Genet.* 46, 1173–1186. doi:10.1038/ng.3097

Yang, T.-L., Guo, Y., Zhang, L.-S., Tian, Q., Yan, H., Guo, Y.-F., et al. (2010). HMGA2 Is Confirmed to Be Associated with Human Adult Height. *Ann. Hum. Genet.* 74 (1), 11–16. doi:10.1111/j.1469-1809.2009.00555.x

Zhang, D., Liu, Y., Zhang, Z., Lv, P., Liu, Y., Li, J., et al. (2018). Basonuclin 1 Deficiency Is a Cause of Primary Ovarian Insufficiency. *Hum. Mol. Genet.* 27 (21), 3787–3800. doi:10.1093/hmg/ddy261

Zhu, Y., Wang, Q., Xu, W., and Li, S. (2008). The Ethanol Response Gene Cab45 Can Modulate the Impairment Elicited by Ethanol and Ultraviolet in PC12 Cells. *J. Genet. Genomics* 35 (3), 153–161. doi:10.1016/S1673-8527(08)60021-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 December 2021
doi: 10.3389/fgene.2021.657282

Unraveling Admixture, Inbreeding, and Recent Selection Signatures in West African Indigenous Cattle Populations in Benin

Sèyi Fridaïus Ulrich Vanvanhossou[1], *Tong Yin*[1], *Carsten Scheper*[1], *Ruedi Fries*[2], *Luc Hippolyte Dossa*[3] *and Sven König*[1]*

[1]*Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, Gießen, Germany,* [2]*Chair of Animal Breeding, Technische Universität München, Freising-Weihenstephan, Germany,* [3]*School of Science and Technics of Animal Production, Faculty of Agricultural Sciences, University of Abomey-Calavi, Abomey-Calavi, Benin*

The Dwarf Lagune and the Savannah Somba cattle in Benin are typical representatives of the endangered West African indigenous Shorthorn taurine. The Lagune was previously exported to African and European countries and bred as Dahomey cattle, whereas the Somba contributed to the formation of two indigenous hybrids known as Borgou and Pabli cattle. These breeds are affected by demographic, economic, and environmental pressures in local production systems. Considering current and historical genomic data, we applied a formal test of admixture, estimated admixture proportions, and computed genomic inbreeding coefficients to characterize the five breeds. Subsequently, we unraveled the most recent selection signatures using the cross-population extended haplotype homozygosity approach, based on the current and historical genotypes. Results from principal component analyses and high proportion of Lagune ancestry confirm the Lagune origin of the European Dahomey cattle. Moreover, the Dahomey cattle displayed neither indicine nor European taurine (EUT) background, but they shared on average 40% of autozygosity from common ancestors, dated approximately eight generations ago. The Lagune cattle presented inbreeding coefficients larger than 0.13; however, the Somba and the hybrids (Borgou and Pabli) were less inbred (≤0.08). We detected evidence of admixture in the Somba and Lagune cattle, but they exhibited a similar African taurine (AFT) ancestral proportion (≥96%) to historical populations, respectively. A moderate and stable AFT ancestral proportion (62%) was also inferred for less admixed hybrid cattle including the Pabli. In contrast, the current Borgou samples displayed a lower AFT ancestral proportion (47%) than historical samples (63%). Irrespective of the admixture proportions, the hybrid populations displayed more selection signatures related to economic traits (reproduction, growth, and milk) than the taurine. In contrast, the taurine, especially the Somba, presented several regions known to be associated with adaptive traits (immunity and feed efficiency). The identified subregion of bovine leukocyte antigen (BoLA) class IIb (including *DSB* and *BOLA-DYA*) in Somba cattle is interestingly uncommon in other African breeds, suggesting further investigations to understand its association with specific adaptation to endemic diseases in Benin.

OPEN ACCESS

Edited by:
Olivier Hubert Hanotte, University of Nottingham, United Kingdom

Reviewed by:
Sophie Thevenon, CIRAD, France
Abdulfatai Tijjani, International Livestock Research Institute, Ethiopia

***Correspondence:**
Sven König
Sven.Koenig@agrar.uni-giessen.de

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *22 January 2021*
Accepted: *07 October 2021*
Published: *08 December 2021*

Citation:
Vanvanhossou SFU, Yin T, Scheper C, Fries R, Dossa LH and König S (2021) Unraveling Admixture, Inbreeding, and Recent Selection Signatures in West African Indigenous Cattle Populations in Benin. Front. Genet. 12:657282. doi: 10.3389/fgene.2021.657282

Overall, our study provides deeper insights into recent evolutionary processes in the Beninese indigenous cattle and their aptitude for conservation and genetic improvement.

Keywords: Dahomey cattle, genomic inbreeding, admixture, haplotype, BoLA genes, adaptation

INTRODUCTION

Western Africa represents a reservoir of the unique and diverse African animal genetic resources, due to a complex history including migration, dispersion, natural and artificial selection, and crossbreeding (Hanotte et al., 2002; Stock and Gifford-Gonzalez, 2013). The region is the exclusive current habitat for indigenous taurine cattle (*Bos taurus*) on the continent (Mwai et al., 2015). African humpless taurine cattle are the earliest known cattle populations on the continent according to historical and archaeological evidence (Blench and MacDonald, 1999). These animals share early common ancestors with European taurine (EUT) cattle dated to before the domestication process (Ho et al., 2008; Murray et al., 2010). They are subdivided into two subgroups, the Longhorn and the Shorthorn taurine. The origins of the two populations are still controversial, but it is scientifically accepted that they have been separately introduced into West Africa several millennia BC (Epstein, 1971; Payne and Hodges, 1997; Hanotte et al., 2002). The N'Dama is the unique reported Longhorn taurine breed, whereas the humpless Shorthorns include the Savannah breeds (Baoulé, Doayo, and Somba) and the Dwarf forest breeds (Dwarf Muturu, Liberian Dwarf, and Lagune), which are barely characterized across West African countries (Rege, 1999).

The Somba and Lagune represent indigenous African Savannah and Dwarf taurine in Benin, respectively. Both are characterized by stocky animals, resistant to diseases (especially trypanosome), and able to survive and produce in harsh environments (Dossa and Vanvanhossou, 2016; Ahozonlin et al., 2020). The Somba cattle are the typical ancestral residue of West African Savannah Shorthorns that migrated to Togo, Ghana, and the Ivory-coast (Rege, 1999). They have been preserved from admixture until recent past in the Atacora mountain area in Northwest Benin (Hall et al., 1995; Rege, 1999). The Lagune cattle are present not only in Southern Benin but also in the coastal areas and near lagoons in West and Central Africa, as indicated by their name (Rege et al., 1994). The Lagune are described as the smallest of the African taurine (AFT) cattle (93 cm of average height at withers) and acquired their Dwarf phenotype through adaptation to environmental constraints in their belt (Rege et al., 1994). According to previous studies, this breed is genetically different from other Shorthorn cattle breeds (Moazami Goudarzi et al., 2001; Gautier et al., 2009). Berthier et al. (2015) reported higher trypanotolerance with lesser anemic condition in Lagune animals in comparison to the Baoule. However, genomic regions affected by divergent selection and environmental adaptation of the Lagune and the Somba remain unknown.

The Lagune cattle from Benin had been exported during the early 20th century (around 1904) to different African countries such as Zaire (current Democratic Republic of Congo), Zambia, Gabon, as well as to Europe where they are known as Dahomey cattle (Rege et al., 1994; Porter et al., 2016), because Dahomey is the former name of the country of Benin. The European Dahomey cattle are presently kept and bred by a breeder association involving 77 farmers (Verband Europäisches Dahomey-Zwergrind, 2019) from four European countries including Germany, Austria, Czech Republic, and Switzerland (http://www.dahomey-zwergrind.com). To date, neither scientific study nor census addressed these animals. According to the breeder association, the current population is characterized by a small size (adult body weight between 150 and 300 kg and sacrum height from 80 to 105 cm), short horn, easy calving, and good temperament. These characteristics are similar to those of the original Lagune population in Benin and may indicate that the Dahomey population conserved its purity through the past decades. However, the European climate is different from the one encountered in Benin, i.e., less heat stress, but varying temperatures between winter and summer, and also varying sunlight duration. In addition, the new production environment of the Dahomey cattle implies the reduction of disease infections with potentially improved feeding and housing systems in opposition to the harsh production conditions in Benin (characterized by food and water scarcity and the risk for disease infections). Consequently, this geographical isolation may have altered frequencies for alleles and haplotypes associated with specific genetic features in the Dahomey population.

The livestock production systems in Benin like in the other African countries have experienced a drastic transformation in the last decades (Mwai et al., 2015; Houessou et al., 2019a). New challenges arise in African pastoral regions along with the increasing demand for animal products, food insecurity, and poverty. Indeed, demographic explosion increases the pressures on animal product markets in West Africa and accentuates the need to develop more productive breeds. Simultaneously, anthropogenic activities such as deforestation and urbanization associated with climate changes have shrunk feed and water resources and increased disease challenges, forcing herders to develop new breeding strategies such as herd mobility or feed supplementation (Houessou et al., 2019a; Ahozonlin and Dossa, 2020). In this context, locally adapted animals are required to cope with the various instabilities in production environments, but the indigenous Shorthorn cattle in Benin are increasingly threatened. The trypanotolerant taurine cattle (without any Zebu ancestry) reported in the region by MacHugh et al. (1997) and Hanotte et al. (2002) are progressively replaced by crossbreeds and trypanosusceptible Zebu cattle, including White Fulani, Sokoto Gudali, and Red Bororo (Dossa and Vanvanhossou, 2016; Houessou et al., 2019b; Ahozonlin and Dossa, 2020). In consequence, significant adoption of prophylactic measures is observed in Beninese herds dominated by crossbreed and Zebu cattle in comparison to other local herds (Houessou et al., 2020).

In regard to the increasing uncontrolled crossbreeding, two other indigenous hybrid cattle, the Borgou and Pabli, are also endangered in Benin. The Pabli originally reported in the region of Kerou (Northwest Benin) is scarcely described and is considered as extinct by absorption from crossbreeding with Borgou (Belemsaga et al., 2005; Egito et al., 2007). However, recent evaluations revealed the existence of a population with slight genetic differences in comparison to the Borgou (Scheper et al., 2020). The Borgou cattle mainly located in the Nord-Eastern and Central regions of Benin were described as an intermediate crossbreed between taurine and indicine cattle (Porter, 1991). The origin of Borgou is still under discussion, but it is assumed to be a product of an admixture between the taurine Somba and the White Fulani Zebu (Porter, 1991; Belemsaga et al., 2005). Flori et al. (2014) characterized the admixture in Borgou as an efficient short-term adaptation strategy to environmental conditions and disease pressures and identified different genomic regions involved in adaptive mechanisms. However, the Borgou cattle population is now highly affected by Zebu cattle influence due to admixture (Scheper et al., 2020). The increased crossbreeding with trypanosusceptible Zebu cattle over a short period questions adaptive features such as resistance to diseases in Borgou, Somba, and Lagune taurine populations. It is therefore urgent to gain more insight into the genetic composition of the current Beninese cattle population to ensure the sustainability of livestock production.

To understand the genetic architecture underlying adaptive and productive abilities of various breeds evolving in challenging environments, selection signature analyses have the potential to detect specific genomic footprints in terms of differences in marker allele frequencies or in haplotypic mosaicism (Qanbari and Simianer, 2014; Aliloo et al., 2020; Kim et al., 2020). According to Freedman et al. (2016), admixture and subsequent recombination break down parental haplotypes and expand mosaic regions through the genome. Thus, extensive admixture in local breeds may reduce the signal of strong homozygosity of extended haplotypes involved in adaptive processes. Moreover, the assessment of homozygous-by-descent (HBD) segments or runs of homozygosity (ROH) is valuable to describe a population and investigate demographic histories. HBD are chromosome segments inherited from an ancestor and may be exploited to estimate inbreeding (Leutenegger et al., 2003).

In this study, we aimed to genetically characterize the current endangered cattle breeds in Benin and evaluate the effects of admixture and environmental factors related to late changes in production systems. Specifically, we first assessed the admixture level in the different populations and compared them with historical samples. Second, through selection signature analysis, we investigated genomic regions and biological mechanisms involved or affected by recent natural or artificial selection and admixture in the Beninese cattle breeds. Subsequently, we investigated the genetic differentiation resulting from the isolation of the Dahomey cattle.

MATERIALS AND METHODS

Sampling Design and Genotype Data

Hair samples were collected from 449 animals from the four local breeds (Borgou 181, Pabli-Kerou 58, Lagune 150, and Somba 60) in Beninese local farms in 2016 and 2017. The sampling locations were identified ensuring a large coverage of the main geographical distribution of the breeds in the country (**Figure 1**). In each herd, one animal representing a (pure) local breed was identified by the farmer and selected. To reduce the relatedness between the samples, one or two animals per village were sampled in 90% of the 298 investigated villages. In the remaining villages, additional animals (three to ten animals in total per village) were sampled for the assessment of further socioeconomic and ecological factors including transhumance and climate (see **Supplementary Table S1A** and Scheper et al. (2020) for more details). Furthermore, thirty Dahomey cattle were sampled in 2019 in 30 different herds across Europe (Germany 23, Austria 4, Switzerland 2, and Czech Republic 1).

The sampled animals were genotyped with the Illumina BovineSNP50 BeadChip. SNP variants with a call rate <95% and a minor allele frequency (MAF < 0.01) as well as animals with a high percentage of missing genotypes (call rate < 90%) were filtered out in PLINK (Purcell et al., 2007). The final genotype dataset including 40,109 SNP from 460 cattle was used for different analyses.

Description of the Current Beninese Cattle Population

Population Structure

The genetic structure of our samples was first assessed by applying the fastStructure software (Raj et al., 2014). The algorithm was run using the simple prior model and testing different numbers of populations (K) ranging from 2 to 10. Afterward, the optimal K was defined using the chooseK.py function. The identified optimal K was finally incorporated into the logistic prior model, to estimate individual genetic proportions describing the population structure, following the recommendations by Raj et al. (2014).

Second, we applied the unsupervised k-means clustering in the "adegenet" R package (Jombart, 2008; Jombart and Ahmed, 2011) to define homogeneous genetic clusters and to exclude potential outgroups animals from the breeds due to sampling. The find.clusters function was applied, considering the following parameters: maximum number of clusters (max.n.clust) = 100, number of principal components (n.pca) = 100, number of iterations in each run (n.iter) = 10^9, number of starting centroids in each run (n.start) = 30, and the default parameters for the remaining arguments (see package description for further details, Jombart, 2008; Jombart and Ahmed, 2011). Finally, the optimal number of genetic clusters was defined by choosing the k-value based on the Bayesian information criterion (BIC), as recommended by the authors (Jombart 2008; Jombart and Ahmed 2011).

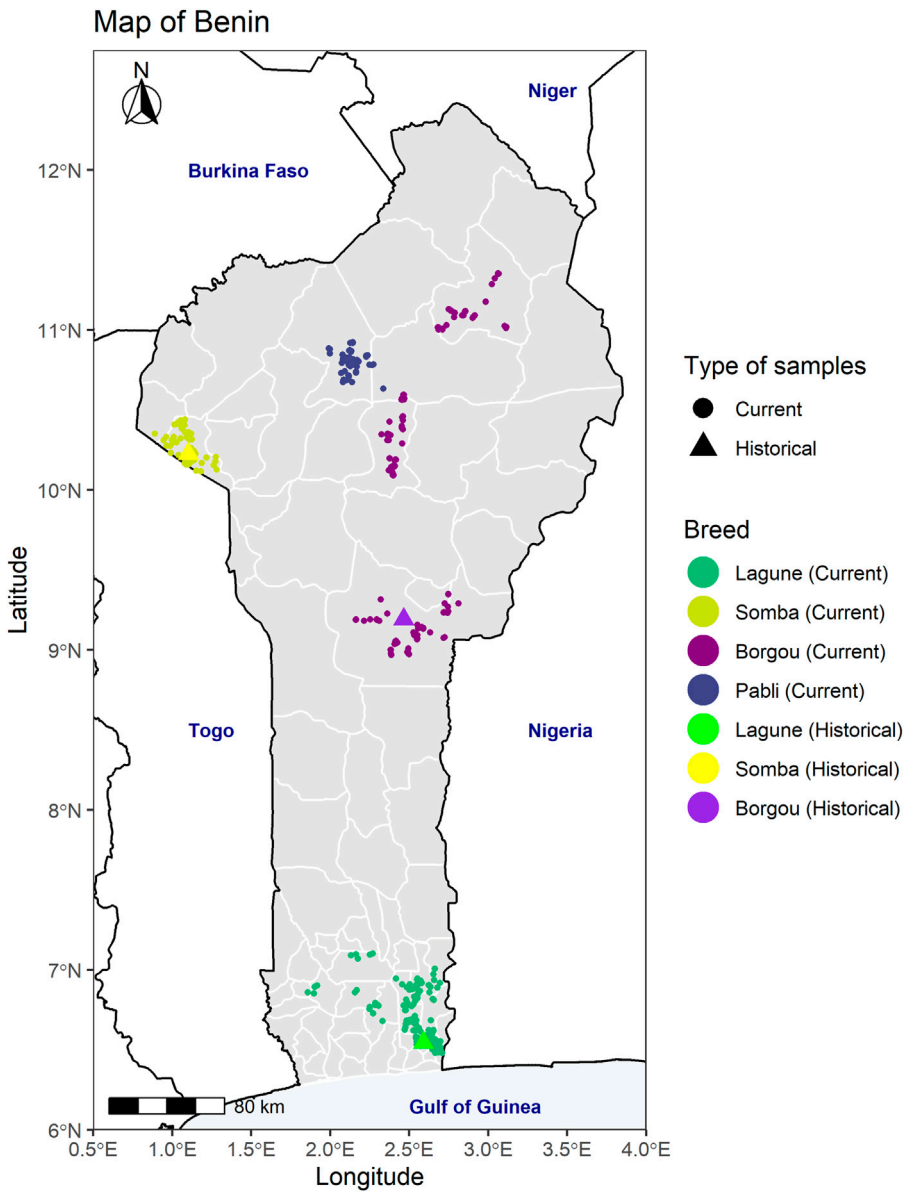

FIGURE 1 | Map of Benin with the sampling locations of the different Beninese cattle populations investigated in the country. For the historical data, the geographical coordinates of each sample are not available. In consequence, the coordinates of the sampling regions as described by Moazami Goudarzi et al. (2001) were plotted: Porto-Novo for the Lagune, Boukombé for the Somba, and the department of Borgou for the Borgou samples.

Genome-Wide Inbreeding Coefficients

We utilized the "RzooRoH" package (Druet and Gautier, 2017; Bertrand et al., 2019) in R to identify HBD and infer the contribution of ancestors from different past generations to inbreeding in our current cattle populations. The software applies a hidden Markov model (HMM) to fit the individual genome as a succession of HBD and non-HBD segments, considering marker allele frequencies, genotyping error rates, and intermarker genetic distances (Druet et al., 2020). In a multiple HBD class model, HBD segments are assigned to K age-related classes associated with different rates (R_k). The class rates (R_k) are related to an expected length and exponential distribution of HBD segments. They are equivalent to twice the number of generations to the common ancestor. The proportion of the genome belonging to a specific HBD class is defined as "realized autozygosity" in the respective HBD class. Similarly, the genome-wide inbreeding coefficient is estimated as the cumulative fraction of the genome that is HBD in the current population with respect to an ancestral base population. As suggested by the authors of the "RzooRoH" package (Druet and Gautier, 2017; Bertrand et al., 2019), we applied a default "MixKR" model with 10 classes (9 HBD classes and 1 non-HBD class) and predefined rates (R_k) 2, 4, 8, 16, 32, 64, 128, 256, 512, and 512, respectively.

Comparison of Current Beninese Cattle Population With Historical Data

Extension of the Genotype Dataset With Historical Data

We contrast the genotype of the current Beninese cattle population with available historical genotypes, retrieved from the WIDDE database (Sempéré et al., 2015). In a first step, the genotype data of 133 animals from Beninese cattle breeds (Borgou 45, Lagune 44, and Somba 44) sampled in different locations (see **Figure 1** and **Supplementary Table S1B**) between 1997 and 2000 (Moazami Goudarzi et al., 2001; Gautier et al., 2009) were merged to the genotype data of our new samples. The dataset with 42,802 SNP from 588 animals was submitted to quality control in PLINK (Purcell et al., 2007), with the parameters --geno 0.05, --mind 0.25, and --maf 0.01. After quality control, the final genotype dataset including 32,533 SNP from 586 cattle was used to apply principal component analysis and unsupervised k-means clustering. Second, further African Shorthorn taurine breed (Baoulé), African Longhorn taurine breed (N'Dama), African indicine breeds (Zebu White Fulani and Zebu Bororo), Asian indicine breeds (Gir, Brahman, Ongole, and Nellore), EUT breeds (Angus, Holstein, Charolais, Shorthorn, and Salers), and African hybrid (Kuri) were included. This extension resulted in 52,341 SNP from 997 animals. However, only 30,637 SNP from 997 animals passed the quality control procedure (with the parameters --geno 0.05, --mind 0.25, and --maf 0.01 in PLINK) and were considered in the principal component analysis, unsupervised k-means clustering, and admixture tests (see below for more details). Finally, 21 Gayal (*Bos frontalis*, Gao et al., 2017) samples were added to the previously extended dataset for the estimation of admixture proportion through the calculation of the f_4-ratio. Likewise, we used the same filtering parameters (--geno 0.05, --mind 0.25, and --maf 0.01) in PLINK. The genotype data consisted of 52,364 SNP and 30,228 SNP from 1,018 animals before and after quality control, respectively. All the genotype data exploited in this study are fully described and available in a public repository (see the section "*Data Availability Statement*"). Before being merged with our samples, the historical genotype data were remapped with the current reference assembly ARS-UCD1.2/bosTau9 (GenBank Bioproject PRJNA391427) and flipped (with the --flip flag in PLINK) to correct for strand inconsistency (Purcell et al., 2007).

Principal Component Analyses and Clustering

The genetic divergence between the Beninese cattle populations and other AFT was assessed using the principal component analysis (PCA) in PLINK (Purcell et al., 2007) and using the unsupervised k-means clustering in "adegenet" (Jombart, 2008; Jombart and Ahmed, 2011) in R. The analyses were subsequently repeated considering EUT and indicine breeds in order to investigate potential introgression of these breeds in our samples, especially in Dahomey. The optimal number of genetic clusters for the unsupervised k-means clustering was defined using the same approach as described above.

Admixture and Estimation of Ancestral Proportion

We tested admixture and inferred admixture proportion in the current and historical Beninese indigenous cattle populations by means of the three-population test (F_3) and the F_4-ratio estimation in Admixtools, respectively (Patterson et al., 2012). The methods are based on f-statistics, corresponding to the average of F values over markers. The Admixtools software uses allele frequencies of the available samples to estimate unbiased f_3 and f_4 statistics (see Patterson et al., 2012, for more details). The three-population test is a formal test of admixture. Negative f_3 (X; B, C) indicates that the allele frequencies in population X tend to be intermediate between B and C and indicates admixture in X populations from populations related to B and C. The f_4 statistics were used to infer admixture proportions in the Beninese cattle populations (target populations) through the calculation of the F_4-ratio or alpha (**Eq. 3**; Patterson et al., 2012). We considered the phylogeny model introduced by Flori et al. (2014) to estimate alpha, as AFT ancestral proportions in our target populations.

$$\text{Alpha} = \frac{f_4(A, O; B, C)}{f_4(A, O; X, C)}, \tag{3}$$

where X is the target population, A is the Salers population as EUT ancestral, O is the Gayal as outgroup population, B is Baoulé (BAO) as the Shorthorn AFT reference population, and C is Gir as the indicine reference population.

Detection of Regions Under Recent Selection—Gene Annotation and QTL Mapping

We investigated divergence in extended haplotype homozygosity between the current and historical Beninese cattle populations, in order to detect positive selection signatures or genomic footprints left by recent demographic events. Specifically, we compared DAH_cur, LA_cur, and LA_out with LA_hist; SO_cur with SO_hist; and BO_cur and Adm_cur with BO_hist. The cross-population–extended haplotype homozygosity (XP-EHH) approach (Sabeti et al., 2002) was implemented in the "Rehh" package in R (version 3.1.2; Gautier and Vitalis, 2012; Gautier et al., 2017). Prior to the analyses, the main genotype was phased and missing variants were imputed in Beagle 5.1 (Browning et al., 2018). Afterward, integrated site-specific extended haplotype homozygosity (iES) was calculated for each focal marker in the respective population with the "Rehh" package. The XP-EHH statistics were computed as the standardized log ratio of the iES of the two populations. One-sided *p*-values were estimated for XP-EHH to identify strong extended homozygosity in our current populations relative to the respective historical populations (as described above). The estimated *p*-values were subsequently adjusted (Benjamini and Hochberg, 1995) in R. Variants with adjusted *p*-values ≤ 0.05 were considered as significant. In addition, we used a conservative approach similar to those described in previous studies (Flori et al., 2014; Bertolini et al., 2020; Han et al., 2020) and defined as candidate regions under selection, sliding windows of 0.5 MB spanning at least three significant markers. The calc_candidate_regions function in "Rehh" was used to detect the candidate regions. Neighboring

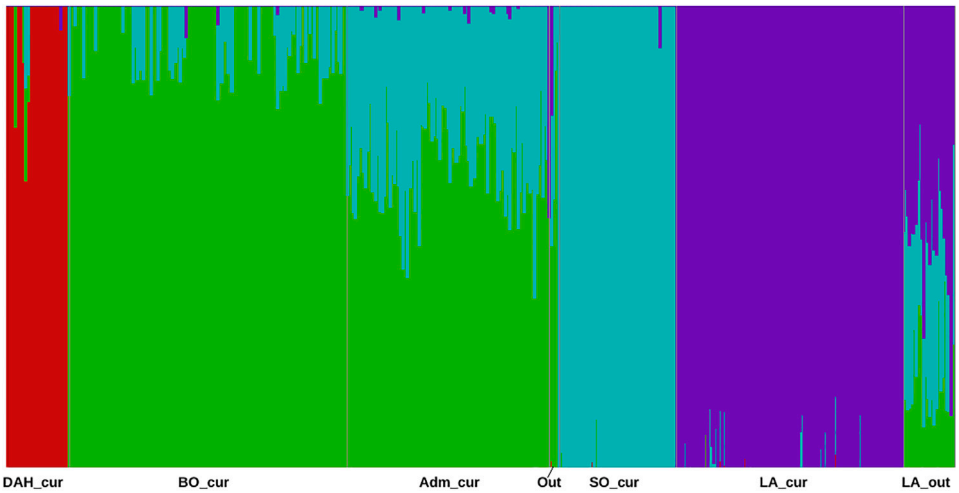

FIGURE 2 | Barplot for admixture proportions inferred by fastStructure in the Beninese cattle populations. Dah_cur = current Dahomey; LA_cur = current pure Lagune; LA_out = admixed Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou, Out = Outgroups (see **Table 1** for more information).

windows with significant SNP were merged to one candidate region. The peak of each candidate region, i.e., the SNP with the lowest adjusted *p*-values in a region, was considered as core SNP.

Genes located in the candidate regions for positive selection signatures were annotated from the Ensembl genome database (2020) (http://www.ensembl.org/biomart/martview/) and submitted to gene ontology (GO) enrichment analysis using the Gene Ontology web-tool (Ashburner et al., 2000; Gene Ontology Consortium, 2019). Fisher's exact threshold of *p*-values < 0.01 was considered to identify overrepresented GO terms for biological processes and Reactome pathways. In addition, QTL that overlapped with the candidate regions under selection were mapped from the online data analysis tools of the cattle database (Hu et al., 2019) and summarized in major production and functional categories: milk, carcass quality, reproduction, body weight, conformation, feed intake, heat tolerance, and health traits (see **Supplementary Table S2** for more details). Subsequently, for each population, we calculated the frequency of the QTL which is equal to the number of candidate regions overlapping with the given QTL per the total number of candidate regions.

RESULTS

Population Structure and Inbreeding in the Current Beninese Cattle Population

Structure of the Current Population

The analyses with the fastStructure algorithm revealed four model components to explain the population structure of our samples. Similarly, the model complexity that maximizes marginal likelihood was equal to four. Considering the posterior mean of admixture proportion in the logistic model, we identified three components representing Somba, Lagune, and Dahomey samples, respectively (**Supplementary Table S3**). The fourth component was mainly made based on Borgou animals. However, some hybrid samples displayed genetic proportions across two or more components. This structure was similar to those obtained with adegenet (**Figure 2**). The unsupervised k-means clustering presented an optimal K (number of clusters, **Supplementary Figure S1**) equal to 6 and displayed the hybrid animals in different three clusters (**Supplementary Figure S1** and **Table 1**). Consequently, a large Lagune group named LA_cur (n = 110) was separated from other Lagune samples called LA_out (n = 25). Forty-two Borgou and 56 Pabli animals (except two outgroups) formed a homogeneous cluster named Adm_cur (n = 98). The remaining Borgou animals were allocated to the BO_cur cluster. The Dahomey and the Somba correspond to the DAH_cur and SO_cur, respectively (**Table 1**). LA_cur, LA_out, DAH_cur, SO_cur, BO_cur, and Adm_cur were considered as populations for further analyses instead of original breed assignments.

Genomic Inbreeding Coefficients

The genome-wide contributions and genomic inbreeding coefficients with respect to different age-related HBD classes in the Beninese cattle populations are presented in **Figures 3A,B**, respectively. Overall, 40% of the genome of DAH_cur samples was HBD and was related only to HBD classes with $R_k \leq 16$. We observed a major contribution of autozygosity from the HBD class with $R_k = 8$, accounting for 63.32% of the total HBD proportion in DAH_cur. The HBD classes with $R_k = 4$ and $R_k = 16$ contributed to 15.70 and 16.80% of the overall autozygosity, respectively. The estimated genomic inbreeding coefficients were 0.17 and 0.13 in LA_cur and LA_out, respectively, when considering all HBD classes (i.e., the most remote base population). The major contribution of autozygosity in LA_cur came from the ancient HBD class with $R_k = 128$ (40% of the HBD proportion). Recent HBD classes with $R_k = 4$ and $R_k = 8$ explained 13.78 and 7.80% of the total autozygosity in LA_cur, respectively, whereas 27.67 and 26.11% of the overall autozygosity in LA_out were derived from the two classes. Tracing back to the oldest ancestors, the fraction of the genome that was HBD in SO_cur was equal to 0.08 and was mainly originated from very ancient HBD

TABLE 1 | Genetic clusters generated from unsupervised k-means clustering applied on the current Beninese cattle population.

Generated genetic clusters		Animals	Breed name assigned by the sampling				
			Borgou	Dahomey	Lagune	Pabli	Somba
Lagune_current	LA_cur	110			110		
Lagune_outgroup	LA_out	25			25		
Dahomey_ current	DAH_cur	30		30			
Somba_ current	SO_cur	57					57
Borgou_ current	BO_cur	135	135			2[a]	
Undescribed_admix	Adm_cur	98	42		2[a]	56	1[a]
Total		455	177	30	137	58	58

[a]*Five animals that did not show clear adherence to the main groups were considered as outgroups and excluded from the generated genetic clusters.*

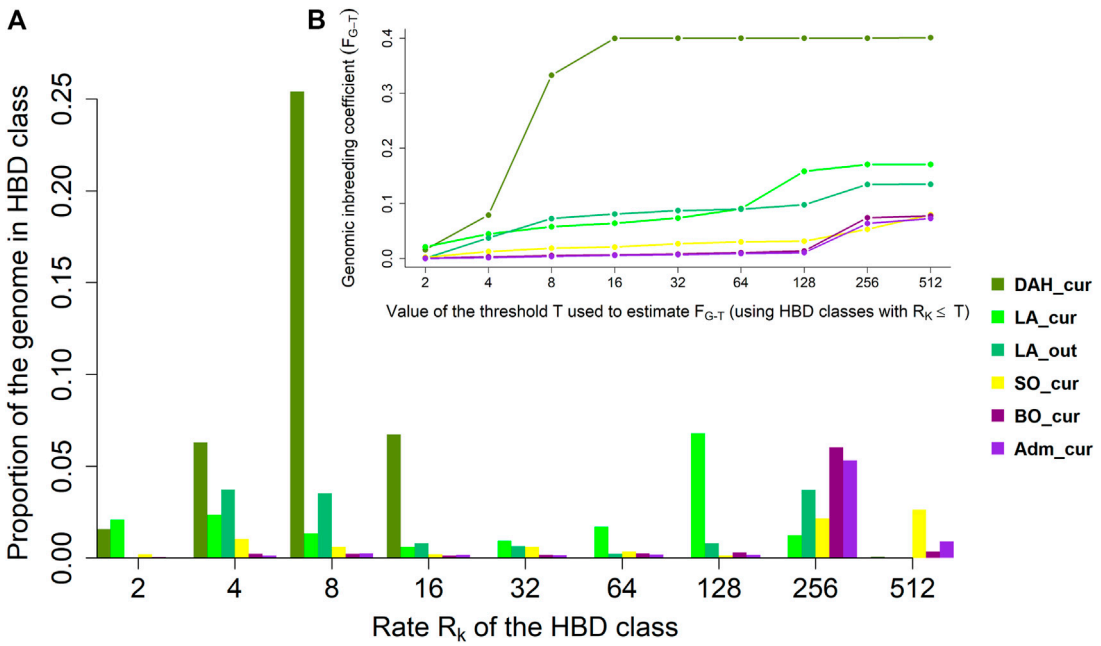

FIGURE 3 | (A) Average realized autozygosity and **(B)** cumulative genomic inbreeding associated with different age-related homozygous-by-descent (HBD) classes in Beninese cattle populations. Dah_cur = current Dahomey; LA_cur = current pure Lagune; LA_out = admixed Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou.

classes with R_k = 256 (27% of the total HBD proportion) and R_k = 512 (33% of the total HBD proportion). The contribution of recent classes to autozygosity in SO_cur was lower (≤13% of the HBD loci). The estimated genomic inbreeding coefficients were 0.08 for BO_cur and 0.07 for Adm_cur. The HBD class with R_k = 256 was the main source of autozygosity, contributing to 77.94 and 73.09% of the total HBD proportions in BO_cur and Adm_cur, respectively.

Comparison of Current Beninese Cattle Population With Historical Data

Principal Component Analyses and Clustering

The first and second principal components (PC) from PCA presented a clear separation of the Beninese breeds (**Figure 4A**). Within each breed, recent and historical samples were distinctly displayed along the second PC. The Dahomey cattle (DAH_cur) were projected next to LA_cur on PC1 but have a major contribution to PC2. Similar differentiations between current and historical samples were observed with the unsupervised k-means clustering, especially in Borgou and Lagune (BIC values suggested seven clusters, see **Supplementary Figure S2A**). Indeed, the unsupervised k-means clustering assigned the animals from Adm_cur, BO_cur, and BO_hist in three different clusters (clusters 2, 6, and 7), respectively (**Supplementary Table S4A**). Similarly, the current Lagune samples (LA_cur and LA_out) were grouped into cluster 4 and historical samples (LA_hist) into cluster 1. DAH_cur formed its own cluster (cluster 4). However, the current and historical Somba samples (SO_cur and SO_hist) were jointly grouped into cluster 5.

We repeated the diversity analyses on the extended dataset including additional AFT, EUT, AFZ, Asian indicine (AID), and

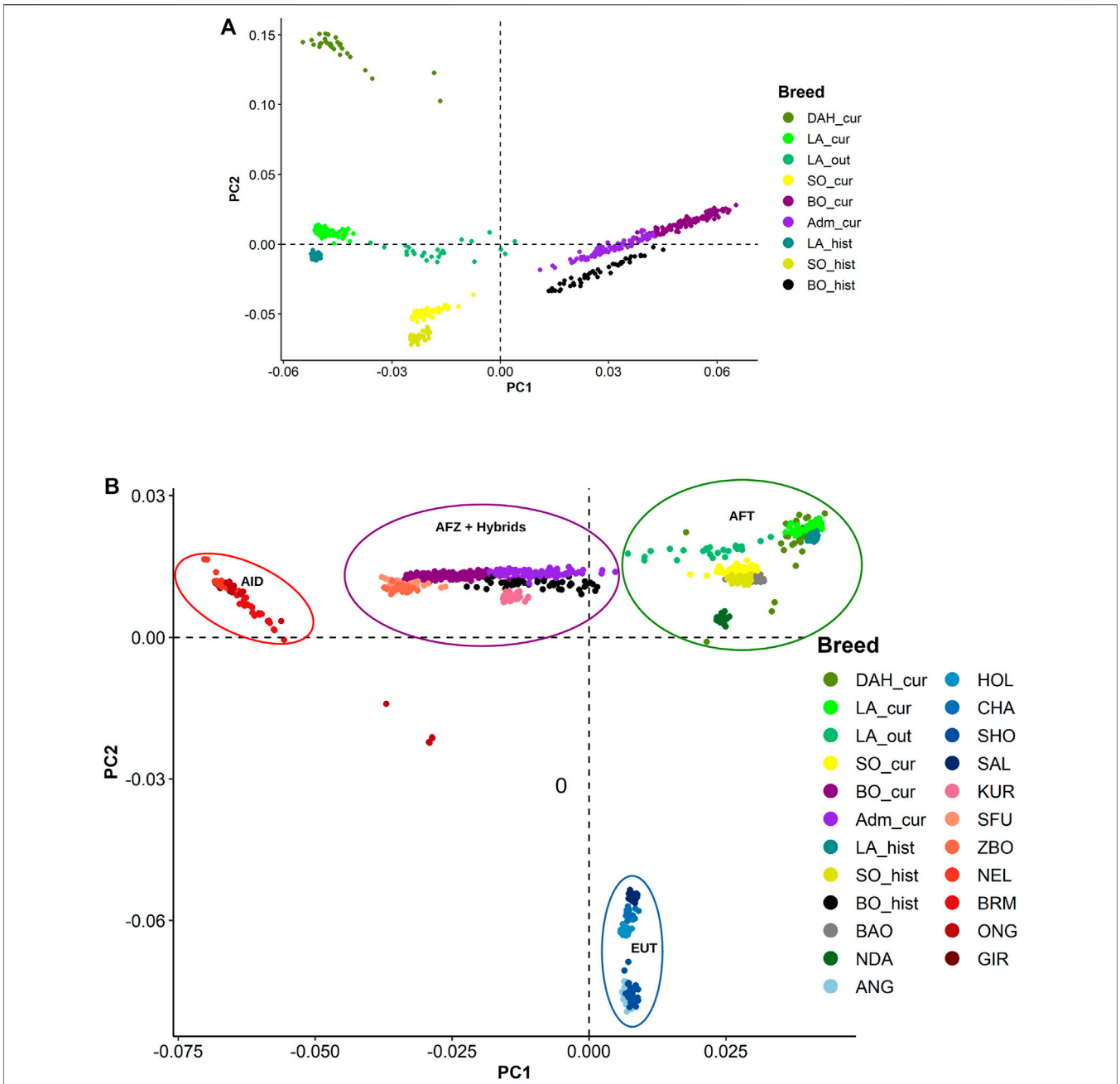

FIGURE 4 | Principal components (PC) displaying genomic divergence: **(A)** in the different Beninese cattle populations; **(B)** in the Beninese cattle with additional African and European taurine (EUT) breeds, African and Asian indicine breeds, and African crossbreeds. Dah_cur = current Dahomey; LA_cur = current pure Lagune; LA_out = admixed Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou; LA_hist = historical Lagune; SO_hist = historical Somba; BO_hist = historical Borgou; ANG = Angus; HOL = Holstein; CHA = Charolais; SHO = Shorthorn; BAO = Baoulé; NDA = N'Dama; GIR = Gir; SFU = Sudanese Fulani; ZBO = Zebu Bororo; BRM = Brahman; ONG = Ongole; NEL = Nellore; KUR = Kuri; SAL = Salers.

hybrids. Considering the PCA, the taurine animals (AFT and EUT) were separated from indicine (AFZ and AID) and hybrids samples along the first component, whereas the second component (PC2) showed a separation between the AFT and EUT (**Figure 4B**). Interestingly, the Dahomey was aggregated with the Lagune far away from EUT. The AID (Gir, Brahman, Ongole, and Nellore) also were clearly separated from the AFZ (ZFU and ZBO) and hybrids. Eight clusters (low BIC at k = 8; **Supplementary Figure S2B**) were identified from the unsupervised k-means clustering. Current and historical Somba samples (SO_cur and SO_hist) formed one genetic group (cluster 2) with the other AFT (BAO and NDA). The

TABLE 2 | f_3 and f_4 statistics for formal test of admixture in Beninese cattle populations.

Target[a]	f_3-ratio	z-score	Alpha[b]
DAH_cur	0.376	26.927	0.974
LA_cur	0.111	35.612	1.000
LA_out	−0.021	−9.677	0.872
SO_cur	0.005	2.492	0.963
BO_cur	−0.083	−39.083	0.474
Adm_cur	−0.048	−28.460	0.615
LA_hist	0.110	35.501	1.008
SO_hist	−0.006	−4.542	0.972
BO_hist	−0.098	−55.732	0.626

[a]*Dah_cur = current Dahomey; LA_cur = current pure Lagune; LA_out = admixed Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou; LA_hist = historical Lagune; SO_hist = historical Somba; BO_hist = historical Borgou.*

[b]*Alpha values represent the estimates of the proportion of AFT in the different populations.*

Lagune samples (LA_cur, LA_out, and LA_hist) were grouped into cluster 2. Similarly, DAH_cur is displayed in one cluster (cluster 1). Moreover, the hybrids (Adm_cur, BO_cur, and BO_hist) were grouped with Kuri, ZFU, and ZBO in cluster 6. The other genetic groups consisted of the EUT and AID (**Supplementary Table S4B**).

Formal Test of Admixture and Inference of Ancestral Proportion

The formal test of admixture of three populations resulted in positive f_3-ratios for DAH_cur, LA_cur, and SO_cur (**Table 2**). In contrast, we obtained negative statistics for LA_out, BO_cur and Adm_cur with the most significantly negative f_3 value for BO_cur ($f_3 = -0.08$; $Z = -39.08$). Considering the historical data, the f_3-ratio test resulted in positive values for LA_hist but was negative for SO_hist and BO_hist.

The estimation of the ancestral AFT with the f_4-ratio test revealed high proportions (alpha values superior to 0.97) in DAH_cur, LA_cur, and SO_cur, respectively (**Table 2**). In contrast, lower alpha values were found for LA_out, Adm_cur, and BO_cur. In comparison to the historical samples, we observed slight reductions of AFT ancestral proportions in Somba (0.96 for SO_cur against 0.97 for SO_hist) and in Adm_cur (0.62 against 0.63 for BO_hist). The current Borgou (BO_cur) displayed an important reduction of AFT ancestral proportions (alpha = 0.47) in comparison to historical samples (BO_hist). For LA_cur and LA_hist, the estimated AFT ancestral proportions were equal to 1.00, respectively.

Selection Signatures

Dahomey and Lagune

We detected no significant SNP presenting strong extended homozygosity in DAH_cur and LA_out relative to LA_hist. In contrast, the current Lagune population (LA_cur) displayed 19 significant SNP for positive selection (**Figure 5A** and **Table 3**). Among the SNP, 12 were positioned in a total of three candidate regions on BTA1, 18, and 21. The region 15.5–16 Mb on BTA21 contained the largest number of significant SNP (six SNP). The region 47.5–48 Mb on BTA18 included the most significant SNP rs110495745 ($p\_adjust = 8.40 \times 10^{-04}$), which is positioned within the *WDR87* gene. In total, 9 candidate genes were annotated within the three regions. GO enrichment analysis identified functional enriched terms such as cellular processes (**Table 4**). QTL associated with reproduction (67%) and conformation (67%) were predominant in the candidate regions (**Figure 6**).

Somba

The analysis of positive selection signatures in SO_cur relative to SO_hist detected 35 significant SNP (**Figure 5B**). Five candidate regions under recent selection in SO_cur were identified on BTA6, 7, 9, and 23, harboring in total 20 significant SNP. The region 4–4.5 Mb on BTA6 displayed the majority of the significant SNP (6 SNP) and the most significant SNP (rs42405104, $p\_adjust = 2.34 \times 10^{-07}$). This SNP is not positioned in any gene. However, the second most significant SNP ($p\_adjust = 3.57 \times 10^{-04}$) is positioned in the vicinity of three genes (*BOLA-DMB*, *BOLA-DMA* and *BRD2*) on BTA 23. In total, 33 genes were mapped in the candidate regions. GO enrichment analysis of the genes revealed highly significant GO biological processes such as antigen processing and adaptive immune response (**Table 4**). We observed that the majority of the candidate regions under positive selection in SO_cur overlapped with QTL affecting heath traits (80%) and carcass quality (60%, **Figure 6**).

Crossbreeds

We detected several genomic regions under recent selection in BO_cur and Adm_cur relative to BO_hist (**Figures 5C,D**). A total of 77 SNP displayed strong homozygosity in BO_cur, with adjusted *p_values* for XP-EHH below the significance threshold ($p\_adjust \leq 0.05$). Among these SNP, 44 were located in eight candidate regions on BTA2, 3, 5, 16, and 27 (**Table 3**). The regions that spanned from 70 to 71 Mb on BTA2 and from 54.5 to 55.5 Mb on BTA5 contained the largest number of significant SNP (9 and 8 significant SNP, respectively). The later region hosted the highest peak with the SNP rs41637710 ($p\_adjust = 0.003$). This SNP was not in the vicinity of any gene. However, three significant SNP (rs110903828 on BTA27, rs41637109 on BTA16, and rs43361717 on BTA3) were positioned within the genes *HTRA4*, *DNAJC11*, and *TMEM53*, respectively. A total of 49 genes were mapped within the genomic candidate regions. GO enrichment analysis revealed several significant GO biological processes including cytoplasmic microtubule organization and inner mitochondrial membrane organization (**Table 4**). Moreover, several QTL affecting animal performances overlapped with the regions under positive selection in BO_cur. QTL associated with milk (75%), carcass (63%), and reproduction traits (63%) were the most represented (**Figure 6**).

In Adm_cur, we identified 69 significant SNP under positive selection. A subset of 42 SNP was located in six candidate regions on BTA3, 4, 5, and 13. Three candidate regions in Adm_cur overlapped with those detected in BO_cur (**Table 3**). Similarly, the SNP rs41637710 located on BTA5 (54.5–55.5 Mb) showed the lowest adjusted *p*-value ($p\_adjust = 3.38 \times 10^{-07}$). The segment on BTA3 (101–102 Mb) hosted the largest number of extreme SNP (13 SNP). The candidate regions in Adm_cur included 63 genes.

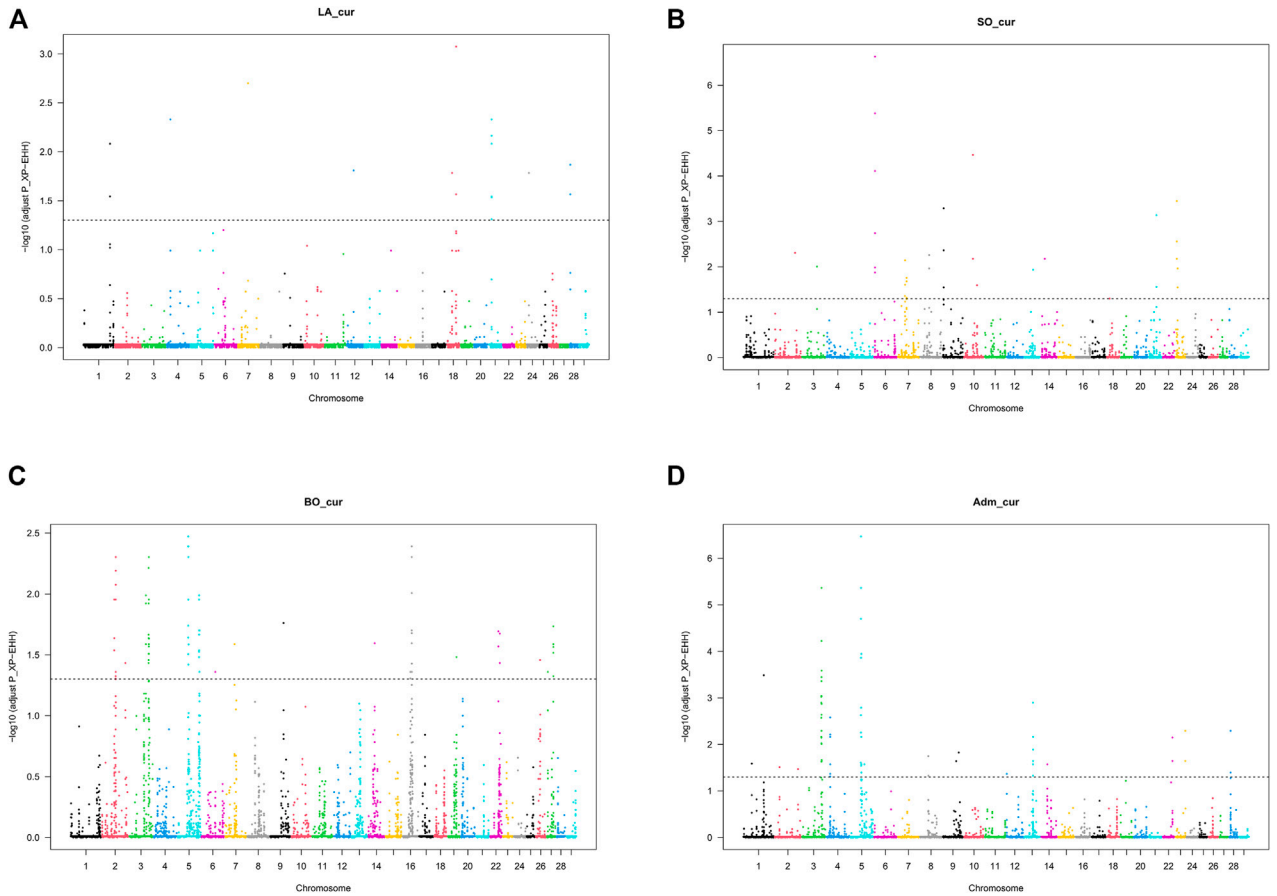

FIGURE 5 | Manhattan plot for cross-population extended haplotype homozygosity (XP-EHH) test for selection signature in Beninese cattle populations: **(A)** LA_cur, current Lagune pure population, **(B)** SO_cur, current Somba population, **(C)** BO_cur, current Borgou highly admixed, and **(D)** Adm_cur, moderately admixed hybrids.

The genes *DNAJC11*, *ENTHD1*, and *MAST2* hosted significant SNP, whereas others (*IL23A*, *PAN2*, and *CNPY2*) were mapped at a close distance to the significant SNP. GO enriched terms included biological processes such as pigment biosynthetic process and response to chemicals (**Table 4**). Like in BO_cur, the majority of QTL, located in the candidate regions of positive selection in Adm_cur, were related to milk (100%), carcass (67%), and reproduction traits (67%, **Figure 6**).

DISCUSSION

Population Structure and Admixture Tests

The results of the population structure analyses are in line with the breed foundation of the different cattle populations in Benin and their divergence from other African and European cattle breeds (Mwai et al., 2015). First, we observed high proximity of the Somba and Baoulé populations as Savannah Shorthorns and their separation to the Dwarf (forest) Lagune Shorthorn (Rege et al., 1994). In addition, the divergence of the Shorthorn breeds from the Longhorn N'Dama illustrates the rich genetic diversity of West African indigenous cattle breeds and the necessity to unravel specific signatures in each population. Second, our study confirms the Lagune origin of the Dahomey cattle. We found neither indicine nor EUT background in the Dahomey population. The genetic purity of the Dahomey cattle is probably due to their promotion in a close production system organized by the Dahomey-Zwergrind breeder association aiming at the conservation of the breed (http://www.dahomey-zwergrind.com). The inference of the oldest common ancestor dated from eight generations ago (approximately 24 years ago). The high genomic inbreeding coefficient in $R_k = 8$ (4 generations equal to 16 years ago) in the Dahomey population may be related to the recent creation of the association in 2001. These results suggest that the founders of the current Dahomey population kept by farmers may have originated from a small number of Dahomey cattle four generations ago (http://www.dahomey-zwergrind.com). However, the low frequency of short ROH segments, resulting in low genomic inbreeding coefficients (<0.1) in very young classes ($R_k \leq 4$), indicates a reduction of mating between related individuals in recent generations (Druet and Gautier, 2017). Exchange of breeding animals between the association members and consideration of Dahomey cattle currently held in different zoos across Germany and other

TABLE 3 | Candidate regions harboring positive recent selection signatures in Beninese cattle populations.

Population[a]	BTA	Region (Mb)		Number of sign. SNP	Adjust p-value		Genes[b]
		Start	End		Min	Max	
LA_cur	1	137.5	138	3	0.008	0.029	
	18	47.5	48	3	0.001	0.027	***WDR87***, *ZNF345, ZFP30*
	21	15.5	16	6	0.005	0.049	*SV2B*
SO_cur	6	4	4.5	6	0.000	0.013	QRFPR
	7	45	45.5	5	0.018	0.049	*FSTL4*
	9	4	4.5	3	0.001	0.028	
	23	7	7.5	3	0.000	0.007	***BOLA-DMA***, ***BOLA-DMB***, ***BRD2***, *PSMB8*
	23	11.5	12	3	0.011	0.028	*MDGA1, ZFAND3*
BO_cur	2	70	71	9	0.005	0.050	EN1, MARCO
	3	86	86.5	3	0.010	0.026	HOOK1
	3	100	100.5	5	0.006	0.035	**MAST2**, RAD54L, POMGNT1
	3	101	101.5	4	0.005	0.027	HECTD3, KIF2C
	5	54.5	55.5	8	0.003	0.031	*LRIG3*
	5	111	111.5	6	0.010	0.033	GRAP2, **ENTHD1**, FAM83F
	16	46.5	47	6	0.004	0.025	***DNAJC11***, RNF207, *PLEKHG5, THAP3*
	27	34	34.5	3	0.026	0.047	*PLEKHA2, ADAM32,*
Adm_cur	3	100	100.5	6	0.001	0.007	*MAST2, RAD54L, POMGNT1*
	3	101	102	13	0.000	0.023	HECTD3, **RNF220**, **IPP**
	4	20	21	6	0.003	0.049	TMEM106B, SCIN, ARL4A
	5	54.5	55.5	8	0.000	0.025	LRIG3
	5	57	57.5	6	0.000	0.040	ANKRD52, SLC39A5, RNF41, DNAJC14
	13	45.5	46	3	0.001	0.047	—

[a]*LA_cur = current pure Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou.*
[b]*Genes harboring the core SNP are displayed in bold. The complete list of the genes located in the candidate regions are presented in* ***Supplementary Table S5****.*

European countries (Zootierliste, 2020: https://www.zootierliste.de) may contribute to control inbreeding and to increase genetic diversity in the population currently managed by the Dahomey–Zwergrind breeder association.

The results of the PCA analyses differentiated historical and current populations as well as populations affected by admixture. The identification of admixed animals (LA_out) from relatively pure Lagune is confirmed by the formal admixture test, the three-population test, and the estimation of admixture proportions. Our findings are in accordance with the increasing crossbreeding in Lagune cattle due to the extension of transhumance as reported by Scheper et al. (2020) and Ahozonlin and Dossa (2020). The large genomic inbreeding in the Lagune (LA_cur) is in line with the small number of the original populations formerly distributed in clusters across West African coastal and forest regions (Rege, 1999). Fortunately, the estimation of high AFT ancestral proportion in LA_cur suggests the existence of a relatively pure population, which may be valuable for the conservation of this indigenous taurine breed. The estimated alpha value of 1.00 may be related to the reference population considered. However, the Baoulé (BAO) is the closest shorthorn taurine with available historical genotype data, whereas the GIR is the purest indicine reference population as the majority of AFZ are admixed (Flori et al., 2014).

We found that Somba cattle are less affected than Lagune by Zebu introgression. Previous studies observed that its habitat in the hilly region of Atacora protected from Zebu introgression (Rege et al., 1994; Hall et al., 1995). In comparison to the location of the other local breeds (Northeastern and Southern Benin), the lower pressure of transhumance in Boukombe and lower economic and demographic pressures, resulting in less "modernization" of cattle management and crossbreeding, are some advantages (Houessou et al., 2019a; Scheper et al., 2020). However, the lower AFT ancestral proportion in SO_cur compared to SO_hist confirms the threat of admixture in Somba cattle mainly caused by entrustment practices (Hall et al., 1995; Dossa and Vanvanhossou, 2016; Vanvanhossou et al., 2021). In addition, the negative and positive f-statistics in the historical (SO_hist) and the current (SO_cur) Somba populations, respectively, indicate former introgression episodes followed by genetic drift (Patterson et al., 2012; Kim et al., 2020). We also observed that Somba cattle remain less inbred despite the reduction of population size and their shrinkage into the unique location of Boukombe (Dossa and Vanvanhossou, 2016).

The results from fastStructure corroborate the Somba background of the Beninese crossbreeds. In addition, the hybrid populations (BO_cur and Adm_cur) presented a relatively low genetic proximity to the AID (Gir, Brahman, Ongole, and Nellore), but they were clustered close to the AFZ (ZFU and ZBO). These results are in agreement with their origin as described by different authors (Atchy, 1976; Flori et al., 2014). The identification of different levels of admixture in the hybrid samples was confirmed by the f_3 and f_4 admixture tests. The lower AFT ancestral proportion in BO_cur compared to the historical population BO_hist confirms the increasing introgression of AFZ in smallholder Borgou herds as indicated by Scheper et al. (2020). BO_cur representing more than 75% of the current Borgou samples suggests an intensive admixture and a high risk of full replacement of the Borgou population by AFZ genotypes. We observed similar AFT ancestral proportions in the historical Borgou (Bo_hist) and Adm_cur. The AFT ancestral

TABLE 4 | Enriched gene ontology (GO) biological process for genes in candidate regions under positive selection in Beninese cattle populations.

Population[a]	GO biological process	p- value	N[b]	Genes
LA_cur	Cellular process (GO:0009987)	5.94E-03	2	*SV2B, ENSBTAG00000054913*
SO_cur	Antigen processing and presentation (GO:0019882)	5.44E-19	11	*BOLA-DOA, BOLA-DMA, TAP1, PSMB8, TAPBP, DSB, BOLA-DMB, BOLA-DYA, BOLA-DOB, BOLA-DMA, BOLA-DIB*
	Adaptive immune response (GO:0002250)	2.74E-11	9	*BOLA-DOA, BOLA-DMA, TAP1, DSB, BOLA-DMB, BOLA-DYA, BOLA-DOB, BOLA-DMA, LOC618733*
	MHC protein complex assembly (GO:0002396)	3.98E-05	2	*BOLA-DMA, TAPBP*
	Cellular response to steroid hormone stimulus (GO: 0071383)	6.61E-03	2	*RXRB, DAXX*
	Proteasomal ubiquitin-independent protein catabolic process (GO:0010499)	7.03E-04	2	*PSMB8, PSMB9*
	DNA conformation change (GO:0071103)	9.03E-03	3	*H2B, BRD2, DAXX*
BO_cur	GO:000700: inner mitochondrial membrane organization (GO:0048519)	2.42E-03	2	*TAZ, DNAJC11*
	Negative regulation of biological process (GO: 0048519)	5.20E-03	2	
	Cytoplasmic microtubule organization (GO: 0031122)	7.61E-03	2	*HOOK1, PLK3*
	Biological regulation (GO:0065007)	6.25E-03	16	*EIF2B3, KLHL21, ADAM9, KIF2C, TNFRSF25, MAST2, THAP3, PTCH2, ZBTB48, HES2, PLK3, TAS1R1, PLEKHG5, EN1, GPBP1L1, FAM83F*
Adm_cur	Pigment biosynthetic process	7.01E-03	2	*UROD, PMEL*
	ATP-dependent chromatin remodeling	7.90E-03	2	*DMAP1, SMARCC2*
	Response to chemical	9.33E-03	3	*SLC39A5, PLK3, ENSBTAG00000051912 (taste receptor type 2)*
	G1/S transition of mitotic cell cycle	9.50E-03	2	*CDK2, PLK3*

[a]*LA_cur = current pure Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou.*
[b]*Number of the identified genes.*

proportion in the later population, comprising the Pabli samples, indicates the existence of a residue of the Beninese indigenous crossbreeds. According to Pecaud (1912) (as reported by Atchy, 1976), the Pabli breed results from crossbreeding between Borgou and Somba around the year 1905. In addition, the region of Kerou hosting the Pabli cattle in Western Benin is also less affected by transhumance. The association between spatial indicine introgression and transhumance in the Beninese cattle population was described by Scheper et al. (2020). Finally, the existence of crossbreed populations with divergent admixture levels offers the opportunity to evaluate the impact of crossbreeding in terms of divergence in extended haplotype homozygosity profiles.

Selection Signatures

Previous studies (Lohmueller et al., 2010; Freedman et al., 2016) observed that admixture or further demographic events (population bottlenecks due to diseases) affect ancestral haplotypes and increase the occurrence of mosaics in the genome (Freedman et al., 2016; Aliloo et al., 2020). This may impede the distinction of genomic footprints left by neutral processes and natural selection. In consequence, we did not expect evidence for historical selections in the Beninese cattle population. Nevertheless, the selection is not dissociable from admixture in several African cattle breeds. Admixture is a historical practice in African cattle production and is considered as a quick means of animal upgrading (Flori et al., 2014). Researchers reported that the selection of the animal or the breed for crossbreeding is driven by farmer interests including desired productive (milk, meat, and reproductive) and adaptive features (left by natural selection) (Boutrais, 2007). In this context, the prevalence of specific genomic regions or functional traits within a population may reflect the production goals of the farmers. In addition, specific features in the genome of African cattle populations (including several populations with various crossbreeding histories) are commonly assessed with selection signature analyses (Gautier et al., 2009; Taye et al., 2018; Aliloo et al., 2020; Kim et al., 2020).

By contrasting the current and historical cattle populations in the Beninese taurine and crossbreeds, we focus on genomic footprints resulting from recent environmental pressures or herd management. Indeed, environmental and socioeconomic factors have induced diverse changes in cattle management practices in Benin. These include the adoption of cattle mobility in taurine or agropastoralist herds (former sedentary), the increase of herd mobility frequencies and amplitudes, the migration and settlement of several agropastoralists from Sahelian countries or Northern Benin into Southern Benin, and the extension of animal entrustment practices (from agropastoralists to traditional pastoralists) (Houessou et al., 2019a; Houessou et al., 2020; Vanvanhossou et al., 2021). Despite the limited period between our and the historical samples, we identified a few candidate regions, providing new insights into the evolutionary process in the indigenous breeds. In comparison to the taurine, we observed that the crossbreeds, especially BO_cur, displayed most of the candidate regions. These results are in line with increased Zebu introgression in BO_cur, leading to a higher admixture proportion than Adm_cur and more genetic divergence from the historical Borgou (BO_hist). Regarding the Somba (SO_cur), the identification of five candidate regions mainly associated with immunity features suggests the

importance of disease pressures in its belt, as extensively discussed below.

Our approach to detect a strong signal of homozygosity considering temporal subpopulations within the same breed is similar to the one applied by Naderi et al. (2020), who contrasted the current German Holstein to one of its recent ancestors, the local dual-purpose German black pied cattle (DSN). In addition, we aimed to reduce the bias due to the uncertainty of the ancestral base population and applied the XP-EHH method to detect complete sweeps in contrast to the Integrated Haplotype Score (IHS) approach (Qanbari and Simianer, 2014). Indeed, the IHS method usually applied to investigate within-breed selection signatures relies on ancestral allele frequency which is inconsistently defined in different studies. For instance, Utsunomiya et al. (2013) derived ancestral alleles from common founders of Bovidae species, namely, *Bos gaurus*, *Bos grunniens*, and *Bubalus bubalis*. Kim et al. (2020) considered fixed alleles in African Buffalo as ancestral alleles to determine selection signatures in the African population, while Gautier et al. (2009) estimated ancestral alleles for West African cattle breeds based on alleles frequencies in indicines and African and European taurine samples. Finally, we expect to reduce false-positive results due to multiple tests, by considering adjusted *p*-values to define significant SNP and by defining only regions with at least three significant SNP as candidate regions. The respective candidate regions identified for each cattle population are discussed in the ongoing sections.

Lagune

We were not able to detect any candidate regions for DAH_cur. The Dahomey cattle may have acquired very few complete selective sweeps that are not detectable by the XP-EHH. This result reflects the breeding strategy with only focus on leisure and is in line with its high genomic inbreeding coefficient (http://www.dahomey-zwergrind.com). Further investigations including a complete sequencing of the Dahomey cattle and other methods of selection signature analyses may help untangle the genetic divergence between the Dahomey and the Lagune.

The admixed LA_out population was not considered for selection signature analysis because of the divergence in introgression levels as shown in the population structure analyses and admixture test. The LA_out group is the product of diverse admixed animals sampled as Lagune and therefore cannot represent any specific cattle population in Benin. On the contrary, LA_cur, as a relatively pure Lagune population, displayed three candidate regions. The region 15.5–16 Mb on BTA21 including the most significant SNP in LA_cur is identical to selection signatures reported in Holstein and North African cattle (Taye et al., 2017b; Ben-Jemaa et al., 2020). The *SV2B* gene in this region is associated with feed intake in cattle (Seabury et al., 2017). In contrast to the other populations investigated in this study, the candidate regions under selection in LA_cur encompassed relatively few genes functionally described in the literature. Nevertheless, the *WDR87* gene, hosting the most significant SNP, is related to carcass traits in cattle (Lim et al., 2013). In addition, the majority of the genes are involved in cellular and regulation of RNA biosynthetic processes, which are responsible for feed efficiency and body weight (Olivieri et al., 2016). Similarly, QTL associated with body weight, milk and reproductive traits are predominant in the candidate regions.

Somba

With regard to the five candidate regions under selection in SO_cur, two were detected on BTA23. This chromosome is

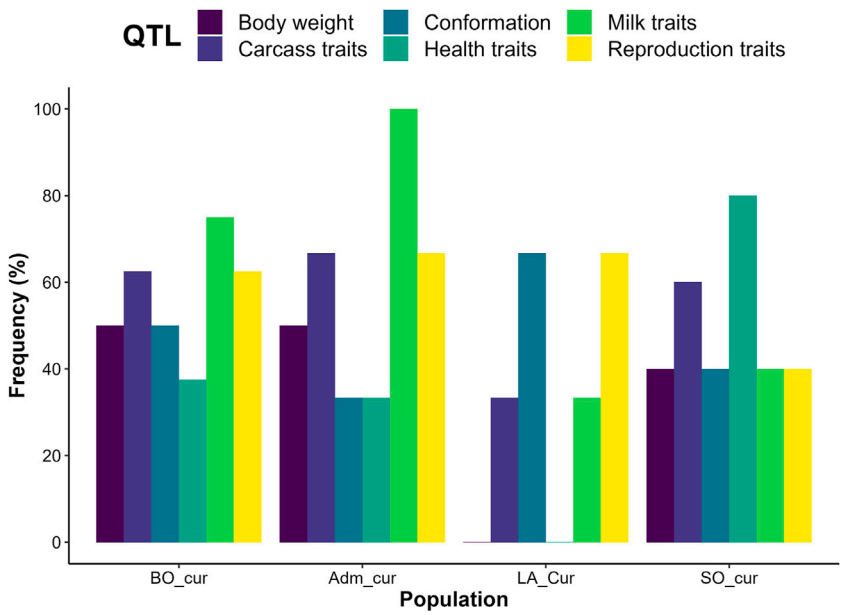

FIGURE 6 | Proportion of candidate regions with quantitative trait loci (QTL) controlling varied trait categories. LA_cur = current pure Lagune; SO_cur = current Somba; BO_cur = current Borgou highly admixed; Adm_cur = current moderately admixed animal including Pabli with some Borgou.

known to have two subregions (classes IIa and IIb) of the bovine leukocyte antigen (BoLA), also called bovine major histocompatibility complex (MHC), which is determinant in the development of acquired immune responses to diverse parasitic and viral diseases (Ellis and Ballingall, 1999; Takeshima and Aida, 2006). The selective region 7–7.5 Mb (BTA23) spanning several genes (e.g., *DSB*, *BOLA-DYA*, *BOLA-DMB*, and *BOLA-DOA*) overlaps with the subregion of BoLA class IIb (Takeshima and Aida, 2006). These BoLA class IIb genes are specific to ruminants but are less characterized, in contrast to genes in the BoLA class IIa (e.g., *BoLA-DRB3* and *BoLA-DQA3*) (Takeshima and Aida, 2006). In addition, Ballingall and McKeever (2005) associated the rare polymorphism of BoLA class IIb genes to evolutionary processes under functional constraints. Selection signatures in this genomic region were only found in Angus and Brangus cattle (Goszczynski et al., 2018; Maiorano et al., 2018), whereas selection for adaptive immunity in African cattle breeds is usually identified within the BoLA class IIa region (Ballingall et al., 1997; Kim et al., 2017; Tijjani et al., 2019). Therefore, recent selection signatures in the specific BoLA class IIb genes may indicate a possible adaptation of the Somba cattle to endemic diseases, especially to anthrax. Indeed, the Somba cattle are affected by several recent episodes of anthrax outbreaks especially in the years 2007, 2009, 2012, and 2013 (Dossa and Vanvanhossou, 2016). Further studies are required to investigate the association between BoLA subregion IIb and resistance to disease in Somba and other African cattle breeds. In addition, other candidate genes include *PSMB9* and *HSD17B8* which are involved in meat and growth traits (Lee et al., 2012; Ma et al., 2015), *FSTL4* associated with milk production (Sanchez et al., 2019) and *ZFAND3* responsible for reproduction (Mohammadi et al., 2020). The predominance of candidate genes associated with immune response is confirmed by enriched terms such as antigen processing and presentation, and adaptive immune response. However, other biological processes include the regulation of cellular metabolic processes and intracellular protein transport mechanisms. The identified bovine QTL suggests selection on body weight, carcass, reproduction, and milk traits in the Somba cattle.

Crossbreeds

The regions of selection (54.5–55.5 Mb on BTA5, 100–100.5 Mb, and 101–101.5 Mb on BTA3) overlapping in the two hybrid populations are in line with their common indicine background. The significant SNP mapped in the region from 100 to 101.5 MB on BTA3 were positioned in various genes including the *RNF220* gene. This gene has been previously identified under selection in West African cattle (Gautier et al., 2009). In addition, the region includes genomic footprints of signatures in South African and East African Shorthorn hybrids with indicine ancestry deficiency in the later breed (Bahbahani et al., 2015; Zwane et al., 2019). These findings suggest that the region may represent an ancient and stable footprint of selection in indigenous African hybrids. The *RNF220* gene is involved in calving performance and milk yield (Abdel-Shafy et al., 2020; Purfield et al., 2020). In addition, the HECTD3 in this region is associated with cell cycle regulation and fat deposition, while PLK3 is related to gain and feed intake in cattle (Yu et al., 2009; Bahbahani et al., 2015; Zarek et al., 2017). We also detected in this region the PTCH2 and *SLC6A9* which are involved in reproduction and Porphyria disease, respectively (Nezamzadeh et al., 2005; Basavaraja et al., 2021). The region 55–55.5 Mb on BTA5 previously displayed evidence of selective sweeps in the EUT Charolais as well as in a tropical crossbreed between Charolais and Zebu, namely, the Canchim (Xu et al., 2015; Urbinati et al., 2016; Naval-Sánchez et al., 2020). Moreover, the *LRIG3* gene in this region is associated with body length in cattle and litter size in pigs (Xu et al., 2015; Metodiev et al., 2018).

Specific selection signatures detected in BO_cur include the selective sweep 46.5–47.5 Mb on BTA16. This region is of great interest. It presents several significant SNP and overlaps with genomic footprints detected in East African Zebu cattle as well as in a subpopulation of the German dual-purpose black and white cattle (Taye et al., 2018; Naderi et al., 2020). The region contains *DNAJC11*, a heat shock protein gene, involved in response to heat stress (Li et al., 2015). In addition, the candidate regions of selection (BTA 5:111.5–111.5 Mb and BTA 2:70–71 Mb) are known candidate regions under selection in different African, European and Asian cattle and sheep breeds (Hudson et al., 2014; Bomba et al., 2015; Wang et al., 2015; Bertolini et al., 2020). The region 111.5–111.5 Mb on BTA5 contributed to positive selection for natural virus resistance and to extensive admixture in West Sahelian African human populations (Cagliani et al., 2011; Pérez-Rivas et al., 2014; Triska et al., 2015). In addition, the segment BTA 2:70–71 Mb covered various candidate genes such as *EN1*, involved in growth traits in cattle (Buroker, 2014). Further regions under selection in BO_cur host candidate genes, significantly associated with different traits, including *ADAM32*, *ADAM9*, *HTRA4*, and *KLHL21* with residual feed intake and immune responses, *PLEKHA2* and *TNFRSF25* with growth and carcass performances, *HTRA4* with milk, and *HOOK1* with heat stress (Fan et al., 2015; Tizioto et al., 2015; Seong et al., 2016; Blanco et al., 2017; Hardie et al., 2017; Hay and Roberts, 2018; Sengar et al., 2018; Skibiel et al., 2018; Oliveira et al., 2019; Braz et al., 2020; Brunes et al., 2021; Soares et al., 2021).

Regarding the selection signatures in Adm_cur, the region 20–21 Mb on BTA4 in Adm_cur is consistent with the region reported by Naderi et al. (2020), who identified selection signatures in close distance to the *TMEM106B* gene in the German black pied cattle. In addition, the *ARL4A* gene was reported in the context of selective sweeps in Australian Holstein (Larkin et al., 2012) and with regard to copy number variations in African Nguni cattle and Polish Holstein (Wang et al., 2015; Mielczarek et al., 2017). The latter gene is associated with milk production in dairy (Raschia et al., 2018; Khan et al., 2020). The *SCIN* gene, also mapped in the region, is involved in residual feed intake in cattle (Salleh et al., 2017). The remaining regions under selection in Adm_cur spanned other candidate genes including *ANKRD52*, *RNF41*, and *MYL6* associated with height and carcass traits (Cai et al., 2019; Moravčíková et al., 2019; Feitosa et al., 2021), and *COQ10A* and *RNF41* related to milk trait and calf mortality (Lázaro et al., 2021; Marín-Garzón et al., 2021).

In addition, the *SARNP* gene is related to animal reproduction (Labrecque et al., 2014), while the *DNAJC14* is involved in heat stress (Bahbahani et al., 2015; Rehman et al., 2020).

Overall, the common and specific candidate regions identified in the hybrid populations confirm selection signatures in African and European crossbreeds (**Supplementary Table S6**). They cover several candidate genes related to economic and functional traits. The enriched biological processes including inner mitochondrial membrane organization and ATP-dependent chromatin remodeling are related to carcass traits, milk production, and reproduction in cattle (Lu et al., 2016; Zhang and Xie, 2019; Shi et al., 2021). We also observed that the hybrids present more candidate regions related to heat response than the taurine populations, which is in line with the admixture with Zebu cattle known to tolerate high environmental heat loads (Taye et al., 2017a; Kim et al., 2020). Moreover, few genes in candidate regions of selection are involved in immune response and feed efficiency. The evidence of selection for adaptive traits in the hybrid populations (including BO_cur that is highly admixed) may be related to the fact that the AFZ introduced in the production environments of the West African taurine for decades have also developed various adaptive features (Atchy, 1976; Houessou et al., 2019b). Consequently, their crossbreeding with indigenous taurine cattle reduces the risk of diluting adaptive traits in local breeds while offering the opportunity to increase animal performances. Our findings are in line with those reported in other African cattle breeds (Aliloo et al., 2020; Kim et al., 2020) and suggest the ability to develop robust and productive breeds via crossbreeding. Nevertheless, the improvement of cattle breeding in West Africa requires the establishment of sustainable crossbreeding programs and the enhancement of genomic selection including genotype by environment interactions in the indigenous breeds. These will be achieved through the determination of suitable breeds and optimal proportions of admixture, considering social-ecological constraints (Wollny, 2003; Wurzinger et al., 2014).

CONCLUSION

In this study, we confirm that the Dahomey cattle currently bred in Europe are a subpopulation of the Beninese indigenous Lagune breed. The high genomic inbreeding in the Dahomey population is due to its current breeding system. The introduction of new animals from zoological parks can increase the diversity of the Dahomey population. Moreover, the Beninese taurine indigenous Lagune and Somba cattle still conserve a high proportion of AFT ancestry, in comparison to the historical population. The Borgou displays a risk of full genetic replacement by African Zebu. However, we observed the existence of a hybrid population relatively less affected by ongoing indicine introgression and comparable to the historical Borgou.

We found no evidence of the negative impact of admixture on the adaptive features in the cattle populations including the crossbreeds, as they all present several genomic footprints involved in immune response, feed efficiency, and heat stress. Moreover, specific candidate regions in the Somba cattle demonstrate selection pressures related to endemic diseases in the habitat areas of the breed. Overall, identified recent selection in Beninese indigenous cattle towards productive traits such as reproduction, milk, and carcass traits favor the improvement of the economic merits of the breeds.

DATA AVAILABILITY STATEMENT

All the data supporting the results of this article are presented within the article or in the additional files. The raw genotypic data used in this study are openly accessible at http://dx.doi.org/10.22029/jlupub-73.

AUTHOR CONTRIBUTIONS

SV, CS, LD, and SK conceived the project. SU collected the data, designed and applied the statistical analyses, and drafted the manuscript. TY and CS supported in statistical analyses and manuscript preparation. SK, RF, and LD reviewed and edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by the VolkswagenStiftung, Hannover, Germany (grant Az 94 829). We acknowledge funding of SV through the German Academic Exchange Service (DAAD) scholarship program: Research Grants-Doctoral Programmes in Germany.

ACKNOWLEDGMENTS

The authors are thankful to the farmers in Benin and the Verband Europäisches Dahomey-Zwergrind for their valuable help and their collaboration during the collection of hair samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.657282/full#supplementary-material

REFERENCES

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., Ismael, A., El-Assal, S. E.-D., and Abou-Bakr, S. (2020). A Single-Step Genomic Evaluation for Milk Production in Egyptian Buffalo. *Livestock Sci.* 234, 103977. doi:10.1016/j.livsci.2020.103977

Ahozonlin, M. C., Dossa, L. H., Dahouda, M., and Gbangboche, A. B. (2020). Morphological Divergence in the West African Shorthorn Lagune Cattle Populations from Benin. *Trop. Anim. Health Prod.* 52, 803–814. doi:10.1007/s11250-019-02071-1

Ahozonlin, M. C., and Dossa, L. H. (2020). Diversity and Resilience to Socio-Ecological Changes of Smallholder Lagune Cattle Farming Systems of Benin. *Sustainability* 12, 7616. doi:10.3390/su12187616

Aliloo, H., Mrode, R., Okeyo, A. M., and Gibson, J. P. (2020). Ancestral Haplotype Mapping for GWAS and Detection of Signatures of Selection in Admixed Dairy Cattle of Kenya. *Front. Genet.* 11, 544. doi:10.3389/fgene.2020.00544

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Atchy, A. A. (1976). Contribution à l'étude de la transhumance en Republique populaire du Bénin. PhD Thesis. Senegal: Université de Dakar.

Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M. N., van Tassell, C., Sonstegard, T., et al. (2015). Signatures of Positive Selection in East African Shorthorn Zebu: A Genome-wide Single Nucleotide Polymorphism Analysis. *Sci. Rep.* 5, 11729. doi:10.1038/srep11729

Ballingall, K. T., Luyai, A., and McKeever, D. J. (1997). Analysis of Genetic Diversity at the DQA Loci in African Cattle: Evidence for a BoLA-DQA3 Locus. *Immunogenetics* 46, 237–244. doi:10.1007/s002510050268

Ballingall, K. T., and McKeever, D. J. (2005). Conservation of Promoter, Coding and Intronic Regions of the Non-classical MHC Class IIDYAgene Suggests Evolution under Functional Constraints. *Anim. Genet.* 36, 237–239. doi:10.1111/j.1365-2052.2005.01281.x

Basavaraja, R., Drum, J. N., Sapuleni, J., Bibi, L., Friedlander, G., Kumar, S., et al. (2021). Downregulated Luteolytic Pathways in the Transcriptome of Early Pregnancy Bovine Corpus Luteum Are Mimicked by Interferon-Tau *In Vitro*. *BMC Genomics* 22, 452. doi:10.1186/s12864-021-07747-3

Belemsaga, D., Lombo, Y., Thevenon, S., and Sylla, S. (2005). "Inventory Analysis of West African Cattle Breeds," in *Applications of Gene-Based Technologies for Improving Animal Production and Health in Developing Countries*. Editors H. P. Makkar and G. J. Viljoen (Berlin/Heidelberg: Springer-Verlag), 167–173.

Ben-Jemaa, S., Mastrangelo, S., Lee, S.-H., Lee, J. H., and Boussaha, M. (2020). Genome-Wide Scan for Selection Signatures Reveals Novel Insights into the Adaptive Capacity in Local North African Cattle. *Sci. Rep.* 10, 19466. doi:10.1038/s41598-020-76576-3

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Berthier, D., Peylhard, M., Dayo, G.-K., Flori, L., Sylla, S., Bolly, S., et al. (2015). A Comparison of Phenotypic Traits Related to Trypanotolerance in Five West African Cattle Breeds Highlights the Value of Shorthorn Taurine Breeds. *PLoS ONE* 10, e0126498. doi:10.1371/journal.pone.0126498

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M. T., Mastrangelo, S., Dall'Olio, S., et al. (2020). Comparative Selection Signature Analyses Identify Genomic Footprints in Reggiana Cattle, The Traditional Breed of the Parmigiano-Reggiano Cheese Production System. *Animal* 14, 921–932. doi:10.1017/S1751731119003318

Bertrand, A. R., Kadri, N. K., Flori, L., Gautier, M., and Druet, T. (2019). RZooRoH: An R Package to Characterize Individual Genomic Autozygosity and Identify Homozygous-by-Descent Segments. *Methods Ecol. Evol.* 10, 860–866. doi:10.1111/2041-210X.13167

Blanco, F. C., Soria, M. A., Klepp, L. I., and Bigi, F. (2017). ERAP1 and PDE8A Are Downregulated in Cattle Protected against Bovine Tuberculosis. *J. Mol. Microbiol. Biotechnol.* 27, 237–245. doi:10.1159/000479183

Bomba, L., Nicolazzi, E. L., Milanesi, M., Negrini, R., Mancini, G., Biscarini, F., et al. (2015). Relative Extended Haplotype Homozygosity Signals across Breeds Reveal Dairy and Beef Specific Signatures of Selection. *Genet. Sel. Evol.* 47, 25. doi:10.1186/s12711-015-0113-9

Boutrais, J. (2007). The Fulani and Cattle Breeds: Crossbreeding and Heritage Strategies. *Africa* 77, 18–36. doi:10.3366/afr.2007.77.1.18

Braz, C. U., Rowan, T. N., Schnabel, R. D., and Decker, J. E. (2020). Extensive Genome-wide Association Analyses Identify Genotype-By-Environment Interactions of Growth Traits in Simmental Cattle. *BioRxiv* 11, 13335. doi:10.1101/2020.01.09.900902

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015

Brunes, L. C., Baldi, F., Lopes, F. B., Lôbo, R. B., Espigolan, R., Costa, M. F. O., et al. (2021). Weighted Single-Step Genome-Wide Association Study and Pathway Analyses for Feed Efficiency Traits in Nellore Cattle. *J. Anim. Breed. Genet.* 138, 23–44. doi:10.1111/jbg.12496

Cagliani, R., Riva, S., Fumagalli, M., Biasin, M., Caputo, S. L., Mazzotta, F., et al. (2011). A Positively Selected APOBEC3H Haplotype Is Associated with Natural Resistance to HIV-1 Infection. *Evolution* 65, 3311–3322. doi:10.1111/j.1558-5646.2011.01368.x

Cai, Z., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2019). Weighting Sequence Variants Based on Their Annotation Increases the Power of Genome-Wide Association Studies in Dairy Cattle. *Genet. Sel. Evol.* 51, 20. doi:10.1186/s12711-019-0463-9

Dossa, L. H., and Vanvanhossou, F. U. S. (2016). The Indigenous Somba Cattle of the Hilly Atacora Region in North-West Benin: Threats and Opportunities for its Sustainable Use. *Trop. Anim. Health Prod.* 48, 349–359. doi:10.1007/s11250-015-0958-5

Druet, T., Oleński, K., Flori, L., Bertrand, A. R., Olech, W., Tokarska, M., et al. (2020). Genomic Footprints of Recovery in the European Bison. *J. Hered.* 111, 194–203. doi:10.1093/jhered/esaa002

Druet, T., and Gautier, M. (2017). A Model-Based Approach to Characterize Individual Inbreeding at Both Global and Local Genomic Scales. *Mol. Ecol.* 26, 5820–5841. doi:10.1111/mec.14324

E. Buroker, N. (2014). Computational STAT4 rSNP Analysis, Transcriptional Factor Binding Sites and Disease. *J. Bioinform. Diabetes* 1, 18–53. doi:10.14302/issn.2374-9431.jbd-15-890

Egito, A. A., Paiva, S. R., Albuquerque, M. D. S. M., Mariante, A. S., Almeida, L. D., Castro, S. R., et al. (2007). Microsatellite Based Genetic Diversity and Relationships Among Ten Creole and Commercial Cattle Breeds Raised in Brazil. *BMC Genet.* 8, 83. doi:10.1186/1471-2156-8-83

Ellis, S. A., and Ballingall, K. T. (1999). Cattle MHC: Evolution in Action? *Immunol. Rev.* 167, 159–168. doi:10.1111/j.1600-065X.1999.tb01389.x

Ensembl genome database (2020). Available at: http://www.ensembl.org/biomart/martview/ (Accessed November 04, 2020).

Epstein, H. (1971). *The Origin of the Domestic Animals of Africa*. New York: Africana Publishing Corporation.

Fan, H., Wu, Y., Zhou, X., Xia, J., Zhang, W., Song, Y., et al. (2015). Pathway-Based Genome-Wide Association Studies for Two Meat Production Traits in Simmental Cattle. *Sci. Rep.* 5, 18389. doi:10.1038/srep18389

Feitosa, F. L. B., Pereira, A. S. C., Mueller, L. F., Fonseca, P. A. d. S., Braz, C. U., Amorin, S., et al. (2021). Genome-wide Association Study for Beef Fatty Acid Profile Using Haplotypes in Nellore Cattle. *Livestock Sci.* 245, 104396. doi:10.1016/j.livsci.2021.104396

Flori, L., Thevenon, S., Dayo, G.-K., Senou, M., Sylla, S., Berthier, D., et al. (2014). Adaptive Admixture in the West African Bovine Hybrid Zone: Insight from the Borgou Population. *Mol. Ecol.* 23, 3241–3257. doi:10.1111/mec.12816

Freedman, A. H., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Davis, B. W., Gronau, I., et al. (2016). Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. *Plos Genet.* 12, e1005851. doi:10.1371/journal.pgen.1005851

Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., et al. (2017). Species Composition and Environmental Adaptation of Indigenous Chinese Cattle. *Sci. Rep.* 7, 16196. doi:10.1038/s41598-017-16438-7

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A Whole Genome Bayesian Scan for Adaptive Genetic Divergence in West African Cattle. *BMC Genomics* 10, 550. doi:10.1186/1471-2164-10-550

Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh2.0: A Reimplementation of the R Packagerehhto Detect Positive Selection from Haplotype Structure. *Mol. Ecol. Resour.* 17, 78–90. doi:10.1111/1755-0998.12634

Gautier, M., and Vitalis, R. (2012). Rehh: An R Package to Detect Footprints of Selection in Genome-Wide SNP Data from Haplotype Structure. *Bioinformatics* 28, 1176–1177. doi:10.1093/bioinformatics/bts115

Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 Years and Still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi:10.1093/nar/gky1055

Goszczynski, D. E., Corbi-Botto, C. M., Durand, H. M., Rogberg-Muñoz, A., Munilla, S., Peral-Garcia, P., et al. (2018). Evidence of Positive Selection Towards Zebuine Haplotypes in the BoLA Region of Brangus Cattle. *Animal* 12, 215–223. doi:10.1017/S1751731117001380

Hall, S. J. G., Gnaho, L. K., and Meghen, C. (1995). Une enquête sur la race bovine Somba au Bénin. *Rev. Elev. Med. Vet. Pays Trop.* 48, 77–83. doi:10.19182/remvt.9492

Han, H., McGivney, B. A., Farries, G., Katz, L. M., MacHugh, D. E., Randhawa, I. A. S., et al. (2020). Selection in Australian Thoroughbred Horses Acts on a Locus Associated with Early Two-Year Old Speed. *PLoS ONE* 15, e0227212. doi:10.1371/journal.pone.0227212

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2002). African Pastoralism: Genetic Imprints of Origins and Migrations. *Science* 296, 336–339. doi:10.1126/science.1069878

Hardie, L. C., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., Armentano, L. E., Wiggans, G. R., et al. (2017). The Genetic and Biological Basis of Feed Efficiency in Mid-Lactation Holstein Dairy Cows. *J. Dairy Sci.* 100, 9061–9075. doi:10.3168/jds.2017-12604

Hay, E. H., and Roberts, A. (2018). Genome-Wide Association Study for Carcass Traits in a Composite Beef Cattle Breed. *Livestock Sci.* 213, 35–43. doi:10.1016/j.livsci.2018.04.018

Ho, S. Y. W., Larson, G., Edwards, C. J., Heupink, T. H., Lakin, K. E., Holland, P. W. H., et al. (2008). Correlating Bayesian Date Estimates with Climatic Events and Domestication Using a Bovine Case Study. *Biol. Lett.* 4, 370–374. doi:10.1098/rsbl.2008.0073

Houessou, S. O., Dossa, L. H., Assogba, C. A., Diogo, R. V. C., Vanvanhossou, S. F. U., and Schlecht, E. (2020). The Role of Cross-Border Transhumance in Influencing Resident Herders' Cattle Husbandry Practices and Use of Genetic Resources. *Animal* 14, 2378–2386. doi:10.1017/S1751731120001378

Houessou, S. O., Dossa, L. H., Diogo, R. V. C., Ahozonlin, M. C., Dahouda, M., and Schlecht, E. (2019b). Confronting Pastoralists' Knowledge of Cattle Breeds Raised in the Extensive Production Systems of Benin with Multivariate Analyses of Morphological Traits. *PLoS ONE* 14, e0222756. doi:10.1371/journal.pone.0222756

Houessou, S. O., Dossa, L. H., Diogo, R. V. C., Houinato, M., Buerkert, A., and Schlecht, E. (2019a). Change and Continuity in Traditional Cattle Farming Systems of West African Coast Countries: A Case Study from Benin. *Agric. Syst.* 168, 112–122. doi:10.1016/j.agsy.2018.11.003

Hu, Z.-L., Park, C. A., and Reecy, J. M. (2019). Building a Livestock Genetic and Genomic Information Knowledgebase through Integrative Developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–D710. doi:10.1093/nar/gky1084

Hudson, N. J., Porto-Neto, L. R., Kijas, J., McWilliam, S., Taft, R. J., and Reverter, A. (2014). Information Compression Exploits Patterns of Genome Composition to Discriminate Populations and Highlight Regions of Evolutionary Interest. *BMC Bioinformatics* 15, 66. doi:10.1186/1471-2105-15-66

Jombart, T. (2008). Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart, T., and Ahmed, I. (2011). Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data. *Bioinformatics* 27, 3070–3071. doi:10.1093/bioinformatics/btr521

Khan, M. Z., Liu, L., Zhang, Z., Khan, A., Wang, D., Mi, S., et al. (2020). Folic Acid Supplementation Regulates Milk Production Variables, Metabolic Associated Genes and Pathways in Perinatal Holsteins. *J. Anim. Physiol. Anim. Nutr.* 104, 483–492. doi:10.1111/jpn.13313

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The Genome Landscape of Indigenous African Cattle. *Genome Biol.* 18, 34. doi:10.1186/s13059-017-1153-y

Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., et al. (2020). The Mosaic Genome of Indigenous African Cattle as a Unique Genetic Resource for African Pastoralism. *Nat. Genet.* 52, 1099–1110. doi:10.1038/s41588-020-0694-2

Labrecque, R., Vigneault, C., Blondin, P., and Sirard, M.-A. (2014). Gene Expression Analysis of Bovine Oocytes at Optimal Coasting Time Combined with GnRH Antagonist during the No-FSH Period. *Theriogenology* 81, 1092–1100. doi:10.1016/j.theriogenology.2014.01.037

Larkin, D. M., Daetwyler, H. D., Hernandez, A. G., Wright, C. L., Hetrick, L. A., Boucek, L., et al. (2012). Whole-Genome Resequencing of Two Elite Sires for the Detection of Haplotypes under Selection in Dairy Cattle. *Proc. Natl. Acad. Sci.* 109, 7693–7698. doi:10.1073/pnas.1114546109

Lázaro, S. F., Tonhati, H., Oliveira, H. R., Silva, A. A., Nascimento, A. V., Santos, D. J. A., et al. (2021). Genomic Studies of Milk-Related Traits in Water Buffalo (Bubalus Bubalis) Based on Single-Step Genomic Best Linear Unbiased Prediction and Random Regression Models. *J. Dairy Sci.* 104, 5768–5793. doi:10.3168/jds.2020-19534

Lee, S. H., van der Werf, J., Lee, S. H., Lim, D. J., Park, E. W., Gondro, C., et al. (2012). Genome Wide QTL Mapping to Identify Candidate Genes for Carcass Traits in Hanwoo (Korean Cattle). *Genes Genom* 34, 43–49. doi:10.1007/s13258-011-0081-6

Leutenegger, A.-L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., et al. (2003). Estimation of the Inbreeding Coefficient through Use of Genomic Data. *Am. J. Hum. Genet.* 73, 516–523. doi:10.1086/378207

Li, L., Sun, Y., Wu, J., Li, X., Luo, M., and Wang, G. (2015). The Global Effect of Heat on Gene Expression in Cultured Bovine Mammary Epithelial Cells. *Cell Stress and Chaperones* 20, 381–389. doi:10.1007/s12192-014-0559-7

Lim, D., Lee, S.-H., Kim, N.-K., Cho, Y.-M., Chai, H.-H., Seong, H.-H., et al. (2013). Gene Co-Expression Analysis to Characterize Genes Related to Marbling Trait in Hanwoo (Korean) Cattle. *Asian Australas. J. Anim. Sci.* 26, 19–29. doi:10.5713/ajas.2012.12375

Lohmueller, K. E., Bustamante, C. D., and Clark, A. G. (2010). The Effect of Recent Admixture on Inference of Ancient Human Population History. *Genetics* 185, 611–622. doi:10.1534/genetics.109.113761

Lu, J., Argov-Argaman, N., Anggrek, J., Boeren, S., van Hooijdonk, T., Vervoort, J., et al. (2016). The Protein and Lipid Composition of the Membrane of Milk Fat Globules Depends on Their Size. *J. Dairy Sci.* 99, 4726–4738. doi:10.3168/jds.2015-10375

Ma, Y., Chen, N., Li, F., Fu, W., Han, Y., Chang, Y., et al. (2015). Bovine HSD17B8 Gene and its Relationship with Growth and Meat Quality Traits. *Sci. Bull.* 60, 1617–1621. doi:10.1007/s11434-015-0882-0

MacHugh, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P., and Bradley, D. G. (1997). Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146, 1071–1086. doi:10.1093/genetics/146.3.1071

Maiorano, A. M., Lourenco, D. L., Tsuruta, S., Ospina, A. M. T., Stafuzza, N. B., Masuda, Y., et al. (2018). Assessing Genetic Architecture and Signatures of Selection of Dual Purpose Gir Cattle Populations Using Genomic Information. *PLoS ONE* 13, e0200694. doi:10.1371/journal.pone.0200694

Marín-Garzón, N. A., Magalhães, A. F. B., Schmidt, P. I., Serna, M., Fonseca, L. F. S., Salatta, B. M., et al. (2021). Genome-Wide Scan Reveals Genomic Regions and Candidate Genes Underlying Direct and Maternal Effects of Preweaning Calf Mortality in Nellore Cattle. *Genomics* 113, 1386–1395. doi:10.1016/j.ygeno.2021.02.021

Metodiev, S., Thekkoot, D. M., Young, J. M., Onteru, S., Rothschild, M. F., and Dekkers, J. C. M. (2018). A Whole-Genome Association Study for Litter Size and Litter Weight Traits in Pigs. *Livestock Sci.* 211, 87–97. doi:10.1016/j.livsci.2018.03.004

Mielczarek, M., Frąszczak, M., Giannico, R., Minozzi, G., Williams, J. L., Wojdak-Maksymiec, K., et al. (2017). Analysis of Copy Number Variations in Holstein-Friesian Cow Genomes Based on Whole-Genome Sequence Data. *J. Dairy Sci.* 100, 5515–5525. doi:10.3168/jds.2016-11987

Moazami Goudarzi, K., Belemsaga, D. M. A., Ceriotti, G., Laloë, D., Fagbohoum, F., Kouagou, N. T., et al. (2001). Caractérisation de la race bovine Somba à l'aide de marqueurs moléculaires. *Rev. Elev. Med. Vet. Pays Trop.* 54, 129. doi:10.19182/remvt.9791

Mohammadi, A., Alijani, S., Rafat, S. A., and Abdollahi-Arpanahi, R. (2020). Genome-Wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle. *Ann. Anim. Sci.* 20, 825–851. doi:10.2478/aoas-2020-0031

Moravčíková, N., Kasarda, R., Vostrý, L., Krupová, Z., Krupa, E., Lehocká, K., et al. (2019). Analysis of Selection Signatures in the Beef Cattle Genome. *Czech J. Anim. Sci.* 64, 491–503. doi:10.17221/226/2019-CJAS

Murray, C., Huerta-Sanchez, E., Casey, F., and Bradley, D. G. (2010). Cattle Demographic History Modelled from Autosomal Sequence Variation. *Phil. Trans. R. Soc. B* 365, 2531–2539. doi:10.1098/rstb.2010.0103

Mwai, O., Hanotte, O., Kwon, Y.-J., and Cho, S. (2015). Invited Review - African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian Australas. J. Anim. Sci.* 28, 911–921. doi:10.5713/ajas.15.0002R

Naderi, S., Moradi, M. H., Farhadian, M., Yin, T., Jaeger, M., Scheper, C., et al. (2020). Assessing Selection Signatures within and between Selected Lines of Dual-Purpose Black and White and German Holstein Cattle. *Anim. Genet.* 51, 391–408. doi:10.1111/age.12925

Naval-Sánchez, M., Porto-Neto, L. R., Cardoso, D. F., Hayes, B. J., Daetwyler, H. D., Kijas, J., et al. (2020). Selection Signatures in Tropical Cattle Are Enriched for Promoter and Coding Regions and Reveal Missense Mutations in the Damage Response Gene HELB. *Genet. Sel. Evol.* 52, 27. doi:10.1186/s12711-020-00546-6

Nezamzadeh, R., Seubert, A., Pohlenz, J., and Brenig, B. (2005). Identification of a Mutation in the Ovine Uroporphyrinogen Decarboxylase (UROD) Gene Associated with a Type of Porphyria. *Anim. Genet.* 36, 297–302. doi:10.1111/j.1365-2052.2005.01301.x

Oliveira, H. R., Lourenco, D. A. L., Masuda, Y., Misztal, I., Tsuruta, S., Jamrozik, J., et al. (2019). Single-Step Genome-Wide Association for Longitudinal Traits of Canadian Ayrshire, Holstein, and Jersey Dairy Cattle. *J. Dairy Sci.* 102, 9995–10011. doi:10.3168/jds.2019-16821

Olivieri, B. F., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Branco, R. H., Bonilha, S. F. M., de Albuquerque, L. G., et al. (2016). Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population. *PLoS ONE* 11, e0164390. doi:10.1371/journal.pone.0164390

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient Admixture in Human History. *Genetics* 192, 1065–1093. doi:10.1534/genetics.112.145037

Payne, W. J. A., and Hodges, J. (1997). *Tropical Cattle: Origins, Breeds and Breeding Policies*. Oxford, United Kingdom: Blackwell Science.

Pecaud, M. G. (1912). *L'élevage des Animaux Domestiques au Dahomey*; Imp. du Gvt. Général de l'A.O.F: Gorée, Senegal, 157p.

Pérez-Rivas, L. G., Jerez, J. M., Carmona, R., de Luque, V., Vicioso, L., Claros, M. G., et al. (2014). A microRNA Signature Associated with Early Recurrence in Breast Cancer. *PLoS ONE* 9, e91884. doi:10.1371/journal.pone.0091884

Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding*. Wallingford: CABI Publishing.

Porter, V. (1991). *Cattle. A Handbook to the Breeds of the World. Facts on File*. London: Christopher Helm.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Purfield, D. C., Evans, R. D., and Berry, D. P. (2020). Breed- and Trait-Specific Associations Define the Genetic Architecture of Calving Performance Traits in Cattle. *J. Anim. Sci.* 98, 1–18. doi:10.1093/jas/skaa151

Qanbari, S., and Simianer, H. (2014). Mapping Signatures of Positive Selection in the Genome of Livestock. *Livestock Sci.* 166, 133–143. doi:10.1016/j.livsci.2014.05.003

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* 197, 573–589. doi:10.1534/genetics.114.164350

Raschia, M. A., Nani, J. P., Maizon, D. O., Beribe, M. J., Amadio, A. F., and Poli, M. A. (2018). Single Nucleotide Polymorphisms in Candidate Genes Associated with Milk Yield in Argentinean Holstein and Holstein X Jersey Cows. *J. Anim. Sci. Technol.* 60, 31. doi:10.1186/s40781-018-0189-1

R. Blench and K. MacDonald (Editors) (1999). *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography* (London, United Kingdom: Routledge).

Rege, J. E., Aboagye, G. S., and Tawah, C. L. (1994). Shorthorn Cattle of West and Central Africa. I. Origin, Distribution, Classification and Population Statistics. *World Anim. Rev.* 78, 2–13.

Rege, J. E. O. (1999). The State of African Cattle Genetic Resources I. Classification Framework and Identification of Threatened and Extinct Breeds. *Anim. Genet. Resour. Inf.* 25, 1–25. doi:10.1017/S1014233900003448

Rehman, S. U., Nadeem, A., Javed, M., Hassan, F.-U., Luo, X., Khalid, R. B., et al. (2020). Genomic Identification, Evolution and Sequence Analysis of the Heat-Shock Protein Gene Family in Buffalo. *Genes* 11, 1388. doi:10.3390/genes11111388

Salleh, M. S., Mazzoni, G., Höglund, J. K., Olijhoek, D. W., Lund, P., Løvendahl, P., et al. (2017). RNA-Seq Transcriptomics and Pathway Analyses Reveal Potential Regulatory Genes and Molecular Mechanisms in High- and Low-Residual Feed Intake in Nordic Dairy Cattle. *BMC Genomics* 18, 258. doi:10.1186/s12864-017-3622-9

Sanchez, M.-P., Ramayo-Caldas, Y., Wolf, V., Laithier, C., El Jabri, M., Michenet, A., et al. (2019). Sequence-Based GWAS, Network and Pathway Analyses Reveal Genes Co-Associated with Milk Cheese-Making Properties and Milk Composition in Montbéliarde Cows. *Genet. Sel. Evol.* 51, 34. doi:10.1186/s12711-019-0473-7

Scheper, C., Bohlouli, M., Brügemann, K., Weimann, C., Vanvanhossou, S. F. U., König, S., et al. (2020). The Role of Agro-Ecological Factors and Transboundary Transhumance in Shaping the Genetic Diversity in Four Indigenous Cattle Populations of Benin. *J. Anim. Breed. Genet.* 137, 622–640. doi:10.1111/jbg.12495

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-Wide Association Study for Feed Efficiency and Growth Traits in U.S. Beef Cattle. *BMC Genomics* 18, 386. doi:10.1186/s12864-017-3754-y

Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE: A Web-Interfaced Next Generation Database for Genetic Diversity Exploration, with a First Application in Cattle. *BMC Genomics* 16, 940. doi:10.1186/s12864-015-2181-1

Sengar, G. S., Deb, R., Singh, U., Raja, T. V., Kant, R., Sajjanar, B., et al. (2018). Differential Expression of microRNAs Associated with thermal Stress in Frieswal (*Bos T* X *Bos I*) Crossbred Dairy Cattle. *Cell Stress and Chaperones* 23, 155–170. doi:10.1007/s12192-017-0833-6

Seong, J., Yoon, H., and Kong, H. S. (2016). Identification of microRNA and Target Gene Associated with Marbling Score in Korean Cattle (Hanwoo). *Genes Genom* 38, 529–538. doi:10.1007/s13258-016-0401-y

Shi, Y., Zhao, P., Dang, Y., Li, S., Luo, L., Hu, B., et al. (2021). Functional Roles of the Chromatin Remodeler SMARCA5 in Mouse and Bovine Preimplantation Embryos†. *Biol. Reprod.* 105, 359–370. doi:10.1093/biolre/ioab081

Skibiel, A. L., Peñagaricano, F., Amorín, R., Ahmed, B. M., Dahl, G. E., and Laporta, J. (2018). In Utero Heat Stress Alters the Offspring Epigenome. *Sci. Rep.* 8, 14609. doi:10.1038/s41598-018-32975-1

Soares, R. A. N., Vargas, G., Duffield, T., Schenkel, F., and Squires, E. J. (2021). Genome-Wide Association Study and Functional Analyses for Clinical and Subclinical Ketosis in Holstein Cattle. *J. Dairy Sci.* 104, 10076–10089. doi:10.3168/jds.2020-20101

Stock, F., and Gifford-Gonzalez, D. (2013). Genetics and African Cattle Domestication. *Afr. Archaeol. Rev.* 30, 51–72. doi:10.1007/s10437-013-9131-6

Takeshima, S.-N., and Aida, Y. (2006). Structure, Function and Disease Susceptibility of the Bovine Major Histocompatibility Complex. *Anim. Sci. J.* 77, 138–150. doi:10.1111/j.1740-0929.2006.00332.x

Taye, M., Lee, W., Caetano-Anolles, K., Dessie, T., Cho, S., Jong Oh, S., et al. (2018). Exploring the Genomes of East African Indicine Cattle Breeds Reveals Signature of Selection for Tropical Environmental Adaptation Traits. *Cogent Food Agric.* 4, 1552552. doi:10.1080/23311932.2018.1552552

Taye, M., Lee, W., Caetano-Anolles, K., Dessie, T., Hanotte, O., Mwai, O. A., et al. (2017a). Whole Genome Detection of Signature of Positive Selection in African Cattle Reveals Selection for Thermotolerance. *Anim. Sci. J.* 88, 1889–1901. doi:10.1111/asj.12851

Taye, M., Lee, W., Jeon, S., Yoon, J., Dessie, T., Hanotte, O., et al. (2017b). Exploring Evidence of Positive Selection Signatures in Cattle Breeds Selected for Different Traits. *Mamm. Genome* 28, 528–541. doi:10.1007/s00335-017-9715-6

Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. *Front. Genet.* 10, 442. doi:10.3389/fgene.2019.00442

Tizioto, P. C., Coutinho, L. L., Decker, J. E., Schnabel, R. D., Rosa, K. O., Oliveira, P. S., et al. (2015). Global Liver Gene Expression Differences in Nelore Steers with Divergent Residual Feed Intake Phenotypes. *BMC Genomics* 16, 242. doi:10.1186/s12864-015-1464-x

Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., and Pereira, L. (2015). Extensive Admixture and Selective Pressure across the Sahel Belt. *Genome Biol. Evol.* 7, 3484–3495. doi:10.1093/gbe/evv236

Urbinati, I., Stafuzza, N. B., Oliveira, M. T., Chud, T. C. S., Higa, R. H., Regitano, L. C. D. A., et al. (2016). Selection Signatures in Canchim Beef Cattle. *J. Anim. Sci Biotechnol* 7, 29. doi:10.1186/s40104-016-0089-5

Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., van Tassell, C. P., do Carmo, A. S., Mészáros, G., et al. (2013). Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. *PLoS ONE* 8, e64280. doi:10.1371/journal.pone.0064280

Vanvanhossou, S. F. U., Koura, I. B., and Dossa, L. H. (2021). The Implications of Herd Entrustment Practice for the Sustainable Use of Cattle Genetic Resources in the (Agro)-Pastoral Systems of West Africa: A Case Study from Benin. *Pastoralism* 11, 8. doi:10.1186/s13570-020-00189-8

Verband Europäisches Dahomey-Zwergrind (2019). Tiere. Available at: http://www.dahomey-zwergrind.com/ (Accessed November 21, 2020).

Wang, M. D., Dzama, K., Hefer, C. A., and Muchadeyi, F. C. (2015). Genomic Population Structure and Prevalence of Copy Number Variations in South African Nguni Cattle. *BMC Genomics* 16, 894. doi:10.1186/s12864-015-2122-z

Wollny, C. B. A. (2003). The Need to Conserve Farm Animal Genetic Resources in Africa: Should Policy Makers Be Concerned? *Ecol. Econ.* 45, 341–351. doi:10.1016/S0921-8009(03)00089-2

Wurzinger, M., Mirkena, T., and Sölkner, J. (2014). Animal Breeding Strategies in Africa: Current Issues and the Way Forward. *J. Anim. Breed. Genet.* 131, 327–328. doi:10.1111/jbg.12116

Xu, L., Bickhart, D. M., Cole, J. B., Schroeder, S. G., Song, J., Tassell, C. P. V., et al. (2015). Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. *Mol. Biol. Evol.* 32, 711–725. doi:10.1093/molbev/msu333

Yu, S.-L., Lee, S.-M., Kang, M.-J., Jeong, H.-J., Sang, B.-C., Jeon, J.-T., et al. (2009). Identification of Differentially Expressed Genes Between Preadipocytes and Adipocytes Using Affymetrix Bovine Genome Array. *J. Anim. Sci. Technol.* 51, 443–452. doi:10.5187/JAST.2009.51.6.443

Zarek, C. M., Lindholm-Perry, A. K., Kuehn, L. A., and Freetly, H. C. (2017). Differential Expression of Genes Related to Gain and Intake in the Liver of Beef Cattle. *BMC Res. Notes* 10, 1. doi:10.1186/s13104-016-2345-3

Zhang, X., and Xie, J. (2019). Analysis of Proteins Associated with Quality Deterioration of Grouper Fillets Based on TMT Quantitative Proteomics during Refrigerated Storage. *Molecules* 24, 2641. doi:10.3390/molecules24142641

Zootierliste (2020). Dahomey-Zwergrind. Available at: https://www.zootierliste.de/ (Accessed December 21, 2020).

Zwane, A. A., Schnabel, R. D., Hoff, J., Choudhury, A., Makgahlela, M. L., Maiwashe, A., et al. (2019). Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa. *Front. Genet.* 10, 273. doi:10.3389/fgene.2019.00273

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fgene.2021.746665

Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification

Enrico Mancin[*†], *Beniamino Tuliozi*[†], *Sara Pegolo, Cristina Sartori and Roberto Mantovani*

Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padua, Legnaro, Italy

OPEN ACCESS

Edited by:
Olivier Hubert Hanotte, University of Nottingham, United Kingdom

Reviewed by:
Monica Correa Ledur, Embrapa Suínos e Aves, Brazil
Lingyang Xu, Institute of Animal Sciences (CAAS), China

***Correspondence:**
Enrico Mancin
enrico.mancin@phd.unipd.it

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *24 July 2021*
Accepted: *02 December 2021*
Published: *04 January 2022*

Citation:
Mancin E, Tuliozi B, Pegolo S, Sartori C and Mantovani R (2022) Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification. Front. Genet. 12:746665. doi: 10.3389/fgene.2021.746665

Knowledge of the genetic architecture of key growth and beef traits in livestock species has greatly improved worldwide thanks to genome-wide association studies (GWAS), which allow to link target phenotypes to Single Nucleotide Polymorphisms (SNPs) across the genome. Local dual-purpose breeds have rarely been the focus of such studies; recently, however, their value as a possible alternative to intensively farmed breeds has become clear, especially for their greater adaptability to environmental change and potential for survival in less productive areas. We performed single-step GWAS and post-GWAS analysis for body weight (BW), average daily gain (ADG), carcass fleshiness (CF) and dressing percentage (DP) in 1,690 individuals of local alpine cattle breed, Rendena. This breed is typical of alpine pastures, with a marked dual-purpose attitude and good genetic diversity. Moreover, we considered two of the target phenotypes (BW and ADG) at different times in the individuals' life, a potentially important aspect in the study of the traits' genetic architecture. We identified 8 significant and 47 suggestively associated SNPs, located in 14 autosomal chromosomes (BTA). Among the strongest signals, 3 significant and 16 suggestive SNPs were associated with ADG and were located on BTA10 (50–60 Mb), while the hotspot associated with CF and DP was on BTA18 (55–62 MB). Among the significant SNPs some were mapped within genes, such as *SLC12A1*, *CGNL1*, *PRTG* (ADG), *LOC513941* (CF), *NLRP2* (CF and DP), *CDC155* (DP). Pathway analysis showed great diversity in the biological pathways linked to the different traits; several were associated with neurogenesis and synaptic transmission, but actin-related and transmembrane transport pathways were also represented. Time-stratification highlighted how the genetic architectures of the same traits were markedly different between different ages. The results from our GWAS of beef traits in Rendena led to the detection of a variety of genes both well-known and novel. We argue that our results show that expanding genomic research to local breeds can reveal hitherto undetected genetic architectures in livestock worldwide. This could greatly help efforts to map genomic complexity of the traits of interest and to make appropriate breeding decisions.

Keywords: genome-wide association, alpine breeds, single step genome-wide association study, local cattle breed, beef traits, time stratification, livestock conservation

1 INTRODUCTION

Genome-wide association is a powerful analysis that allows to identify genomic regions associated with phenotype variations in a target population to understand better the genetic architecture of the phenotype (Begum et al., 2012); such analysis has proved to be invaluable in the study of the genetic architecture of livestock species traits, especially cattle (Schmid and Bennewitz, 2017). Most of the target traits in livestock are polygenic phenotypes (de Oliveira Silva et al., 2017), which are suitable for investigation with robust GWAS. However, the GWAS is only the start of the investigation of the target traits genetic architecture (Atwell et al., 2010). Weaker signals that would be missed by GWAS analysis can be identified and described via pathways enrichment analysis, under the assumption that these signals are related to genes involved in complex pathways and biological processes (Buitenhuis et al., 2014; Pegolo et al., 2020). In beef cattle, traits such as growth or carcass conformation are critical to the profitability of meat production since greater growth means a shorter fattening period, and more conformed animals have higher economic value (Samorè et al., 2016). GWAS analysis in different species highlighted the strongly polygenic nature of these traits (Mateescu et al., 2017; Huang et al., 2018; Falker-Gieske et al., 2019; Gershoni et al., 2021).

In recent years, many studies have proposed more advanced approaches to investigate these phenotypes, such as the inclusion of whole genome sequences (Mao et al., 2016) or the analysis of growth traits in a longitudinal perspective (Yin and König, 2019). This latter approach has been scarcely used in beef cattle breeding (Yin and König, 2019; Gershoni et al., 2021), but there are dramatic differences in the functional elements involved in determining morphological traits at different ages (Helgeland et al., 2019): these differences could be investigated by separate analyses of the same trait collected at various ages. Investigations on beef traits (Mudadu et al., 2016) have been extensively performed in cattle, but most studies have regarded few cosmopolitan, specialized breeds. Dual-purpose breeds, which consist of local populations apart from a few exceptions (such as Simmental cattle), have rarely been the target of GWAS. Local breeds are genetically more diverse than the cosmopolitan ones and have generally better health parameters and fitness due to a much-reduced specialization (Biscarini et al., 2015). Also, the negative genetic correlations occurring between dairy and beef traits make the genetic improvement of both aptitudes in dual-purpose populations far from its optimum (Frigo et al., 2013; Mazza et al., 2016; Sartori et al., 2018). Moreover, such breeds often present unique characteristics that allow them to adapt to harsher conditions (Krupová et al., 2016; Sutera et al., 2021) and better respond to environmental shifts or challenges (Biscarini et al., 2015). Thus, these dual-purpose local breeds represent an unexploited source of diversity for the animal breeding sector and a rare opportunity to conduct GWAS on key economic traits that have not been under excessive specialization.

Rendena is an autochthonous breed from Alpine regions of North-East of Italy with a dual-purpose aptitude for meat and milk still maintained through the current selection scheme, assigning 65% of the economic weight to milk and 35% to meat (Guzzo et al., 2019; for further details on the selection scheme see Mantovani et al., 1997; and **Supplementary Figure S1**).

The dual-purpose aptitude also allows to counteract inbreeding erosion and maintain good genetic variability despite the small population size (the current number of animals is around 7,000 of which 4,000 are cows). Rendena also presents good fertility and longevity parameters and excellent adaptability to local environments, ranging from plains to Alpine pastures (Ovaska and Soini, 2017; Guzzo et al., 2018). As in various other local breeds, genomic information of Rendena has started to be available just recently, after implementing a routine activity of genotyping. This information might allow identifying and describing genes and functional pathways involved in the genomic architecture of traits of economic or functional interest (Senczuk et al., 2020). Moreover, as genomic selection has just been implemented in Rendena (Mancin et al., 2021a), investigating these traits could also be helpful to increase the prediction accuracy (see Tiezzi and Maltecca, 2015).

In this study, we performed a single-step GWAS and pathway analysis in Rendena cattle to investigate the genetic architecture of growth and carcass conformation traits, i.e., body weight, average daily gain, *in vivo* dressing percentage, and *in vivo* fleshiness (SEUROP grade). Additionally, body weight and average daily gain were analyzed using records taken at different ages, to study possible temporal variation in the genetic architecture of growth at the early stages.

2 MATERIALS AND METHODS

2.1 Animals and Phenotypes

All phenotypic records were collected at the performance test (PT) station of the National Breeders Association of Rendena cattle—ANARE, Trento Italy (www.ANARE.it). All phenotypes belonged to young (on average of 1 month of age) candidate bulls. About 60 young bulls are tested every year at the PT station for a total period of 11 months, following the criteria reported in Mantovani et al. (1997). Records have been collected since 1985, when PT started, until present times. The phenotypes collected during the PT are body weight (BW), average daily gain (ADG), carcass fleshiness (CF) and dressing percentage (DP). Both CF and DP are evaluated *in vivo* by 3 skilled operators at the end of the PT period and averaged to obtain the final score. The CF evaluation applies the same scores of post-mortem carcass appraisal established by the European Union Council (SEUROP), where the middle class (R) is equal to 100 points and other classes (upper or lower classes) correspond to 10-points-variations. Furthermore, the evaluation also considers sub-classes (e.g., R+ and R-for the middle class) that are spaced 3. 33 points from the class score. DP is a visual prediction of the post-mortem measure of DP: the operator makes a visual appraisal of the individual at the end of the performance test, offering an estimate of the expected DP—i.e., conformation—at slaughter (Mantovani et al., 1997). Average daily gain (ADG) is calculated as the linear regression of weight (BW) on age. For this

study, ADG and BW were collected at different stages of PT. ADG has been divided into ADG_i and ADG_f: ADG_i covers the daily gain of the first half of the testing period (since entering the PT station until the 6th month), while ADG_f covers the daily gain of the second half (from the 6th month to the end of the period). ADG covering the entire PT test was labeled as ADG_tot. BW was split along the same timeline as ADG: body weight at the entrance to the station (BW_i), at 6 months (BW_m) and at the end of PT (BW_f). Data cleaning consisted of removing animals with a regression of weight on age showing a coefficient of determination below 0.9 (for further details, see Guzzo et al., 2019).

2.2 Genomic Data and Quality Control

The biological material of the animals chosen for the genotyping resulted from salivary swab, hair (at least 30 bulbs), or ear tissue from biopsy brand, collected by ANARE on females and young candidate bulls at PT, as well as from semen of proven bulls, already subjected in the past to PT and progeny test for milk and to a large extent now eliminated. The Bovine 150K Array GGPv3 Bead Chip (HD, 138,974 SNPs), and Illumina Bovine LD GGPv3 (LD, 26,497 SNPs), were used for genotyping (Illumina, Illumina Inc., San Diego, CA, United States). The overlapping between the two panels is about 60%. The HD platform was used for 554 young bulls, while 1,416 individuals (174 males and 1,242 females) were genotyped with LD chips. To achieve a reliable genomic imputation accuracy, the 174 males were animals with at least one parent and one half-sib genotyped with HD chips. The genotyped females were individuals with a kinship of at least 0.2 with phenotyped animals.

Before proceeding with imputation, we performed a preliminary quality control removing SNPs with a minor allele frequency (MAF) < 0.01 and call rate lower than 0.90, using Plink program (Purcell et al., 2007). Only the 29 autosomal chromosomes (BTA) were used for association, and progeny conflicts were fixed using the seekparentsf90 program (Aguilar et al., 2018).

AlphaImpute2 was used for imputation (Whalen and Hickey, 2020), as it combines a population imputation algorithm (Positional Burrows Wheeler Transform) with pedigree-based imputation (iterative peeling); we used the same parameters as in Mancin et al. (2021a). The accuracy of the imputations was roughly estimated as a correlation between true and imputed SNPs. To this aim, ten rounds of cross-validation were performed: in each round the overlapping SNPs between the two panels were removed in ten animals and then imputed using the HD panel from young bulls as reference population (**Supplementary Table S1**). Subsequently, the correlation between the true and the imputed genotypes was calculated on these animals.

After imputation, we performed a second genomic quality control with the preGSf90 program (Aguilar et al., 2018): the SNPs with MAF lower than 0.05 and SNPs that deviated too much for the expected value of heterozygosis (i.e., Hardy-Weinberg Equilibrium) were removed. In accordance to Wiggans et al. (2012) the threshold for was set to 0.15: SNPs were deleted if $|\frac{n\ of\ heterozigois}{n} - 2pq| > 0.15$. In addition, SNPs with a call-rate < 0.90 and animals with a call rate < 0.95 were excluded. The final genomic database contained 1,690 animals (698 with both genotypic and phenotypic information), and 113,279 SNPs. Genome-wide linkage disequilibrium (LD) within chromosome was also calculated, as the squared correlation of allele counts for two SNP. Principal Components Analysis (PCA) of **G** matrix and LD were also calculated with pregsGSf90.

2.1 Single Step Genome-wide Association Analyses

Single step genome-wide association (ssGWAS) models were used to estimate allele substitution effect. In ssGWAS, the estimation of allele substitution effects was obtained from a linear transformation of the BLUP of breeding value under ssGBLUP model (Aguilar et al., 2019). Mancin et al. (2021b) showed the advantages of this method in terms of QTL detection and control of populations structure over two-step methods in which de-regression of breeding value as pseudo phenotype is required. This issue is particularly evident in the presence of unbalanced data (i.e., sex-limited traits). In fact, the ssGWAS allows the use of both male and female genomes even when analyzing a phenotype collected only in individuals of one sex.

The ssGBLUP model used in this analysis, written in matrix form, is the following:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{H}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'y \\ \boldsymbol{Z}'y \end{bmatrix} \quad (1)$$

Where phenotypes are included in vector **y**, **X** is the incidence matrix of fixed effects (group of contemporaries, cow parity class and months of birth), **b** is the vector of these effects. The contemporary group has 147 levels, with each level consisting of bulls grouped together at the Performance Test because homogeneous by age (i.e., born within 1 month of each other; 82.

Animals per group on average, minimum 5 and maximum 142). The parity order of cow has four classes (first parity; second parity; third to seventh parity; above the eighth parity), and the classes of months of birth correspond to the single months, as in Guzzo et al. (2019).

Z represents the incident matrix that relates the random genetic additive effects to the phenotype, with effects represented by vector ***a***. The vector of random residual error (**e**) has a normal distribution $N(0, I\sigma_e^2)$, where σ_e^2 is the residual variance. In the ssGBLUP vector of additive genetic effects is distributed as $N(0, H\sigma_a^2)$, where σ_a^2 is the additive genetic variance and **H** is the (co)variances structure which combines pedigree and genomic relationships (Aguilar et al., 2010). Its inverse, used in **Eq. 1** is described as:

$$\boldsymbol{H}^{-1} = \boldsymbol{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1} \end{bmatrix} \quad (2)$$

where $\boldsymbol{A}^{-1}$ and $\boldsymbol{A}_{22}^{-1}$ are the inverse of the pedigree kindship matrix respectively for all animals and for only genotyped animals. Since the frequencies of current genotyped population are used to

center **G** and pedigree and genomic matrices have different bases, **G** was adjusted so the average diagonal and off-diagonal matches the averages of $\mathbf{A}^{22}$. Pedigree kinship (sub) matrix was estimated tracing back the pedigree up to 7 generations, i.e., 6,644 animals. **G** matrix was built using the methods proposed by VanRaden (2008), as follows:

$$\boldsymbol{G}_0 = \frac{\mathbf{MM}'}{2\sum \boldsymbol{p}_i\,(1-\boldsymbol{p}_i)} \tag{3}$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and $\boldsymbol{p}_i$ is the allele frequency for the ith SNP (VanRaden, 2008).

Additionally, to avoid singularity problems, the final **G** was computed as

$$\boldsymbol{G} = \boldsymbol{\lambda G}_0 + \boldsymbol{\beta I} \tag{4}$$

Where **G** is the matrix present in the **Eq. 2**, **I** is an identity matrix of the same dimensions, λ and β are two weighting coefficients, with λ = 0.99 and β = 0.01. These values were chosen due to their influence on the power of signal detection of the GWAS, and because they resulted in inflation close to optimum values. In addition, G was adjusted to a better blending with diagonal and off-diagonal of A^{22} as described in Vitezica et al. (2011):

$$\delta = 1 - \frac{0.5}{n^2}\left(\sum_i \sum_j \boldsymbol{A}_{22\,(\mathrm{i,j})} - \sum_i \sum_j \boldsymbol{G}_{\mathrm{i,j}}\right) \tag{5}$$

Then, the vector of estimated breeding values was obtained as:

$$\hat{\boldsymbol{g}} = \lambda\delta \frac{1}{2\sum pq} \boldsymbol{M}'\boldsymbol{G}^{-1}\hat{a}_{22} \tag{6}$$

Where $\hat{\boldsymbol{a}}_{22}$ is the vector of estimated breeding values of genotyped animals. The prediction error variances $\hat{\boldsymbol{g}}$, necessary to calculate the *p*-values, were calculated following Gualdrón Duarte et al. (2014) and computed as in Aguilar et al. (2019), where:

$$Var\left(\hat{\boldsymbol{g}}\right) = Var\left(\lambda\delta \frac{1}{2\sum pq} \boldsymbol{M}'\boldsymbol{G}^{-1}\hat{a}_{22}\right) \tag{7}$$

$$Var\left(\hat{\boldsymbol{g}}\right) = \lambda\delta \frac{1}{2\sum pq} \boldsymbol{M}'\boldsymbol{G}^{-1}\mathbf{Var}\left(\hat{\boldsymbol{a}}_{22}\right)\boldsymbol{G}^{-1}\boldsymbol{M}\,\lambda\delta \frac{1}{2\sum pq} \tag{8}$$

Since $\boldsymbol{Var}\left(\hat{\boldsymbol{a}}_{22}\right)$ is equal to $\mathrm{PEV}\left(\hat{\boldsymbol{a}}_{22}\right) - \mathrm{var}\left(a_{22}\right)$; thus $\boldsymbol{Var}\left(\hat{\boldsymbol{a}}_{22}\right) = \mathbf{G}\hat{\sigma}_a^2 - \mathbf{C}^{22}$. It follows that formula **Eq. 8** becomes:

$$Var\left(\hat{\boldsymbol{g}}\right) = \lambda\delta \frac{1}{2\sum pq} \boldsymbol{M}'\boldsymbol{G}^{-1}\left(\boldsymbol{G}\hat{\sigma}_a^2 - \boldsymbol{C}^{22}\right)\boldsymbol{G}^{-1}\boldsymbol{M}\,\lambda\delta \frac{1}{2\sum pq} \tag{9}$$

$\mathbf{C}^{22}$ is a submatrix of **C** belonging to the genotyped animals and represents the prediction error variances of $\hat{\boldsymbol{a}}_{22}$. The *p*-values are then calculated as

$$\boldsymbol{p} - \boldsymbol{value}_i = 2\left(1 - \boldsymbol{\Phi}\left(\left|\frac{\hat{\boldsymbol{g}}_i}{\boldsymbol{sd}\left(\hat{\boldsymbol{g}}_i\right)}\right|\right)\right) \tag{10}$$

Where $\hat{g}_i$ is the allele substitution effect of SNP i and $sd\left(\hat{g}_i\right)$ represents the square root of **Eq. 9**, Φ (·) is the cumulative density function (CDF) of the normal distribution.

Two thresholds were used for the association tests: a genome-wide 5% significant level of −log10(p) = 5.55 (0.05/17,766) and a suggestive association with −log10(p) = 4.29 (0.1/17,766). These are the thresholds corrected for multiple tests i.e., $\frac{p}{n}$ where p is the probability level of significance and n is the corresponding number of independent SNPs (n = 17,766) calculated using the "poolR" R package (https://cran.r-project.org/web/packages/poolr; R Core Team, 2021), according to Li and Ji (2005). The number of independent tests was calculated based on the number of eigenvalues. Instead of the standard approach of Cheverud (2001), we used the approach by Li and Ji (2005), a function that decomposes the eigenvalues in the integral part (Effective Number Independent Test) and the nonintegral part.

The (co)variance components have been estimated with REML using Average-Information algorithm (Gilmour et al., 1995). Approximate standard error of (co)variance components has also been estimated through Monte Carlo sampling as in Houle and Meyer (2015), in which standard deviations were calculated from Monte Carlo chains sampled from multinormal distribution with covariance being the inverse of the Average Information Matrix and the estimated variances as the expectation. Then the heritability for the 3 phenotypes was calculated under single trait models as in **Eq. 1**. Heritability was calculated as: $h^2 = \frac{\sigma_a^2}{(\sigma_a^2+\sigma_e^2)}$; where σ_a^2 and σ_e^2 are, respectively, the additive genetic and the residual variances.

Genetic and phenotypic correlations were estimated with bi-traits models, which are equivalent to **Eq. 1** except for the animal additive genetic and residual variance, assumed to follow a multivariate normal distribution with mean 0 and variances **G** ⊗ **H**, and **R** ⊗ **I**, where

$$\boldsymbol{G} = \begin{vmatrix} \sigma_{a1}^2 & \sigma_{a1a2} \\ \sigma_{a1a2} & \sigma_{a2}^2 \end{vmatrix}; \boldsymbol{R} \begin{vmatrix} \sigma_{e1}^2 & \sigma_{e1e2} \\ \sigma_{e2e1} & \sigma_{e2}^2 \end{vmatrix} \tag{11}$$

where **G** is the matrix of additive genetic (co)variances σ_{a1}^2, σ_{a2}^2, σ_{a1a2} of traits 1 and 2, **R** the matrix of residual (co)variances σ_{e1}^2, σ_{e2}^2 and σ_{e1e2} of traits 1 and 2.

The correlation was estimated as: $cov = \frac{\sigma_{i1i2}}{(\sigma_{i,1}\ \sigma_{i,2})}$ where *i* stands for the genetic and phenotypic correlation; 1 and 2 refer to the different performance test traits, and σ_{i1i2} is the covariance between traits 1 and traits 2, off diagonal of **Eq. 11**. For phenotypic (co)variance, we mean the sum of the genetic and the phenotypic (co)variances. Traits that do not include zero in their correlations Higher Posterior Density Interval (HPD) were declared significantly correlated.

All the genomic analyses were carried out with BLUPF90 family software (Aguilar et al., 2018) following the procedure described in Lourenco et al. (2020). Manhattan plots were drawn using "ggplot" R package (Wickham, 2016), as were the LD graphs.

2.2 Pathway Analysis

Pathway's enrichment analysis was conducted to identify which biological pathways and functional elements were enriched for the investigated traits. From GWAS results, we selected SNPs with nominal *p*-values of < 0.01 which were mapped to genes based on a distance of 15 kb from the coding region using the

"biomaRt" R package (Drost and Paszkowski, 2017) and Bos taurus UMD3.1 assembly as in Pegolo et al. (2020). Functional enrichment analysis was carried out on the list of significant genes using the Cytoscape plugin ClueGo (Bindea et al., 2009). As functional categories, we used cellular component, biological process, and molecular functions within the Gene Ontology (GO, http://geneontology.org) database and the Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/). The Benjamini-Hochberg correction was applied to declare significant pathways: only pathways showing FDR < 0.05 were retained. The minimum number of genes in the pathway was set to 3; the minimum percentage of genes present in the pathway was set to 4%. To simplify the redundance of GO terms we provide figures with similar terms grouped based on their semantic similarity using the R packages "rrvgo" (Sayols, 2020). In addition, we investigated if the candidate regions declared as significant by our GWAS overlapped with QTL in animal QTLdb, identified with R package "GALLO" (Fonseca et al., 2020).

TABLE 1 | Summary statistics for phenotypic data of animals with both genotypic and phenotypic information (n = 689).

Traits	Mean	SD	Min	Max
BW_i (kg)	65.72	14.64	37	139
BW_m (kg)	183.40	30.53	83	317
BW_f (kg)	376.20	43.60	203	576
ADG_i (g/d)	939.20	167.90	138	1,388
ADG_f (g/d)	1,082	157.30	365	1756
ADG_tot (g/d)	1,024	124.20	474	1,562
CF (score)	99.05	3.80	80	111
DP (score)	54.18	0.94	50	57

BW_i, body weight at the entrance at performance test stations; BW_m, body weight at 6 months; BW_f, at the end of performance test; ADG_i, average daily gains covering the first half of the period (since entering into the PT, station until the 6th month); ADG_f, average daily gain covering the daily gain of the second half (since the 6th month to the end of the period), ADG_tot average daily gain covering the entire period; DP, Dressing Percentage; CF, Carcass Fleshiness; SD, Standard deviation; Min, minimum; Max, maximum.

3 RESULTS AND DISCUSSION

3.1 Heritability and Genetic Correlations

Descriptive statistics after data editing of the phenotypes are shown in **Table 1**. Phenotypic and genetic correlations and the heritability (h^2) for the analyzed traits are reported in **Table 2**. Body weight traits presented an average value of h^2 lower than other traits: BW_i showed the lower heritability (0.130), while BW_m and BW_f had heritability of 0.220. In fact, as reported in literature, a large discrepancy of values has been observed for heritability of body weights, and generally, traits similar to birth weight or weaning weight have a slightly lower heritability than weight measured in more advanced stages (Yin and König, 2018). Average daily gain (ADG_tot) presented an intermediate heritability of 0.322 partitioned into 0.164 and 0.220 for ADG in the first and last period. As for body weight, ADG presents lower h^2 in first stages of the performance test, and h^2 values agree with what has been found in the literature (Yin and König, 2018). The highest heritabilities were found for the traits related to the carcass conformation, with a value of 0.45 and 0.47 respectively for CF and DP, close to what was observed in other local dual-purpose or beef cattle (Albera et al., 2001; Sbarra et al., 2013; Mancin et al., 2021c). These traits also appeared highly genetic correlated. All ADG traits were moderately genetically correlated with them, with a value of 0.5 on average. On the contrary, body weight measured at the beginning of the performance test was not significantly correlated with CF and DP. Interestingly, the weights measured in more advanced periods showed an increase of genetic correlation with a value close to 0.7. Body weight and ADG also presented a strong genetic correlation with body weight traits, especially for the traits measured at the final stages of the performance test. In terms of genetic correlations, the results agree with what was found in other local dual-purpose or beef breeds (Veselá et al., 2011; Filipčík et al., 2020). Phenotypic correlation followed the same trends of genetic correlation but with a lower magnitude (**Table 2**, under diagonal).

3.2 Genomic Architecture and Imputation

A homogeneous density distribution (number of SNPs per Mb) was found throughout the genome, apart from few relatively small blank areas in 12 chromosomes. For further details on SNP density on each chromosome after imputation and quality control, see **Supplementary Figure S2**. The new imputed panel had a SNPs density close to the one found in the young bulls genotyped with HD platforms. A value of imputation accuracy of 0.95 ± 0.05 was observed via cross-validation in the HD males (**Supplementary Figure S2**). Combined with the high correlation between the A and G matrix, these results confirm the reliability of the new Alphaimpute2 algorithm for this population.

The PCA scatterplots (**Figure 1**) illustrate a homogenous distribution of allele frequencies in individuals that comprised our study population. No stratification has been observed in the first two components, suggesting that most G matrix variance is explained by many eigenvalues with small effect. Genome-wide linkage disequilibrium and MAF have also been explored since the availability of high-density SNP platforms permits to explore the LD decay at an unprecedented resolution. In addition, MAF and LD are useful for understanding differences in population history and demography and for its impacts for genome-wide mapping studies. LD decay per each chromosome is reported in **Supplementary Figure S3**. As expected, most tightly linked SNPs presented strong levels of LD while it rapidly declines when the distance increases. A within-chromosome LD average value of 0.19 ± 0.12 has been observed. When the distance between markers is lower than 1 Mb, the LD squared correlation between pairs of loci across autosomes (r^2) (Hill and Robertson, 1968) reached an average value of 0.17 ± 0.27, and when the distance was > 1 Mb LD decreased to 0.04 ± 0.09 (**Supplementary Figure S3**). Larger levels of LD have been observed for chromosome 6 (0.20), while lower levels of LD were observed for chromosome 28 (0.18). An average value of 0.29 ± 0.12 was observed for minor allele frequency;

TABLE 2 | Mean of genetic (over diagonal) and phenotypic (under diagonal) correlations, and heritability (diagonal) with the respective standard deviations in target traits in Rendena population, estimated under ssGBLUP models. (^{NS}) stands for non-significant correlations.

	BW_i	BW_m	BW_f	ADG_i	ADG_f	ADG_tot	CF	DP
BW_i	0.13 ± 0.08	0.99 ± 0.17	0.80 ± 0.10	0.52 ± 0.96	0.44 ± 0.85	0.50 ± 0.60^{NS}	0.33 ± 0.71	0.53 ± 0.80
BW_m	0.41 ± 0.05	0.22 ± 0.09	0.87 ± 0.11	0.81 ± 0.41	0.68 ± 0.36	0.78 ± 0.59	0.69 ± 0.58	0.73 ± 0.44
BW_f	0.29 ± 0.07	0.79 ± 0.03	0.22 ± 0.09	0.78 ± 0.43	0.97 ± 0.17	0.97 ± 0.28	0.62 ± 0.21	0.63 ± 0.23
ADG_i	0.17 ± 0.07	0.77 ± 0.03	0.86 ± 0.02	0.16 ± 0.10	0.64 ± 0.12	0.81 ± 0.21	0.62 ± 0.43	0.67 ± 0.25
ADG_f	−0.04 ± 0.08	0.09 ± 0.08	0.68 ± 0.04	0.14 ± 0.08	0.23 ± 0.08	0.97 ± 0.1	0.43 ± 0.23	0.47 ± 0.22
ADG_tot	0.11 ± 0.08	0.47 ± 0.06	0.84 ± 0.02	0.68 ± 0.04	0.80 ± 0.03	0.32 ± 0.09	0.55 ± 0.16	0.6 ± 0.15
CF	0.14 ± 0.08	0.4 ± 0.07	0.49 ± 0.09	0.3 ± 0.08	0.37 ± 0.08	0.42 ± 0.08	0.46 ± 0.09	0.98 ± 0.02
DP	0.05 ± 0.09	0.35 ± 0.08	0.51 ± 0.07	0.26 ± 0.09	0.38 ± 0.08	0.98 ± 0.02	0.73 ± 0.05	0.46 ± 0.09

BW_i, body weight at the entrance at performance test stations; BW_m, body weight at 6 months; BW_f, at the end of performance test; ADG_i, average daily gains covering the first half of the period (since entering into the PT, station until the 6th month); ADG_f, average daily gain covering the daily gain of the second half (since the 6th month to the end of the period), ADG_tot average daily gain covering the entire period; DP, Dressing Percentage; CF, Carcass Fleshiness.

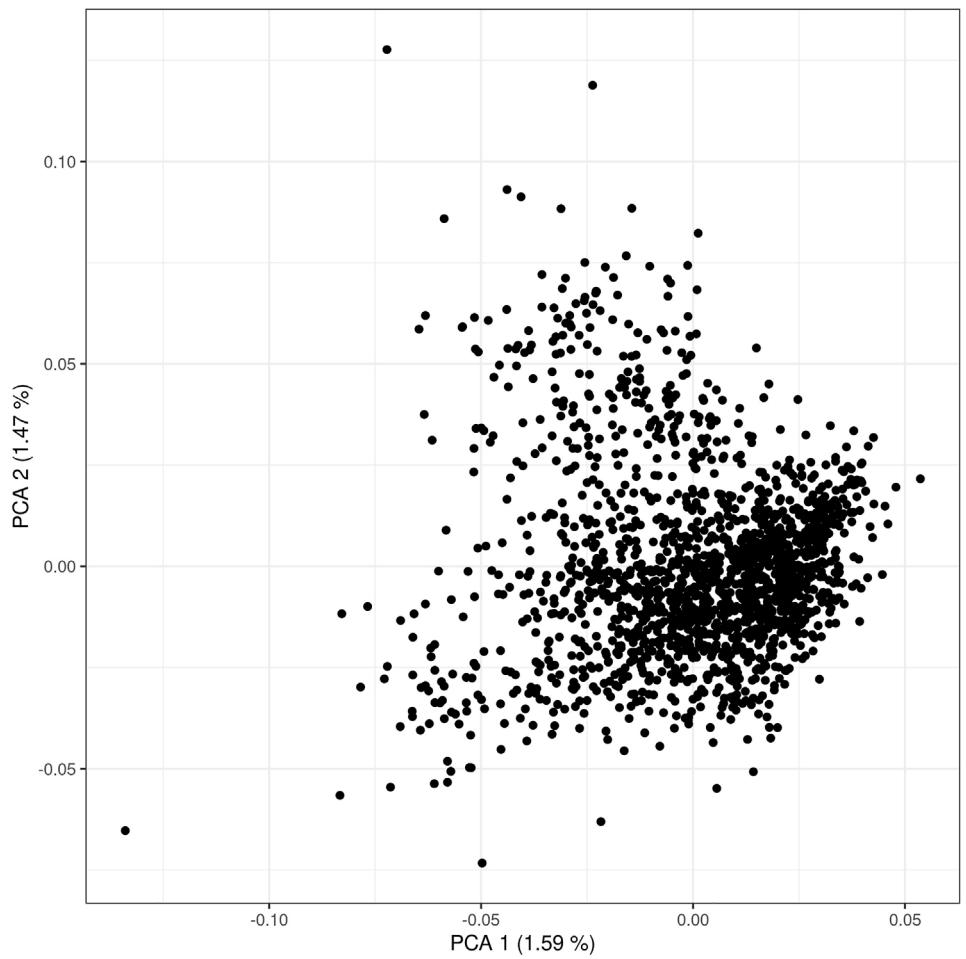

FIGURE 1 | Scatter plot of first and second principal components of the genomic relationship matrix (the G matrix) used in the ssGBLUP. A total of 113,279 SNPs and 1,690 cattle were used to perform the principal component analysis.

no noticeable difference has been observed along the 29 chromosomes, with MAF values ranging from 0.28 ± 0.12 (chromosome 12) to 0.30 ± 0.12 (chromosome 19). With respect to the other local Italian breeds (i.e., Fabbri et al., 2020), Rendena presents a lower level of LD. This issue implicitly underlines the reassuring demographic situation of Rendena compared with other indigenous cattle of Italy, as it demonstrates a lower risk of inbreeding depression.

TABLE 3 | Significant and suggestively SNPs found on the GWAS study.

Trait	BTA	Position of the SNP (bp)	Significance of the SNP (–log (*p*-value))	Nearest gene(s)	Distance to nearest gene (kb)	Other traits associated	Variance explained (%)
Body weight							
BW_i	**9**	**64,611,352**	**3.04E-06**	***TBX18***	**0.589**		0.22
BW_i	9	64,599,056	2.37E-05	*TBX18*	12.885		
BW_i	9	64,557,321	2.81E-05	*TBX18*	54.620		
BW_i	24	49,394,386	3.43E-05	*ACAA2*	48.389		
BW_i	24	49,493,559	4.43E-05	*MYO5B*	within		
BW_m	7	32,306,269	8.65E-06	*FTMT*	321.80	BW_f; ADG_i	
BW_m	1	67,212,088	3.27E-05	*DIRC2*	2.783	ADG_i	
BW_m	21	22,956,171	5.11E-05	*CPEB1*	within		
BW_m	24	49,735,783	5.55E-05	*MYO5B*	within		
BW_f	26	6,437,290	7.50E-06	*MBL2*	3.483	ADG_tot	
BW_f	7	32,306,269	8.39E-06	*FTMT*	321.80	BW_m, ADG_i	
BW_f	21	17,568,377	3.44E-05	*AGBL1*	within		
BW_f	24	24,130,452	4.56E-05	*CCDC178*	within		
BW_f	14	60,644,816	4.62E-05	*RIMS2*	within		
Average Daily Gain							
ADG_i	**1**	**67,212,088**	**2.84E-06**	***DIRC2***	**2.783**	**BW_m**	**0.441**
ADG_i	7	32,306,269	1.99E-05	*FTMT*	321.80	BW_m; BW_f	
ADG_i	7	32,009,625	3.03E-05	*FTMT*	25.152		
ADG_i	4	91,417,417	3.11E-05	*GRM8*	within		
ADG_f	**10**	**62,113,751**	**1.81E-07**	***SLC12A1***	within	**ADG_tot**	**0.073**
ADG_f	**10**	**52,785,760**	**1.29E-06**	***CGNL1***	within		**0.203**
ADG_f	**10**	**54,787,499**	**1.75E-06**	***PRTG***	within		**0.435**
ADG_f	10	55,502,036	3.42E-06	*UNC13C*	135.046		
ADG_f	10	55,510,249	3.56E-06	*UNC13C*	126.833		
ADG_f	10	55,535,781	4.35E-06	*UNC13C*	101.301		
ADG_f	10	57,348,706	6.68E-06	*LOC101904374*	248.031		
ADG_f	26	8,564,813	5.92E-06	*A1CF; ASAH2*	17.739; 32.479	ADG_tot	
ADG_f	10	52,777,666	9.27E-06	*CGNL1*	within		
ADG_f	10	57,311,183	9.77E-06	*LOC101904374*	285.554		
ADG_f	10	52,023,061	1.35E-05	*AQP9*	65.881		
ADG_f	10	56,585,283	1.56E-05	*WDR72*	within		
ADG_f	10	61,604,387	2.24E-05	*LOC104973175; FBN1*	20.944; 51.118		
ADG_f	10	58,180,258		*MYO5C; GNB5*	1.494; 11.943		
ADG_f	10	63,669,471	3.56E-05	—			
ADG_f	10	52,284,899	4.06E-05	*ALDH1A2*	within		
ADG_f	10	57,890,651	4.13E-05	*MYO5A*	within		
ADG_f	11	78,877,665	4.48E-05	*WDR35*	within		
ADG_f	10	55,830,543	4.90E-05	*UNC13C*	within		
ADG_f	10	57048787	4.98E-05	*LOC101904374*	547.950		
ADG_tot	**10**	**62,113,571**	**2.07E-06**	***SLC12A1***	within	**ADG_f**	**0.501**
ADG_tot	26	8,564,813	1.66E-05	*A1CF; ASAH2*	17.739; 32.479	ADG_f	
ADG_tot	11	21,542,682	3.41E-05	*CDKL4; MAP4K3*	7.971; 11.618		
ADG_tot	26	6,437,290	6.03E-05*	*MBL2*	3.483	BW_f	
Dressing Percentage							
DP	**18**	**62,412,976**	**4.51E-07**	***NLRP2***	within	**CF**	**0.640**
DP	**18**	**55,878,286**	**2.40E-06**	***CDC155***	within	**CF**	**0.731**
DP	1	148,893,434	8.77E-06	*SIM2*	80.004		
DP	18	58,645,859	1.06E-05	*LOC101904435*	within	CF	
DP	18	61,137,684	1.15E-05	*LOC513941*	within	CF	
DP	4	99,574,406	2.34E-05	*LOC112446424*	within		
DP	18	57,735,853	3.03E-05	*LOC787554*	within	CF	
DP	18	62,427,814	4.49E-05	*NLRP2*	within		
DP	18	63,362,491	4.97E-05	*LOC107131476*	560		
DP	17	72055006	5.07E-05	*YPEL1*	23.650		
DP	18	62,428,754	5.25E-05	*NLRP2*	within		

(Continued on following page)

TABLE 3 | (*Continued*) Significant and suggestively SNPs found on the GWAS study.

Trait	BTA	Position of the SNP (bp)	Significance of the SNP (–log (*p*-value))	Nearest gene(s)	Distance to nearest gene (kb)	Other traits associated	Variance explained (%)
Carcass Fleshiness							
CF	**18**	**61,137,684**	**5.62E-08**	***LOC513941***	within	**DP**	**0.450**
CF	**18**	**62,412,976**	**9.40E-07**	***NLRP2***	within	**DP**	**0.670**
CF	18	58,645,859	4.71E-06	*LOC101904435*	within	DP	
CF	18	55,878,286	7.67E-06	*CCDC155*	within	DP	
CF	18	61,920,892	9.57E-06	*ZNF784*	895		
CF	18	57,735,853	1.05E-05	*LOC787554*	within	DP	
CF	18	57,516,245	1.66E-05	*LOC618268*	within		
CF	14	45,804,718	2.30E-05	*SAMD12*	within		
CF	28	14,722,675	2.48E-05	*LOC101906006*	within		
CF	18	57,565,406	3.23E-05	*SIGLEC5*	within		
CF	12	27,043,078	3.38E-05	—			
CF	18	57,008,781	4.83E-05	*KLK12*	within		
CF	28	14,788,560	5.31E-05	*PHYHIPL*	within		

*Significant SNPs are reported in bold. Gene with * were just outside suggestive association range for one trait; it was retained in the table because significant for another trait. The threshold of significance chosen for our analysis was* p *= 3.162 * 10-6, obtained through Bonferroni correction, while threshold for Bonferroni suggestive* p*-values was p = 5.629 * 10-5.*

3.3 GWAS and Pathway Analysis

The full results of GWAS are reported in **Table 3**. We found a total of 8 SNP significantly associated with 5 of the investigated traits, and 47 SNPs suggestively associated with all 7 investigated traits (**Figure 2**). Pathway analysis revealed that out of 113,279 SNPs, 77,506 were located within a 15 kb window of annotated genes; in the end, 14,380 annotated genes were used as a background for each trait. On average, 628 genes near significant SNPs (<0.01) were identified and subsequently used for pathway analysis of each trait. All traits presented an inflation factor close to optimum values of 1 (**Figure 2**) calculated based on the median chi-squared test. In addition, analysis on localized linkage disequilibrium (0.5 Mb form significant SNP), has been carried out (**Figures 3**–**7**), and results indicated that all significant candidate genes are extremely close to the significant SNPs, except for candidate gene *ZNF784*, which is situated between two significant SNP (**Figure 6**).

3.3.1 Body Weight

Significant SNPs contributing to the genetic effect of body weight are listed in **Table 3**. Body weight measured at first stage was the only BW trait in which significant SNPs were identified, while body weight measured at the half of the performance test period presented the lowest number of suggestive SNPs and biological pathways enriched. The significant peak for BW_i was located at 64 Mb on BTA9, in the vicinity of gene *TBX18* (**Figure 3**; **Table 3**). This gene is mainly involved in controlling the first stages of embryonic development and in the morphogeny of the embryonic epithelium (Consortium, 2021). To our knowledge, no previous connection with body weight had ever been found for *TBX18*; however, a study found an association between this gene and development in dual-purpose Simmental breed but not in other specialized breeds (Doyle et al., 2020a). We hypothesize that a possible mechanism for the connection between *TBX18* and body weight could lie in the fact that it is a strict paralogue of *TBX15*, a gene linked to obesity-related traits in humans and mice (Ejarque et al., 2019; Sun et al., 2019); it is demonstrated that *TBX15* regulates processes related to the skeletal muscles metabolism, which is in turn linked to animals' body size (Lee et al., 2015). However, studies on the relationship between *TBX15* and *TBX18* in cattle and the impact of *TBX15/18* on the regulation of muscle metabolism are needed to validate this hypothesis. We identified several known cattle QTLs in QTLdb overlapping with our candidate region (**Supplementary Table S2A**): the majority of these QTLs were linked to morphology (47.5%), followed by beef production (22.5%).

MYO5B is a candidate gene for both BW_m and BW_i (**Table 3**), identified by the presence of two suggestively associated SNPs located on chromosome 24. *MYO5B* is related to the development of skeletal muscle for what concerns actin and myosin organization and with the binding of ATP (Consortium, 2021). Interestingly, this gene was also identified in GWAS conducted on dual-purpose Simmental breeds (Doyle et al., 2020b).

The analysis of the enriched pathways, represented in **Figure 8**, reinforced what has been mentioned for the single genes, namely that in our study the mechanisms regulating body weight were mainly those linked to the development of muscle masses. Among the GO terms enriched (**Figure 8**; **Supplementary Figures S4A,B**), there were: organization of cytoskeleton (GO:0007010), actomyosin structure (GO: 0031032), actin filament bundle (GO:0061572), and contractile actin filament bundle assembly (GO:0051017). The pathways analysis revealed a further biological process related to the metabolism of lipids on skeletal muscles (GO:0055088, GO: 0055092, GO:0042632). Regulation of the selection of appropriate nutrients by the skeletal muscle is essential both in terms of muscle energy metabolism and in terms of general regulation of whole-body supply and use of fuel (Hocquette et al., 1998): again, this enriched pathway was also found in Srivastava et al. (2020).

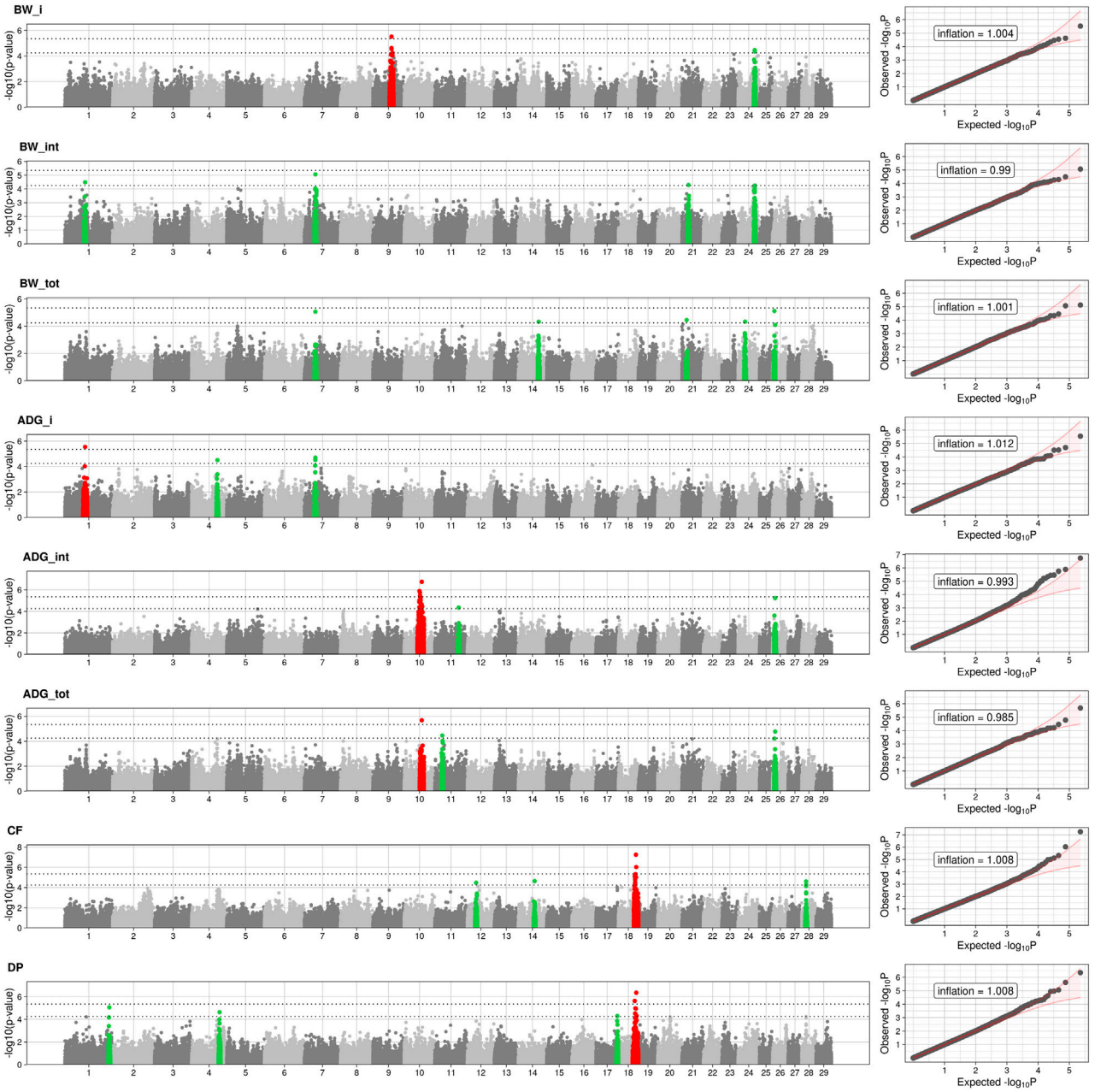

FIGURE 2 | Manhattan and Q-Q plots of BW_i: body weight at the entrance at performance test stations; BW_m: body weight at 6 months; BW_f: body weight at the end of performance test. Average daily gain: ADG_i, covers of the first half of the period (since entering into the PT station until the 6th month); ADG_f, covers the daily gain of the second half (from the 6th month to the end of the period); ADG_tot is the average daily gain throughout the entire period. DP, Dressing Percentage; CF, Carcass Fleshiness. Dotted lines represent the suggestive and the significant threshold. Red dot represented the significant SNPs and neighboring SNPs (±1 Mb) while green dot are the SNPs and neighboring SNPs (±1 Mb). Q-Q plots are displayed as scatter plots of observed and expected –log10 (*p*-values) **(right)**. Values of inflation are reported within the QQplots.

Aside from the already mentioned *MYOB5,* two candidate genes within suggestively associated SNPs were identified for BW_m: *CPEB1* and *DIRC2*, found on BTA1 and 21, respectively (**Table 3**). While these genes are not directly involved with body weight, we found them related to factors with a potential secondary impact on growth. For example, the *CPEB1* gene is involved in the regulation of mRNA translation and cell proliferation, with an influence on the molecular mechanisms associated with superior resilience to heat stress in cattle (Livernois et al., 2021). Moreover, *CPEB1* was also detected by other GWAS studies in cattle in which the target phenotype was residual feed intake (Lapierre et al., 1995). *DIRC2* has been associated with lipid storage in geese's (*Anser anser domesticus*) liver (Yang et al., 2020), given its role as a substrate carrier.

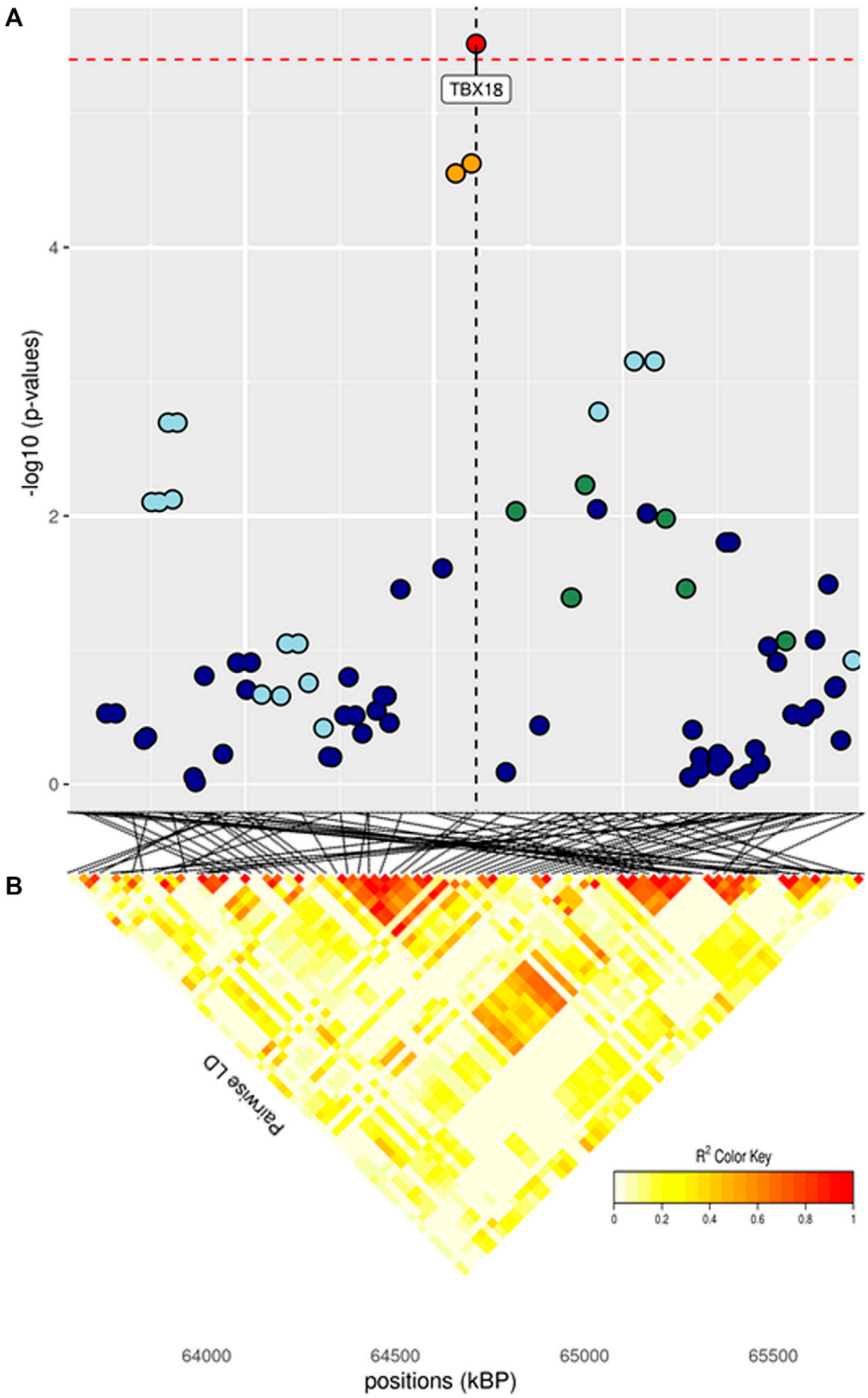

FIGURE 3 | (A) Localized linkage disequilibrium analysis of BW_i. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate gene *TBX18*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** Represents linkage disequilibrium of that area.

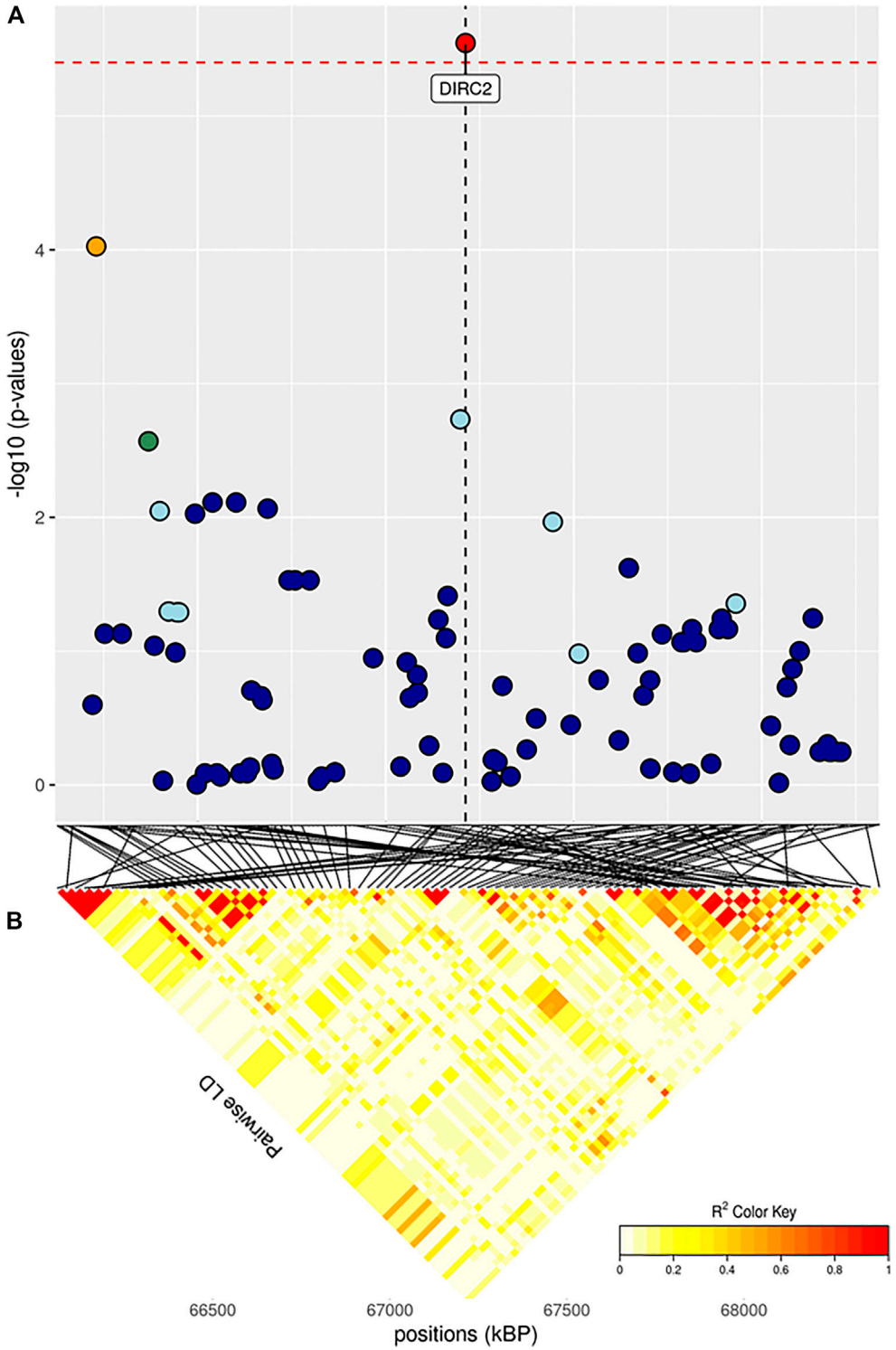

FIGURE 4 | (A) Localized linkage disequilibrium analysis of ADG_i. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate gene *DIRC2*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** Represents linkage disequilibrium of that area.

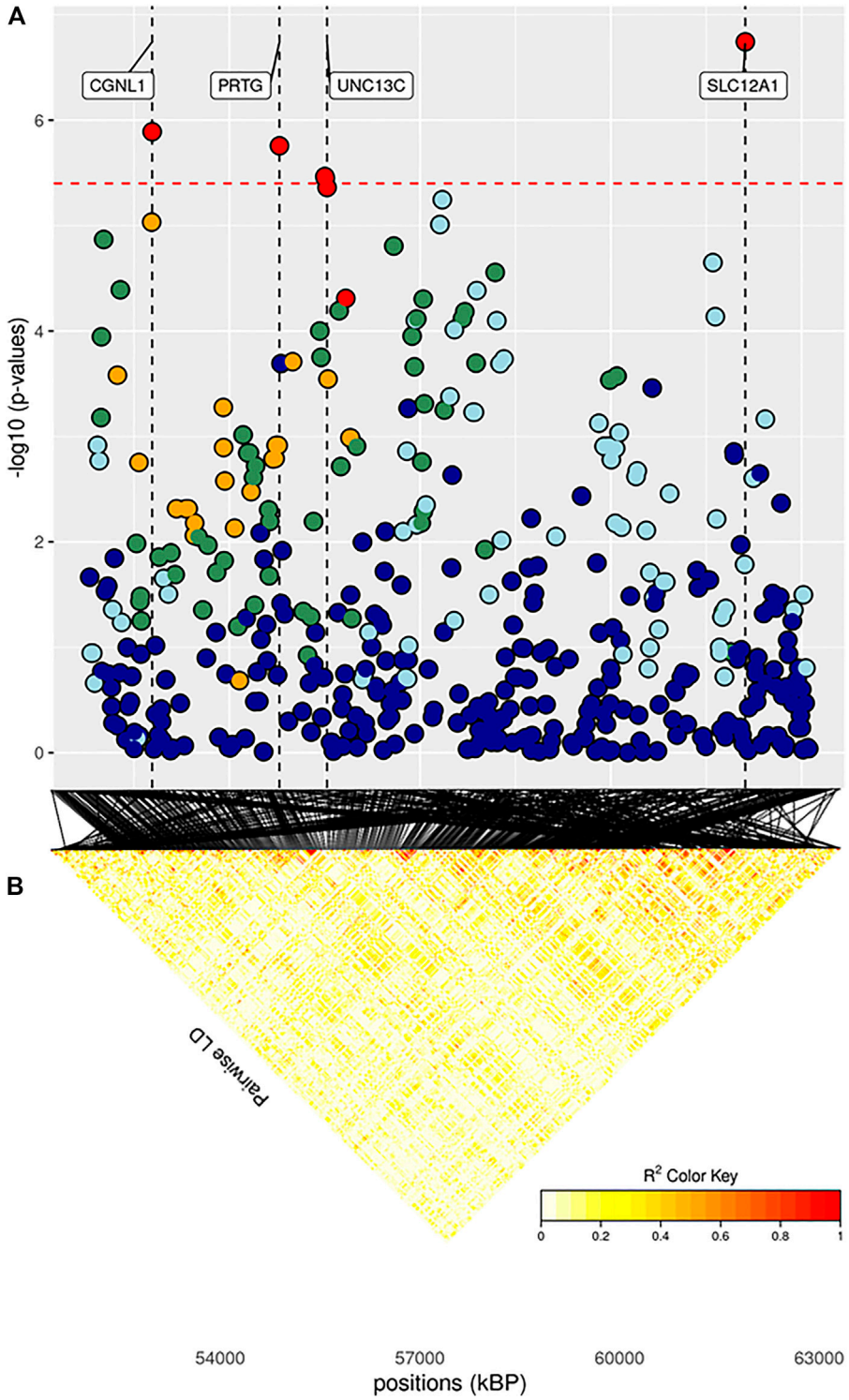

FIGURE 5 | (A) Localized linkage disequilibrium analysis of ADG_f. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate genes *CGNL1*, *PRTG*, *UNC13C* and *SLC12A1*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** the represents Linkage disequilibrium present of that area.

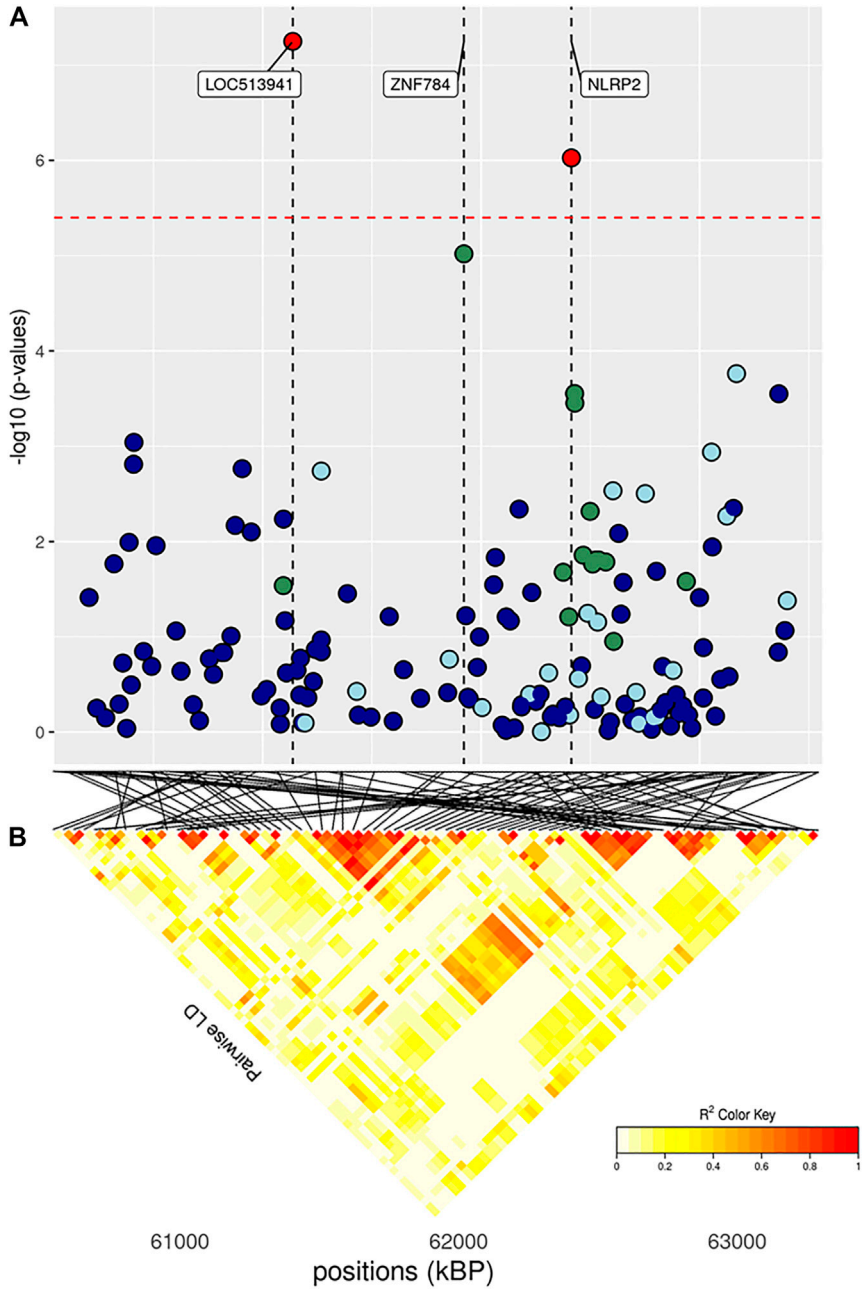

FIGURE 6 | (A) Localized linkage disequilibrium analysis of DP. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate genes *LOC513941*, *ZNF784* and *NLRP2*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** the represents Linkage disequilibrium present of that area.

In BW_f, as in the other phenotypes, several genes identified by suggestively associated SNPs (**Table 3**) had never been associated before with body size traits. Moreover, connections between such candidate genes and body weight were not straightforward. One suggestively associated gene for BW_f, *CCDC178*, was identified in some GWA studies on disease resistance in local cattle (Kosińska-Selbi et al., 2020). The *MBL2* gene, a candidate gene suggestively associated to BW_f (and almost suggestive for ADG_tot), also seems to have an indirect connection with body weight: *MBL2* plays a central role in the activation of the mannose-binding lectin or mannose-binding protein; this protein is involved in processes that regulate the immune system, preventing infection from bacteria, virus, and yeast (Consortium, 2021).

No biological process strictly related to muscle mass development was identified (**Figure 8**; **Supplementary Figure**

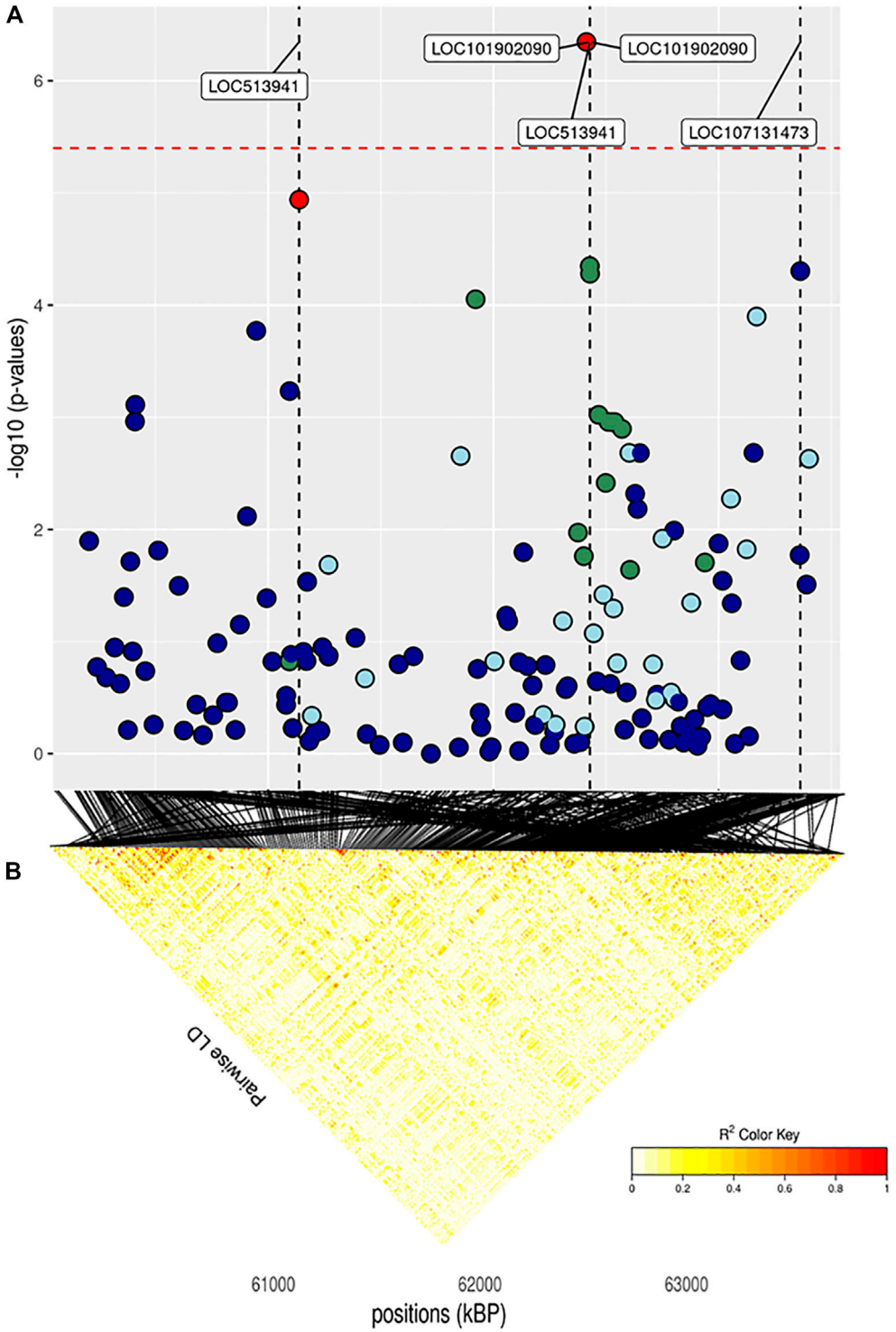

FIGURE 7 | (A) Localized linkage disequilibrium analysis of CF. Manhattan plots displaying the level of significance (y-axis) over genomic positions (x-axis) in a window of 0.5 Mb upstream and downstream of the most significantly SNP. Vertical line represents the position of candidate genes *LOC513941*, *NLRP2* and *LOC107131373*. Different colors are used to represent the pairwise LD with the closest significant SNPs: blue < 0.2; light blue < 0.4; green < 0.6; yellow < 0.8 and red > 0.8. **(B)** the represents Linkage disequilibrium present of that area.

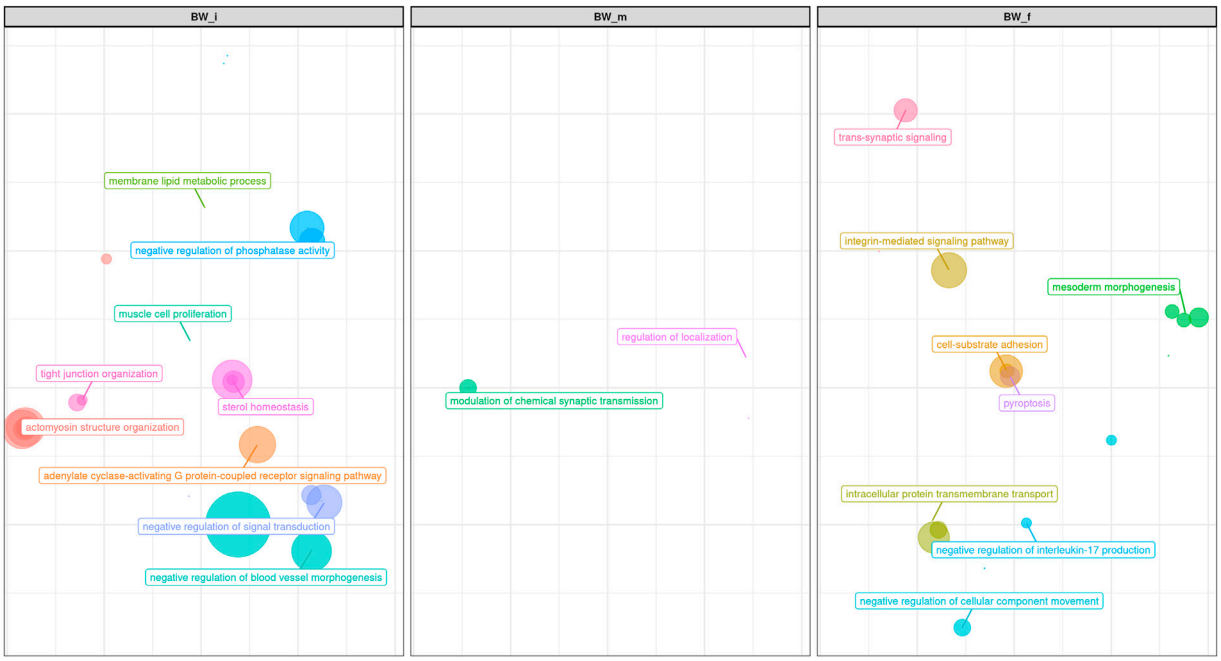

FIGURE 8 | Scatter plot representing the main groups of biological pathways enriched for Body Weight traits measured at first, half and final period of performance test (BW_i, BW_m, BW_f); the area represents the number of pathways in that group, among the total. For a detailed list of the pathways enriched by these traits see **Supplementary Figures S4A–C**.

S4C), but many processes related to other aspects of growth and body weight have been found. Several pathways were involved in GABA processes (**Figure 8**; **Supplementary Figures S4A–C**): GABA is actively involved in regulating leptin, the satiety hormone, which has an essential role in nutrient intake and feeding motivation (Miller 2017). Some pathways also appear to be associated with processes such as morphogenesis of the epithelium (GO:0048791, GO:0007492, GO:0048332, GO: 0001707 GO:0035987; mesoderm morphogenesis in **Figure 8**), which has a connection with body weight (increased paracellular permeability for the absorption of nutrients leads to augmented energy intake (Vanvanhossou et al., 2020).

Finally, many enriched terms were related to neuronal aspects (i.e., GO:0043005 GO:0097060, GO:0099537; **Figure 8**; **Supplementary Figures S4A–C**): this may find justification in the many studies underlining how these pathways are linked to the complex interaction between physio- and behavioral components that control the intake of food and energy expenditure (Martinez, 2000).

3.3.2 Average Daily Gain

Both GWAS and pathway analyses of Average Daily Gain showed different results depending on the age at which the trait was recorded, similarly to what resulted from our analysis of BW. In particular, the only GO terms in common between ADG_i and ADG_f were GO:0031175 (neuron projection development) and its associated terms; all the other 105 GO, and KEGG terms were not (**Supplementary Figure S4D**). The result of the GWAS also highlighted SNPs present in wholly different BTAs (**Table 3**). ADG_i had only one significant SNP (also suggestively associated with BW_m) situated on BTA1 (**Figure 4**), 0.2 Mb away from gene *DIRC2* (also associated with BW_m) and 1.1 Mb away from gene *HSPBAP*. Both loci can be in some ways considered candidate genes for growth, as also *HSPBAP* has already been associated with residual feed intake from birth to 12 months (Cohen-Zinder et al., 2016). One suggestively associated SNP for ADG_i on BTA4 (**Table 3**) was within candidate gene *GRM8*, associated with body size in cattle (Chen et al., 2020) and eating behavior in other mammals (Gast et al., 2013). Again, in agreement with what was found for BW_m (the measure of ADG_i is based on the difference between BW_m and BW_i measurement), the results of the pathway analysis for ADG_i were less extensive than for other ADG traits (**Figure 9**; **Supplementary Figures S4D–F**); moreover, out of 20 pathways (**Supplementary Figure S4D**), those readily associable with ADG were GO:0004629 phospholipase activity (crucial for lipid metabolism) and GO:0043124, responsible for negative regulation of l-kB kinase/NF-κB signaling (involved with metabolic regulation, especially in cases of overnutrition; Kracht et al., 2020).

The same trait recorded at a later age, ADG_f, showed a much greater number of results, similarly to what transpired with BW_f (**Figure 9**; **Supplementary Figure S4E**; **Table 3**). For trait ADG_f the region with the greatest number of signals was on BTA10, roughly between 50 and 60 Mb (**Figure 5**; **Table 3**). This region contains a QTL that has already been associated to growth in cattle (Mao et al., 2016), although not in the present study. The three significant SNPs and 14 out of 16 suggestively associated

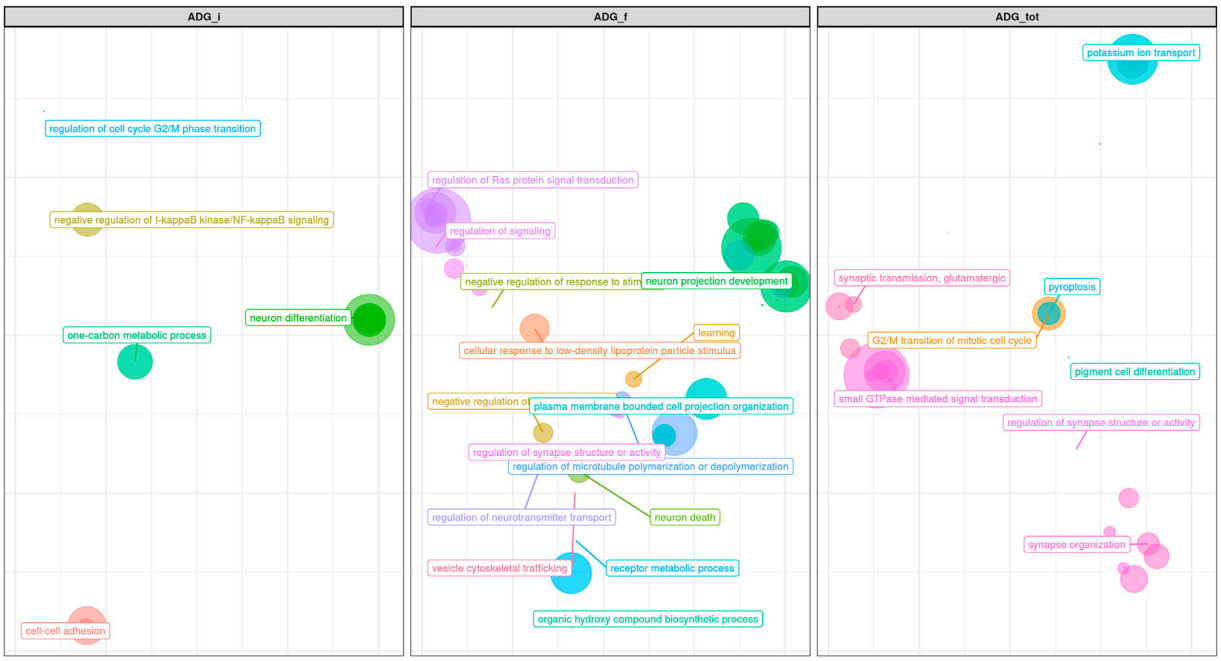

FIGURE 9 | Scatter plot representing the main groups of biological pathways enriched for average daily gain traits measured at first, half and total period of performance test (ADG_i, ADG_f, ADG_tot); the area represents the number of pathways in that group, among the total. For a detailed list of the pathways enriched by these traits see **Supplementary Figures S4D–F**.

SNPs were found in this region. Significant SNPs were situated within *SLC12A1*, *CGNL1* and *PRTG* genes (**Figure 5**). While the latter two have already been associated respectively with growth (Londoño-Gil et al., 2021) and backfat thickness in cattle (Júnior et al., 2016), *SLC12A1,* to our knowledge, has never been associated with growth or weight traits in cattle (but see Kemter et al., 2014, for evidence in mice). However, among the suggestively associated SNPs on BTA10 (**Table 3**), several were within or close genes highly important for ADG, such as *ALDH1A2*, *FBN1,* and *AQP9* (Hirano et al., 2012; Liu et al., 2019; Londoño-Gil et al., 2021; Zhang et al., 2021). **Figure 9** shows that enriched pathways spanned several macro-categories (**Figure 9**; **Supplementary Figure S4E**): these results suggest that, as for BW, during the late months of the first year, a complex interplay of different biological processes takes place in growing bulls. For what concerned the overlapping of our QTLs associated with ADG_f with the animal QTLdb, we identified QTLs from several studies: 28.77% associated with morphology, 21.92% associated with beef production, 19.18% associated with milk, and 8.22% associated with meat and carcass (**Supplementary Table S2B**).

Finally, for the total ADG, ADG_tot, the results obtained mirrored those obtained with final ADG, both in terms of significant and suggestive SNPs (on BTA10 and BTA26; **Table 3**) and in terms of GO terms (**Figure 9**; **Supplementary Figure S4F**) and candidate genes, such as *SLC12A1*. Interestingly, one signal reported in ADG_tot was not present in ADG_f: on BTA11, one single suggestively associated SNP was located close to two genes well known for their effect on feed intake and weight (*CDKL4* and *MAP4K3*; Edea et al., 2020). Apart from this exception, our results show conclusively that total average daily gain mirrored the final part of the daily gain, i.e., that the last months were decisive in shaping the total weight gain trajectory of the bulls.

3.3.2 Carcass Traits

The main region of interest for both CF and DP traits was situated on a gene-rich region of BTA18, between 55 and 62 Mb, where 3 significant and 9 suggestively associated SNPs allowed to locate several candidate genes (**Figure 6**; **Table 3**). The QTL with the highest significance for CF (suggestively associated for DP) was located within candidate gene *LOC513941* (**Figure 7**), translating into a cationic amino acid transporter 3-like. This type of transporters regulates the metabolism of cationic amino acids, a key factor for growth and beef characteristics in cattle (Liao et al., 2009). Further corroboration of the importance of this metabolic pathway for CF was the enrichment of 10 GO terms (**Figure 10**; **Supplementary Figure S4H**), within the group of "amino acid transport," such as amino acid transmembrane transporter activity (GO:0015171), and amino acid transmembrane transport (GO:0003333).

A second SNP in the same region (significant for DP and suggestively associated for CF; **Table 3**) was located within gene *CCDC155* (Coiled-coil domain containing 155). This gene encodes for a protein involved in dynein complex binding and actin filament organization and it has been associated with beef conformation (Lemos et al., 2016; Hardie et al., 2017). Apart from being the main component of the cytoskeleton, actin constitutes together with myosin the myofilaments, which grant muscle cells

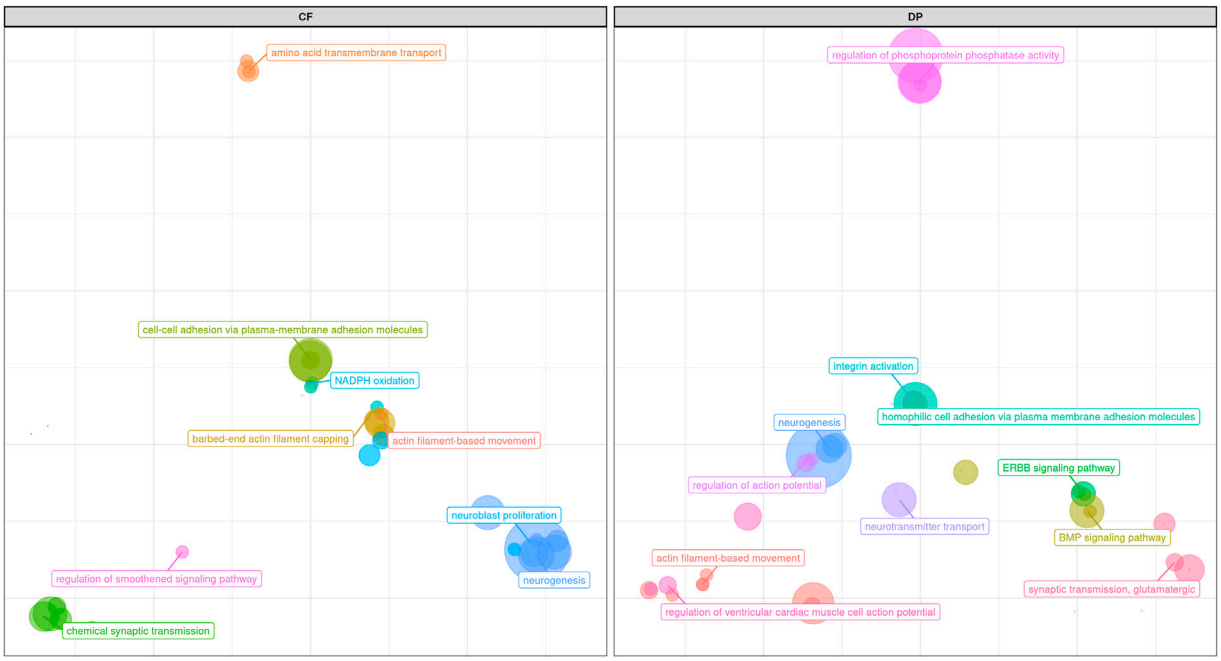

FIGURE 10 | Scatter plot representing the main groups of biological pathways enriched for carcass traits (carcass fleshiness and dressing percentage). For a detailed list of the pathways enriched by these traits see **Supplementary Figures S4G,H**.

their mobility and thus ultimately their organization and dynamics. The association of actin filaments and carcass traits was again made apparent also by the number (more than 30) and diversity of enriched GO terms related to actin (**Figure 10**; **Supplementary Figures S4G,H**): for example, those related to GO:0098858 (CF), actin-based cell projection; GO:0030048 (CF and DP), actin filament-based movement; GO:0070161 (CF and DP), anchoring junction; GO:0030833 regulation of actin filament polymerization; GO:0005912 (CF and DP), adherens junction (Londoño-Gil et al., 2021). Similarly, for DP 20 terms were enriched for pathways associated with actin filament-based GO terms (**Supplementary Figure S4G**).

In the same region of BTA18, our analysis found two more candidate genes with a known association with size and growth traits, all with one or more suggestively associated SNPs for CF. *Siglec-5* is a gene commonly found in GWAS concerning cattle size and growth traits; its over-expression indicates a deficiency of leptin, and thus longer gestation time and bigger fetuses (Hardie et al., 2017). *KLK12* is a kallikrein gene, a serin protease associated with food intake and feed efficiency at the transcript level in backfat and rumen (Kern et al., 2016). *LOC101904435* and *ZNF784* are zinc-finger proteins: the former is suggestively associated with both CF and DP; the latter only with CF but is linked to food intake in cattle (Olivieri et al., 2016).

Finally, three more SNPs (one significant both for CF and DP and two SNPs suggestively associated for DP) were situated within *NLRP2* gene (NACHT, LRR and PYD domains-containing protein 2), a key player in early embryogenesis, maternal effects, immune response, and inflammasome (Peng et al., 2012).

Taken together, these results about carcass traits have numerous substantial implications. Firstly, we highlight how the 57–62 Mb region on BTA18 can truly be considered a hotspot of genetic diversity in this breed (as it is for several others; Grigoletto et al., 2020; Purfield et al., 2020). Secondly, as expected with strongly correlated traits, CF and DP shared part of their genetic architecture, as significant SNPs for the two traits are mostly in the same region. Only another region was shared, as both traits reported two suggestively associated SNPs close to each other on BTA28 (**Table 3**). The region encompasses the *PHYHIPL* gene, which influences feed efficiency (Abo-Ismail et al., 2018), whose link with carcass traits has recently been established (Seabury et al., 2017).

CF was associated only with two more SNPs, one on BTA12 and the other on BTA14 (**Table 3**). While the former was more than 1 Mb far away from any annotated functional element, the latter fell within *SAMD12*, a gene already found to have a significant dominance signal to chuck roll and be associated with 18-months weight in Simmental (Zhuang et al., 2020). On the other hand, DP had an almost significant signal on BTA1: the gene closest to the SNP was *SIM2*, already known to be associated with carcass quality, differentiation of *longissimus,* and *semimembranosus* muscle (De Las Heras-Saldana et al., 2019; Edea et al., 2020). To conclude, the strongest of the remaining suggestively associated signals for DP came from BTA4, within *LOC112446424*, a non-coding RNA close to candidate gene *SLC13A4*, a cationic canal important both for muscle traits in sheep and growth and development in cattle (Carvalho et al., 2020; Kaur et al., 2020).

While, as we mentioned, results from pathway analyses (represented in **Figure 10**; **Supplementary Figures S4G,H**),

and GWAS were often complementary, pathway analyses for both CF and DP resulted in the enrichment of a robust number of pathways related to neuron activity, not really pointed out by GWAS results. Such pathways referred to the regulation of neuroblast proliferation (GO:1902692 for CF), chemical synaptic transmission (GO:0007268 for CF), neurogenesis (GO:0022008 for CF and DP), neuron projection (GO:0043005 for DP), synapse (GO:0045202 for DP) and especially synaptic transmission, glutamatergic (GO:0035249 for DP and, to a lesser extent, CF).

Glutamatergic synapses guide the development of growth neurons and regulate feeding motivation in the hippocampus (Huang et al., 2017). The relation between feeding motivation and nutrient intake is crucial to maintaining energy intake and storage (Illius et al., 2002). Such relationship is complex, involving leptin (see above-mentioned gene *Siglec-5*), and the NPY/AgRP system, which makes food intake-stimulating peptides, which can dramatically influence metabolism and consequently carcass traits (Seabury et al., 2017; Ruud et al., 2020). Among the genes more often represented in the glutamatergic synapse network enriched in our analysis, several were linked with food intake and metabolism (for example, *GRM8*), eating behavior (*GRIK3*), insulin secretion, and lipolysis (*ADCY1*, Olivieri et al., 2016). In support of this hypothesis, we also found out that the enriched KEGG term for DP Glutamatergic synapse (KEGG:04724) belonged to the same group of Circadian entrainments (KEGG:04713) and Apelin signaling pathway (KEGG:04371), both also enriched. Circadian rhythm has a strong connection with feeding behavior (Mrode et al., 2019), and apelin is a peptide connected with food intake and lipid metabolism (Bertrand et al., 2015). The same was true also for CF, with KEGG term Hippo signaling pathway (KEGG:04390) appearing multiple times (**Supplementary Figure S4H**). This might reflect a greater role of regulatory systems of feeding motivation, nutrient intake, and storage in shaping the variability of these traits. On the other hand, glutamatergic synapses are also involved in physiological responses to stressors and environmental changes. QTLs from the QTLdb associated to our candidate regions for these two traits are reported in **Supplementary Tables S2C,D**.

3.4 Traits and Time Stratification

The results of our study can help frame the genetic architecture of our between-traits correlation, including such traits that are measures of the same trait in different time points or intervals (the three BW and the three ADG). Within BW, we demonstrated how also from the genomic point of view the weight at the half of the PT was underlined by a mixture of QTLs that were also found either at the start or at the end of the PT. On the other hand, no common SNPs resulted significant both for BW_i and BW_f, and the number of enriched pathways in common was very low (**Supplementary Figures S4A–C**; **Table 3**). For what concerns ADG, there was also a deep difference between the signals found for ADG_i and ADG_f, with the latter reflecting much more closely the total ADG, and again no SNPs were shared by ADG_i and ADG_f (**Table 3**). Moreover, the lowest number of significant SNPs and pathways for BW was at BW_m, and for ADG was ADG_i, with these two traits sharing a temporal correspondence.

Interestingly, we found many genes in common between measures of different traits taken at the same time. For example, both SNPs on BTA7 and BTA1 were significant both for BW_m and ADG_i. Also, one SNP on BTA26 was suggestively associated both for BW_f and ADG_f (**Table 3**).

These results have several implications: firstly, from an economic point of view, they show that the timing of the trait measurement is crucial. Different life stages can result in different genetic signals; if used for a selection program, this can have an economic and conservation impact. While this is of course expected, given the succession of different biological processes during development, very few studies include such a time stratification in their analysis of productive traits. Even if such a process is difficult to infer, our results show that complexity—intended as the number of functional elements, their diversity, and pathways involved—might increase with age.

4 CONCLUSION AND IMPLICATIONS FOR LOCAL BREEDS

There are four main takeaways that could be extracted from our study. Firstly, our analysis detected a significant signal for body weight (recorded when bulls were 1 month old) on BTA9; a significant signal of average daily gain (recorded at 7 months of age) on BTA1 and three significant signals of average daily gain (recorded at 1 year of age) on BTA10. Three significant signals for carcass traits (one signal each for dressing percentage and carcass fleshiness, plus one in common between the two) were all situated on BTA18.

Secondly, the variety of GO terms and functional elements involved in the beef-related traits under study was staggering. We could detect in multiple traits key roles of pathways related to actin, lipid transport, and several types of channels. Moreover, our analysis detected—alongside many genes often found in relation to the investigated traits—multiple pathways, genes, and functional elements of unclear attribution, for example with links to early development and maternal effect (such as *TBX18, NLRP2, SLCA12*), or to pathogen resistance (*MBL2*). This issue underlines how even research of well-studied traits can turn out unexpected results, especially if performed in rarely investigated breeds. In additions, the fact that Rendena has been bred not only for the considered traits, but also for antagonistic could have added a layer of complexity to our results.

Thirdly, we detected for almost all traits several pathways and genes linked with neuroblast development and synaptic transmission, especially (but not exclusively) glutamatergic, which added to the intricacy of the gene networks involved in these traits. Pathways linked to both neuroblast proliferation and synaptic communication have been tied in recent years to selection for environmental condition (Rowan et al., 2020) differences in behavioral temperament (Gutiérrez-Gil et al., 2008) and adaptability (Taye et al., 2017).

Finally, as discussed above, we found that even when focusing on widely investigated traits the influence of time stratification was fundamental. We argue that future studies on this issue

should include an analysis of time stratification of their trait to fully report their complexity during development.

A greater diffusion of adaptable and diversified local breeds, with characteristics allowing for lower environmental impact, better survival and greater production in challenging environments might be crucial in staving off the negative effects of intensive beef farming. To achieve this, however, there is urgent need for further studies of the genetic basis of productive and life-history trait, which are still lacking. Moreover, these studies could help uncovering several novel gene networks associations and pathways, thanks to the less intensive selection for production occurring in local breed. Finally, they would help to map the diversity of such breeds, in an unvaluable help for their conservation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because this study did not require any specific ethics permit. The cattle sampled belonged to commercial private herds and were not experimentally manipulated. Samples were collected by technicians from the Breeders Association of Rendena.

AUTHOR CONTRIBUTIONS

Conceptualization, RM, BT, and EM; methodology, BT and EM; formal analysis, EM and BT; support to analysis SP, investigation, SP, BT, CS, RM, and EM; resources, RM; data curation, EM and RM; writing original draft preparation, EM and BT writing—review and editing SP, CS, and RM. All authors have read and agreed to the published version of the manuscript.

FUNDING

The study was funded by the DUALBREEDING project (CUP J61J18000030005) and by BIRD183281.

ACKNOWLEDGMENTS

Authors are grateful to National Breeders Association of Rendena cattle breed (ANARE) for data support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.746665/full#supplementary-material

REFERENCES

Abo-Ismail, M. K., Lansink, N., Akanno, E., Karisa, B. K., Crowley, J. J., Moore, S. S., et al. (2018). Development and Validation of a Small SNP Panel for Feed Efficiency in Beef Cattle1. *J. Anim. Sci.* 96, 375–397. doi:10.1093/jas/sky020

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., and Misztal, I. (2019). Frequentist P-Values for Large-Scale-Single Step Genome-wide Association, with an Application to Birth Weight in American Angus Cattle. *Genet. Sel. Evol.* 51, 1–8. doi:10.1186/s12711-019-0469-3

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot Topic: A Unified Approach to Utilize Phenotypic, Full Pedigree, and Genomic Information for Genetic Evaluation of Holstein Final Score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A., and Misztal, I. (2018). BLUPF90 Suite of Programs for Animal Breeding. 11th World Congr. *Genet. Appl. Livest. Prod.* 11, 751.

Albera, A., Mantovani, R., Bittante, G., Groen, A. F., and Carnier, P. (2001). Genetic Parameters for Daily Live-Weight Gain, Live Fleshiness and Bone Thinness in Station-Tested Piemontese Young Bulls. *Anim. Sci.* 72, 449–456. doi:10.1017/S1357729800051961

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide Association Study of 107 Phenotypes in *Arabidopsis thaliana* Inbred Lines. *Nature* 465, 627–631. doi:10.1038/nature08800

Begum, F., Ghosh, D., Tseng, G. C., and Feingold, E. (2012). Comprehensive Literature Review and Statistical Considerations for GWAS Meta-Analysis. *Nucleic Acids Res.* 40, 3777–3784. doi:10.1093/nar/gkr1255

Bertrand, C., Valet, P., and Castan-Laurell, I. (2015). Apelin and Energy Metabolism. *Front. Physiol.* 6, 1–5. doi:10.3389/fphys.2015.00115

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., and Gandini, G. (2015). Challenges and Opportunities in Genetic Improvement of Local Livestock Breeds. *Front. Genet.* 6, 1–16. doi:10.3389/fgene.2015.00033

Buitenhuis, B., Janss, L. L., Poulsen, N. A., Larsen, L. B., Larsen, M. K., and Sørensen, P. (2014). Genome-wide Association and Biological Pathway Analysis for Milk-Fat Composition in Danish Holstein and Danish Jersey Cattle. *BMC Genomics* 15, 1112. doi:10.1186/1471-2164-15-1112

Carvalho, F. E., Espigolan, R., Berton, M. P., Neto, J. B. S., Silva, R. P., Grigoletto, L., et al. (2020). Genome-wide Association Study and Predictive Ability for Growth Traits in Nellore Cattle. *Livestock Sci.* 231, 103861. doi:10.1016/j.livsci.2019.103861

Chen, Q., Huang, B., Zhan, J., Wang, J., Qu, K., Zhang, F., et al. (2020). Whole-genome Analyses Identify Loci and Selective Signals Associated with Body Size in Cattle. *J. Anim. Sci.* 98 (3), skaa068. doi:10.1093/jas/skaa068

Cheverud, J. M. (2001). A Simple Correction for Multiple Comparisons in Interval Mapping Genome Scans. *Heredity* 87, 52–58. doi:10.1046/j.1365-2540.2001.00901.x

Cohen-Zinder, M., Asher, A., Lipkin, E., Feingersch, R., Agmon, R., Karasik, D., et al. (2016). FABP4is a Leading Candidate Gene Associated with Residual Feed Intake in Growing Holstein Calves. *Physiol. Genomics* 48, 367–376. doi:10.1152/physiolgenomics.00121.2015

Consortium, T. U. (2021). UniProt: the Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi:10.1093/nar/gkaa1100

De Las Heras-Saldana, S., Chung, K. Y., Lee, S. H., and Gondro, C. (2019). Gene Expression of Hanwoo Satellite Cell Differentiation in Longissimus Dorsi and Semimembranosus. *BMC Genomics* 20, 1–15. doi:10.1186/s12864-019-5530-7

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Evans, R. D., Walsh, S. W., et al. (2020b). Genomic Regions Associated with Muscularity in Beef Cattle Differ in Five Contrasting Cattle Breeds. *Genet. Sel Evol.* 52 (1), 2. doi:10.1186/s12711-020-0523-1

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Walsh, S. W., Evans, R. D., et al. (2020a). Genomic Regions Associated with Skeletal Type Traits in Beef and Dairy Cattle Are Common to Regions Associated with Carcass Traits, Feed Intake and Calving Difficulty. *Front. Genet.* 11. doi:10.3389/fgene.2020.00020

Drost, H.-G., and Paszkowski, J. (2017). Biomartr: Genomic Data Retrieval with R. *Bioinformatics* 33, btw821–1217. doi:10.1093/bioinformatics/btw821

Edea, Z., Jung, K. S., Shin, S.-S., Yoo, S.-W., Choi, J. W., and Kim, K.-S. (2020). Signatures of Positive Selection Underlying Beef Production Traits in Korean Cattle Breeds. *J. Anim. Sci. Technol.* 62, 293–305. doi:10.5187/JAST.2020.62.3.293

Ejarque, M., Ceperuelo-Mallafré, V., Serena, C., Maymo-Masip, E., Duran, X., Díaz-Ramos, A., et al. (2019). Adipose Tissue Mitochondrial Dysfunction in Human Obesity Is Linked to a Specific DNA Methylation Signature in Adipose-Derived Stem Cells. *Int. J. Obes.* 43, 1256–1268. doi:10.1038/s41366-018-0219-6

Engström Ruud, L., Pereira, M. M. A., de Solis, A. J., Fenselau, H., and Brüning, J. C. (2020). NPY Mediates the Rapid Feeding and Glucose Metabolism Regulatory Functions of AgRP Neurons. *Nat. Commun.* 11 (1), 1–12. doi:10.1038/s41467-020-14291-3

Fabbri, M. C., Dadousis, C., and Bozzi, R. (2020). Estimation of Linkage Disequilibrium and Effective Population Size in Three Italian Autochthonous Beef Breeds. *Animals* 10 (6), 1034. doi:10.3390/ani10061034

Falker-Gieske, C., Blaj, I., Preuß, S., Bennewitz, J., Thaller, G., and Tetens, J. (2019). GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3 Genes, Genomes, Genet.* 9, 2823–2834. doi:10.1534/g3.119.400452

Filipčík, R., Falta, D., Kopec, T., Chládek, G., Večeřa, M., and Rečková, Z. (2020). Environmental Factors and Genetic Parameters of Beef Traits in Fleckvieh Cattle Using Field and Station Testing. *Animals* 10 (11), 2159. doi:10.3390/ani10112159

Fonseca, P. A. S., Suárez-Vega, A., Marras, G., and Cánovas, Á. (2020). GALLO: An R Package for Genomic Annotation and Integration of Multiple Data Sources in Livestock for Positional Candidate Loci. *GigaScience* 9 (12). doi:10.1093/gigascience/giaa149

Frigo, E., Samorè, A. B., Vicario, D., Bagnato, A., and Pedron, O. (2013). Heritabilities and Genetic Correlations of Body Condition Score and Muscularity with Productive Traits and Their Trend Functions in Italian Simmental Cattle. *Ital. J. Anim. Sci.* 12, e40–246. doi:10.4081/ijas.2013.e40

Gast, M.-T., Tönjes, A., Keller, M., Horstmann, A., Steinle, N., Scholz, M., et al. (2013). The Role of Rs2237781 withinGRM8in Eating Behavior. *Brain Behav.* 3 (5), 495–502. doi:10.1002/brb3.151

Gershoni, M., Weller, J. I., and Ezra, E. (2021). *Genetic and Genome-wide Association Analysis of Yearling Weight Gain in Israel Holstein Dairy Calves*.

Gilmour, A. R., Robin, T., and Cullis, B. R. (1995). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics* 51 (4), 1440–1450. doi:10.2307/2533274

Grigoletto, L., Ferraz, J. B. S., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical Composite Beef Cattle. *Front. Genet.* 11, 1–13. doi:10.3389/fgene.2020.00123

Gualdrón Duarte, J. L., Cantet, R. J., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid Screening for Phenotype-Genotype Associations by Linear Transformations of Genomic Evaluations. *BMC Bioinformatics* 15, 1–11. doi:10.1186/1471-2105-15-246

Gutiérrez-Gil, B., Ball, N., Burton, D., Haskell, M., Williams, J. L., and Wiener, P. (2008). Identification of Quantitative Trait Loci Affecting Cattle Temperament. *J. Hered.* 99, 629–638. doi:10.1093/jhered/esn060

Guzzo, N., Sartori, C., and Mantovani, R. (2019). Analysis of Genetic Correlations between Beef Traits in Young Bulls and Primiparous Cows Belonging to the Dual-Purpose Rendena Breed. *animal* 13, 694–701. doi:10.1017/S1751731118001969

Guzzo, N., Sartori, C., and Mantovani, R. (2018). Heterogeneity of Variance for Milk, Fat and Protein Yield in Small Cattle Populations: The Rendena Breed as a Case Study. *Livestock Sci.* 213, 54–60. doi:10.1016/j.livsci.2018.05.002

Hardie, L. C., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., Armentano, L. E., Wiggans, G. R., et al. (2017). The Genetic and Biological Basis of Feed Efficiency in Mid-lactation Holstein Dairy Cows. *J. Dairy Sci.* 100, 9061–9075. doi:10.3168/jds.2017-12604

Helgeland, Ø., Vaudel, M., Juliusson, P. B., Lingaas Holmen, O., Juodakis, J., Bacelis, J., et al. (2019). Genome-wide Association Study Reveals Dynamic Role of Genetic Variation in Infant and Early Childhood Growth. *Nat. Commun.* 10. doi:10.1038/s41467-019-12308-0

Hill, W. G., and Robertson, A. (1968). Linkage Disequilibrium in Finite Populations. *Theoret. Appl. Genet.* 38, 226–231. doi:10.1007/BF01245622

Hirano, T., Matsuhashi, T., Kobayashi, N., Watanabe, T., and Sugimoto, Y. (2012). Identification of an FBN1 Mutation in Bovine Marfan Syndrome-like Disease. *Anim. Genet.* 43, 11–17. doi:10.1111/j.1365-2052.2011.02209.x

Hocquette, J. F., Ortigues-Marty, I., Pethick, D., Herpin, P., and Fernandez, X. (1998). Nutritional and Hormonal Regulation of Energy Metabolism in Skeletal Muscles of Meat-Producing Animals. *Livestock Prod. Sci.* 56, 115–143. doi:10.1016/S0301-6226(98)00187-0

Huang, S., He, Y., Ye, S., Wang, J., Yuan, X., Zhang, H., et al. (2018). Genome-wide Association Study on Chicken Carcass Traits Using Sequence Data Imputed from SNP Array. *J. Appl. Genet.* 59, 335–344. doi:10.1007/s13353-018-0448-3

Huang, W., Guo, Y., Du, W., Zhang, X., Li, A., and Miao, X. (2017). Global Transcriptome Analysis Identifies Differentially Expressed Genes Related to Lipid Metabolism in Wagyu and Holstein Cattle. *Sci. Rep.* 7, 1–11. doi:10.1038/s41598-017-05702-5

Illius, A. W., Tolkamp, B. J., and Yearsley, J. (2002). The Evolution of the Control of Food Intake. *Proc. Nutr. Soc.* 61, 465–472. doi:10.1079/pns2002179

Júnior, G. A. F., Costa, R. B., De Camargo, G. M. F., Carvalheiro, R., Rosa, G. J. M., Baldi, F., et al. (2016). Genome Scan for Postmortem Carcass Traits in Nellore Cattle1. *J. Anim. Sci.* 94, 4087–4095. doi:10.2527/jas.2016-0632

Kaur, M., Kumar, A., Siddaraju, N. K., Fairoze, M. N., Chhabra, P., Ahlawat, S., et al. (2020). Differential Expression of miRNAs in Skeletal Muscles of Indian Sheep with Diverse Carcass and Muscle Traits. *Sci. Rep.* 10, 1–11. doi:10.1038/s41598-020-73071-7

Kemter, E., Rathkolb, B., Becker, L., Bolle, I., Busch, D. H., Dalke, C., et al. (2014). Standardized, Systemic Phenotypic Analysis of Slc12a1 I299F Mutant Mice. *J. Biomed. Sci.* 21 (1), 1–10. doi:10.1186/s12929-014-0068-0

Kern, R. J., Lindholm-Perry, A. K., Freetly, H. C., Kuehn, L. A., Rule, D. C., and Ludden, P. A. (2016). Rumen Papillae Morphology of Beef Steers Relative to Gain and Feed Intake and the Association of Volatile Fatty Acids with Kallikrein Gene Expression. *Livestock Sci.* 187, 24–30. doi:10.1016/j.livsci.2016.02.007

Kosińska-Selbi, B., Suchocki, T., Egger-Danner, C., Schwarzenbacher, H., Frąszczak, M., and Szyda, J. (2020). Exploring the Potential Genetic Heterogeneity in the Incidence of Hoof Disorders in Austrian Fleckvieh and Braunvieh Cattle. *Front. Genet.* 11, 1423. doi:10.3389/fgene.2020.577116

Kracht, M., Müller-Ladner, U., and Schmitz, M. L. (2020). Mutual Regulation of Metabolic Processes and Proinflammatory NF-Kb Signaling. *J. Allergy Clin. Immunol.* 146 (4), 694–705. doi:10.1016/j.jaci.2020.07.027

Krupová, Z., Krupa, E., Michaličková, M., Wolfová, M., and Kasarda, R. (2016). Economic Values for Health and Feed Efficiency Traits of Dual-Purpose Cattle in Marginal Areas. *J. Dairy Sci.* 99, 644–656. doi:10.3168/jds.2015-9951

Lapierre, H., Pelletier, G., Abribat, T., Fournier, K., Gaudreau, P., Brazeau, P., et al. (1995). The Effect of Feed Intake and Growth Hormone-Releasing Factor on Lactating Dairy Cows. *J. Dairy Sci.* 78, 804–815. doi:10.3168/jds.S0022-0302(95)76692-9

Lee, K. Y., Singh, M. K., Ussar, S., Wetzel, P., Hirshman, M. F., Goodyear, L. J., et al. (2015). Tbx15 Controls Skeletal Muscle Fibre-type Determination and Muscle Metabolism. *Nat. Commun.* 6. doi:10.1038/ncomms9054

Lemos, M. V. A., Chiaia, H. L. J., Berton, M. P., Feitosa, F. L. B., Aboujaoud, C., Camargo, G. M. F., et al. (2016). Genome-wide Association between Single Nucleotide Polymorphisms with Beef Fatty Acid Profile in Nellore Cattle Using the Single Step Procedure. *BMC Genomics* 17, 1–16. doi:10.1186/s12864-016-2511-y

Li, J., and Ji, L. (2005). Adjusting Multiple Testing in Multilocus Analyses Using the Eigenvalues of a Correlation Matrix. *Heredity* 95, 221–227. doi:10.1038/sj.hdy.6800717

Liao, S. F., Vanzant, E. S., Harmon, D. L., McLeod, K. R., Boling, J. A., and Matthews, J. C. (2009). Ruminal and Abomasal Starch Hydrolysate Infusions Selectively Decrease the Expression of Cationic Amino Acid Transporter mRNA by Small Intestinal Epithelia of Forage-Fed Beef Steers. *J. Dairy Sci.* 92, 1124–1135. doi:10.3168/jds.2008-1521

Liu, Y., Xu, L., Wang, Z., Xu, L., Chen, Y., Zhang, L., et al. (2019). Genomic Prediction and Association Analysis with Models Including Dominance Effects

for Important Traits in Chinese Simmental Beef Cattle. *Animals* 9, 1055. doi:10.3390/ani9121055

Livernois, A. M., Mallard, B. A., Cartwright, S. L., and Cánovas, A. (2021). Heat Stress and Immune Response Phenotype Affect DNA Methylation in Blood Mononuclear Cells from Holstein Dairy Cows. *Sci. Rep.* 11, 11371. doi:10.1038/s41598-021-89951-5

Londoño-Gil, M., Rincón Flórez, J. C., Lopez-Herrera, A., and Gonzalez-Herrera, L. G. (2021). Genome-Wide Association Study for Growth Traits in Blanco Orejinero (Bon) Cattle from Colombia. *Livestock Sci.* 243, 104366–104369. doi:10.1016/j.livsci.2020.104366

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., and Misztal, I. (2020). Single-Step Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in BLUPF90. *Genes* 11, 790. doi:10.3390/genes11070790

Mancin, E., Lourenco, D., Bermann, M., Mantovani, R., and Misztal, I. (2021b). Accounting for Population Structure and Phenotypes from Relatives in Association Mapping for Farm Animals: A Simulation Study. doi:10.3389/fgene.2021.642065

Mancin, E., Sartori, C., Guzzo, N., Tuliozi, B., and Mantovani, R. (2021c). Selection Response Due to Different Combination of Antagonistic Milk, Beef, and Morphological Traits in the Alpine Grey Cattle Breed. *Animals* 11, 1340. doi:10.3390/ani11051340

Mancin, E., Tuliozi, B., Sartori, C., Guzzo, N., and Mantovani, R. (2021a). Genomic Prediction in Local Breeds: The Rendena Cattle as a Case Study. *Animals* 11, 1815. doi:10.3390/ani11061815

Mantovani, R., Gallo, L., Carnier, P., Cassandro, M., and Bittante, G. (19971997). Vienna, Austria. The Use of a Juvenile Selection Scheme for Genetic Improvement of Small Populations: the Example of Rendena Breed, *Proc. 48th EAAP Annu. Meet.* 25–28 .

Mao, X., Sahana, G., De Koning, D.-J., and Guldbrandtsen, B. (2016). Genome-wide Association Studies of Growth Traits in Three Dairy Cattle Breeds Using Whole-Genome Sequence Data1. *J. Anim. Sci.* 94, 1426–1437. doi:10.2527/jas.2015-9838

Martinez, J. (2000). Body-weight Regulation: Causes of Obesity. *Proc. Nutr. Soc.* 59 (3), 337–345. doi:10.1017/S0029665100000380

Marshall, K., Gibson, J. P., Mwai, O., Mwacharo, J. M., Haile, A., Getachew, T., et al. (2019). Livestock Genomics for Developing Countries - African Examples in Practice. *Front. Genet.* 10, 297. doi:10.3389/fgene.2019.00297

Mateescu, R. G., Garrick, D. J., and Reecy, J. M. (2017). Network Analysis Reveals Putative Genes Affecting Meat Quality in Angus Cattle. 8. doi:10.3389/fgene.2017.00171

Mazza, S., Guzzo, N., Sartori, C., and Mantovani, R. (2016). Genetic Correlations between Type and Test-Day Milk Yield in Small Dual-Purpose Cattle Populations: The Aosta Red Pied Breed as a Case Study. *J. Dairy Sci.* 99, 8127–8136. doi:10.3168/jds.2016-11116

Medeiros de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., Miguel Ferreira de Camargo, G., Matos Ceacero, T., Noely dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. *PLoS One* 12, e0169860–14. doi:10.1371/journal.pone.0169860

Miller, G. D. (2017). Appetite Regulation: Hormones, Peptides, and Neurotransmitters and Their Role in Obesity. *Am. J. lifestyle Med.* 13 (6), 586–601. doi:10.1177/1559827617716376

Mrode, R., Ojango, J. M. K., Okeyo, A. M., and Mwacharo, J. M. (2019). Genomic Selection and Use of Molecular Tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects. *Front. Genet.* 9. doi:10.3389/fgene.2018.00694

Mudadu, M. A., Porto-Neto, L. R., Mokry, F. B., Tizioto, P. C., Oliveira, P. S. N., Tullio, R. R., et al. (2016). Genomic Structure and Marker-Derived Gene Networks for Growth and Meat Quality Traits of Brazilian Nelore Beef Cattle. *BMC Genomics* 17, 1–16. doi:10.1186/s12864-016-2535-3

Olivieri, B. F., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Branco, R. H., Bonilha, S. F. M., De Albuquerque, L. G., et al. (2016). Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population. *PLoS One* 11, e0164390. doi:10.1371/journal.pone.0164390

Ovaska, U., and Soini, K. (2017). Local Breeds - Rural Heritage or New Market Opportunities? Colliding Views on the Conservation and Sustainable Use of Landraces. *Sociologia Ruralis* 57, 709–729. doi:10.1111/soru.12140

Pegolo, S., Momen, M., Morota, G., Rosa, G. J. M., Gianola, D., Bittante, G., et al. (2020). Structural Equation Modeling for Investigating Multi-Trait Genetic Architecture of Udder Health in Dairy Cattle. *Sci. Rep.* 10, 7751. doi:10.1038/s41598-020-64575-3

Peng, H., Chang, B., Lu, C., Su, J., Wu, Y., Lv, P., et al. (2012). Nlrp2, a Maternal Effect Gene Required for Early Embryonic Development in the Mouse. *PLoS One* 7, e30344. doi:10.1371/journal.pone.0030344

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Purfield, D. C., Evans, R. D., and Berry, D. P. (2020). Breed- and Trait-specific Associations Define the Genetic Architecture of Calving Performance Traits in Cattle. *J. Anim. Sci.* 98, 1–18. doi:10.1093/JAS/SKAA151

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rowan, T. N., Durbin, H. J., Seabury, C. M., Schnabel, R. D., and Decker, J. E. (2020). Powerful Detection of Polygenic Selection and Evidence of Environmental Adaptation in US Beef Cattle. doi:10.1101/2020.03.11.988121

Samorè, A. B., Fontanesi, L., Fontanesi, L., and Samore, A. B. (2016). *Genomic Selection in Pigs: State of the Art and Perspectives*, 15, 211–232. doi:10.1080/1828051X.2016.1172034Genomic Selection in Pigs: State of the Art and Perspectives*Ital. J. Anim. Sci.*

Sartori, C., Guzzo, N., Mazza, S., and Mantovani, R. (2018). Genetic Correlations Among Milk Yield, Morphology, Performance Test Traits and Somatic Cells in Dual-Purpose Rendena Breed. *Animal* 12, 906–914. doi:10.1017/S1751731117002543

Sayols, S. (2020). Rrvgo: a Bioconductor Package to Reduce and Visualize Gene Ontology Terms. https://ssayols.github.io/rrvgo.

Sbarra, F., Mantovani, R., Quaglia, A., and Bittante, G. (2013). Genetics of slaughter Precocity, Carcass Weight, and Carcass Weight Gain in Chianina, Marchigiana, and Romagnola Young Bulls under Protected Geographical Indication1. *J. Anim. Sci.* 91, 2596–2604. doi:10.2527/jas.2013-6235

Schmid, M., and Bennewitz, J. (2017). Invited Review: Genome-wide Association Analysis for Quantitative Traits in Livestock - a Selective Review of Statistical Models and Experimental Designs. *Arch. Anim. Breed.* 60, 335–346. doi:10.5194/aab-60-335-2017

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide Association Study for Feed Efficiency and Growth Traits in U.S. Beef Cattle. *BMC Genomics* 18, 1–25. doi:10.1186/s12864-017-3754-y

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., Ciampolini, R., et al. (2020). The Genetic Heritage of Alpine Local Cattle Breeds Using Genomic SNP Data. *Genet. Sel. Evol.* 52, 1–12. doi:10.1186/s12711-020-00559-1

Srivastava, S., Srikanth, K., Won, S., Son, J.-H., Park, J.-E., Park, W., et al. (2020). Haplotype-Based Genome-wide Association Study and Identification of Candidate Genes Associated with Carcass Traits in Hanwoo Cattle. *Genes* 11, 551. doi:10.3390/genes11050551

Sun, W., Zhao, X., Wang, Z., Chu, Y., Mao, L., Lin, S., et al. (2019). Tbx15 Is Required for Adipocyte browning Induced by Adrenergic Signaling Pathway. *Mol. Metab.* 28, 48–57. doi:10.1016/j.molmet.2019.07.004

Sutera, A. M., Moscarelli, A., Mastrangelo, S., Sardina, M. T., Di Gerlando, R., Portolano, B., et al. (2021). Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. *Front. Genet.* 12, 409. doi:10.3389/fgene.2021.643531

Taye, M., Lee, W., Jeon, S., Yoon, J., Dessie, T., Hanotte, O., et al. (2017). Exploring Evidence of Positive Selection Signatures in Cattle Breeds Selected for Different Traits. *Mamm. Genome* 28, 528–541. doi:10.1007/s00335-017-9715-6

Tiezzi, F., and Maltecca, C. (2015). Accounting for Trait Architecture in Genomic Predictions of US Holstein Cattle Using a Weighted Realized Relationship Matrix. *Genet. Sel. Evol.* 47, 24. doi:10.1186/s12711-015-0100-1

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., and König, S. (2020). A Multi-Breed GWAS for Morphometric Traits in Four Beninese Indigenous Cattle Breeds Reveals Loci Associated with Conformation, Carcass and Adaptive Traits. *BMC Genomics* 21, 783. doi:10.1186/s12864-020-07170-0

Veselá, Z., Vostrý, L., and Šafus, P. (2011). Linear and Linear-Threshold Model for Genetic Parameters for SEUROP Carcass Traits in Czech Beef Cattle. *Czech J. Anim. Sci.* 56, 414–426. doi:10.17221/1292-cjas

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in Genomic Predictions for Populations under Selection. *Genet. Res.* 93, 357–366. doi:10.1017/S001667231100022X

Whalen, A., and Hickey, J. M. (2020). AlphaImpute2: Fast and Accurate Pedigree and Population Based Imputation for Hundreds of Thousands of Individuals in Livestock Populations. *bioRxiv*. doi:10.1101/2020.09.16.299677

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New Tork. Retrieved from https://ggplot2.tidyverse.org.

Wiggans, G. R., Cooper, T. A., VanRaden, P. M., Olson, K. M., and Tooker, M. E. (2012). Use of the Illumina Bovine3K BeadChip in Dairy Genomic Evaluation. *J. Dairy Sci.* 95 (3), 1552–1558. doi:10.3168/jds.2011-4985

Yang, Y., Wang, H., Li, G., Liu, Y., Wang, C., and He, D. (2020). Exploring the Genetic Basis of Fatty Liver Development in Geese. *Sci. Rep.* 10, 1–12. doi:10.1038/s41598-020-71210-8

Yin, T., and König, S. (2018). Genetic Parameters for Body Weight from Birth to Calving and Associations between Weights with Test-Day, Health, and Female Fertility Traits. *J. Dairy Sci.* 101, 2158–2170. doi:10.3168/jds.2017-13835

Yin, T., and König, S. (2019). Genome-wide Associations and Detection of Potential Candidate Genes for Direct Genetic and Maternal Genetic Effects Influencing Dairy Cattle Body Weight at Different Ages. *Genet. Sel. Evol.* 51, 1–14. doi:10.1186/s12711-018-0444-4

Zhang, J., Tan, J., Zhang, C., Wang, Y., Chen, X., Lei, C., et al. (2021). Research on Associations between Variants and Haplotypes of Aquaporin 9 (AQP9) Gene with Growth Traits in Three Cattle Breeds. *Anim. Biotechnol.* 32, 185–193. doi:10.1080/10495398.2019.1675681

Zhuang, Z., Xu, L., Yang, J., Gao, H., Zhang, L., Gao, X., et al. (2020). Weighted Single-step Genome-wide Association Study for Growth Traits in Chinese Simmental Beef Cattle. *Genes* 11, 189. doi:10.3390/genes11020189

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 February 2022
doi: 10.3389/fgene.2021.730599

Identification of Internal Reference Genes in Peripheral Blood Mononuclear Cells of Cattle Populations Adapted to Hot Arid Normoxia and Cold Arid Hypoxia Environments

*Preeti Verma[1], Ankita Sharma[1], Monika Sodhi[1], Manish Tiwari[1,2], Prince Vivek[1], Ranjit S. Kataria[1], S. K. Nirajan[1], Vijay K. Bharti[3], Pawan Singh[2,4], S. S. Lathwal[2,4], Vishal Sharma[1], Nampher Masharing[1,2] and Manishi Mukesh[1]**

[1]Animal Biotechnology Division, ICAR-National Bureau of Animal Genetic Resources, Karnal, India, [2]Animal Biotechnology Center, ICAR-National Dairy Research Institute, Karnal, India, [3]DRDO-Defense Institute of High-altitude Research, Leh, India, [4]Department of Livestock Production Management, ICAR-National Dairy Research Institute, Karnal, India

OPEN ACCESS

Edited by:
Mohammed Ali Al Abri, Sultan Qaboos University, Oman

Reviewed by:
Muhammad Ihsan Andi Dagong, Hasanuddin University, Indonesia
Sajad Ahmad Wani, Ohio State University, United States

***Correspondence:**
Manishi Mukesh
mmukesh_26@hotmail.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *25 June 2021*
Accepted: *03 December 2021*
Published: *01 February 2022*

Citation:
Verma P, Sharma A, Sodhi M, Tiwari M, Vivek P, Kataria RS, Nirajan SK, Bharti VK, Singh P, Lathwal SS, Sharma V, Masharing N and Mukesh M (2022) Identification of Internal Reference Genes in Peripheral Blood Mononuclear Cells of Cattle Populations Adapted to Hot Arid Normoxia and Cold Arid Hypoxia Environments. Front. Genet. 12:730599. doi: 10.3389/fgene.2021.730599

To estimate gene expression in a reliable manner, quantitative real-time polymerase chain reaction data require normalisation using a panel of stably expressed reference genes (RGs). To date, information on an appropriate panel of RGs in cattle populations reared at cold arid high-altitude hypoxia and hot arid tropical normoxia environments is not available. Therefore, the present study was carried out to identify a panel of stably expressed RGs from 10 candidate genes (*GAPDH, RPL4, EEF1A1, RPS9, HPRT1, UXT, HMBS, B2M, RPS15,* and *ACTB*) in peripheral blood mononuclear cells (PBMCs) of cattle populations reared at cold arid high-altitude hypoxia and hot arid normoxia environments. Four different statistical algorithms: geNorm, NormFinder, BestKeeper, and RefFinder were used to assess the stability of these genes. A total of 30 blood samples were collected: six adult heifers each of Ladakhi (LAC) and Holstein Frisian crosses (HFX) and 4 Jersey (JYC) cows from cold arid high-altitude hypoxia environments (group I) and five adult heifers each of Sahiwal (SAC), Karan Fries (KFC), and Holstein Friesian (HFC) cows from hot arid normoxia environments (group II). Combined analysis of group I and group II resulted in identification of a panel of RGs like *RPS9, RPS15,* and *GAPDH* that could act as a useful resource to unravel the accurate transcriptional profile of PBMCs from diverse cattle populations adapted to distinct altitudes.

Keywords: reference genes, expression stability, qRT-PCR, normalization, hot arid, cold arid, cattle, PBMCs

INTRODUCTION

India has been blessed with several cattle breeds of *Bos indicus* lineage adapted to various agro-climatic zones from high land to hot tropical regions. Leh-Ladakh, also known as a "COLD DESERT" is a part of the western Himalayan agro-climatic and high-altitude temperate sub-agro-climatic zone in India. Ladakh is situated at an altitude of 3,500–5,500 m with difficult terrain and harsh climate

conditions such as extreme temperature (−40°C in winter and 35°C in summer), low humidity (25–40%), low precipitation (80–300 mm), and low oxygen level (nearly 60–70% of the oxygen concentration at sea level). In spite of harsh weather, this region is blessed with several unique livestock populations such as yak, cattle, horses, sheep, goat, donkey, and double-hump camel. Over thousands of years of the evolutionary process, these animals have developed the special ability to survive in cold and hypoxia environments prevalent in Ladakh. Amongst all the livestock species, the native cattle of Ladakh known as "Ladakhi cattle" are the major livestock species that play an important role in the agriculture economy of the region. The cattle from the Trans-Himalayan region of Ladakh are short in stature and well-adapted to the high-altitude environment. Based on morphometric data on 275 animals and genetic characterization using microsatellite markers (unpublished data), this cattle population was observed to be highly distinct from native cattle breeds adapted to other agroclimatic zones of India. Recently, our group was successful in delineating the distinct transcriptome signatures of peripheral blood mononuclear cells (PBMCs) of high-altitude-adapted Ladakhi cattle and tropically adapted Sahiwal cattle (Verma et al., 2018a; Verma et al., 2018b). In the last few years, the purity of Ladakhi cattle is believed to have declined due to widespread intermixing with Jersey cattle. However, considering the unique hypoxia-tolerant characteristics of Ladakhi cattle, preserving its purity will be a key for long-term conservation and sustainable utilization.

On the other hand, India has also been blessed with a huge native cattle genetic resource base which has adapted to hot and tropical conditions. For example, native cattle breeds like Sahiwal, Tharparkar, Rathi, Gir, Ongole etc. are known for their superior thermotolerance as compared to their exotic counterparts of the *Bos taurus* lineage. The superior cellular tolerance ability of PBMCs from native cattle has been shown in a few studies published by our group (Kishore et al., 2013; Kishore et al., 2014; Sharma et al., 2019) using PBMCs as the cellular model. However, native cattle breeds are also facing genetic dilution due to crossbreeding with exotic germplasm in order to enhance milk production. These native cattle populations that are adapted to distinct altitudes might have acquired a distinct gene pool during the course of the evolutionary process. Such genetic resources with remarkable adaptive traits could be an interesting resource to mine gene expression and the mechanism underlying changes associated with adaptation to cold arid and hot arid environments. It would be interesting to define the importance of various genes in conferring adaptation to cattle populations adapted to diverse altitudes through targeted gene expression analysis. Quantitative real-time polymerase chain reaction (qRT-PCR) has been widely employed to quantify the expression of target genes of interest in different tissues/cells exposed to a variety of experimental conditions. In spite of many of its advantages, this tool is prone to analytical variations arising due to differing amounts of starting material, pipetting errors, and differing efficiencies of RNA extraction and reverse transcription (Vandesompele et al., 2002; Huggett et al., 2005; Bustin et al., 2009). To overcome the limitations of experimental variation, the use of appropriate internal control genes (ICGs) or reference genes (RGs) to successfully normalise the RT-qPCR data has been reported in several studies (Bustin 2010; Castigliego et al., 2010; Crookenden et al., 2017; Die et al., 2017; Sang et al., 2018). The approach to identify a suitable panel of RGs during various experimental/physiological conditions has also been reported in different livestock species (Aggarwal et al., 2013; Kapila et al., 2013; Zhu et al., 2015; Jatav et al., 2016; Li et al., 2016; Kaur et al., 2018). However, to the best of our knowledge, no comparative data on suitable RGs are available for cattle populations reared at cold arid high-altitude and hot arid tropical regions. The present study was planned to identify a panel of stably expressed RGs in PBMCs of six cattle populations from the high-altitude cold arid region of Leh-Ladakh and hot arid tropical climate of India.

MATERIALS AND METHODS

Ethics Statement and Animal Selection

The blood sampling of animals was performed in accordance with the relevant guidelines and regulations as approved by the Institutional Animal Ethics Committee (IAEC) of ICAR-National Bureau of Animal Genetics Resources (NBAGR), Karnal. The study has included three cattle populations from the cold arid high-altitude region of Leh-Ladakh, *viz*., Ladakhi cattle (native), Jersey cattle (exotic), and Frieswal cattle {Sahiwal x Holstein Frisian cross}, and three cattle populations from the hot dry and semi-arid condition of Haryana state, *viz*., Sahiwal cattle (native), Holstein Frisians cattle (exotic), and Karan Fries cattle {Tharparkar x Holstein Frisian}. Ladakhi cattle (LAC) are the native breed of the *Bos indicus* lineage and are naturally adapted to the high-altitude region of Ladakh. Jersey cattle (JYC) are of the *Bos taurus* lineage and originated from temperate regions, while Frieswal cattle (HFX) are a cross-bred cattle population. Both Jersey and Frieswal cattle are non-native to Leh-Ladakh and have been reared in limited organized farms in the region since the last 2 decades. Amongst the cattle populations selected from the hot semi-arid region, Sahiwal cattle (SAC) are a very popular native cattle breed of *Bos indicus* lineage and are known for their adaptation potential to hot dry and tropical conditions. Holstein Frisian cattle (HFC) are of *Bos taurus* lineage and are non-native to India. However, Holstein Frisian cattle have been widely used in India in various cross-breeding programmes to enhance milk production of local breeds. Karan Fries cattle (KFC) are a popular cross-breed that were developed several decades ago by crossing Tharparkar cattle (native) with Holstein Frisian. Therefore the study has included two populations of native cattle: one from high-altitude (Ladakhi cattle; LAC) and the other from hot arid tropical regions (Sahiwal cattle; SAC); and two populations of cross-breeds and two populations of exotic cattle from two extreme altitudes. The geographical coordinates of the sampling site representing the hot arid climate were latitude– 29° 3′ 56.7828″ N, and longitude 76° 2′ 25.7892″ E. The geographical coordinates of the sampling site from the high-altitude region of Ladakh were latitude– 34° 9′ 9.3168″ N, and longitude 77° 34′ 37.3764″ E. About 7–8 ml of whole blood samples was collected aseptically from the external jugular vein of the animals in sterile EDTA-coated vacutainer tubes.

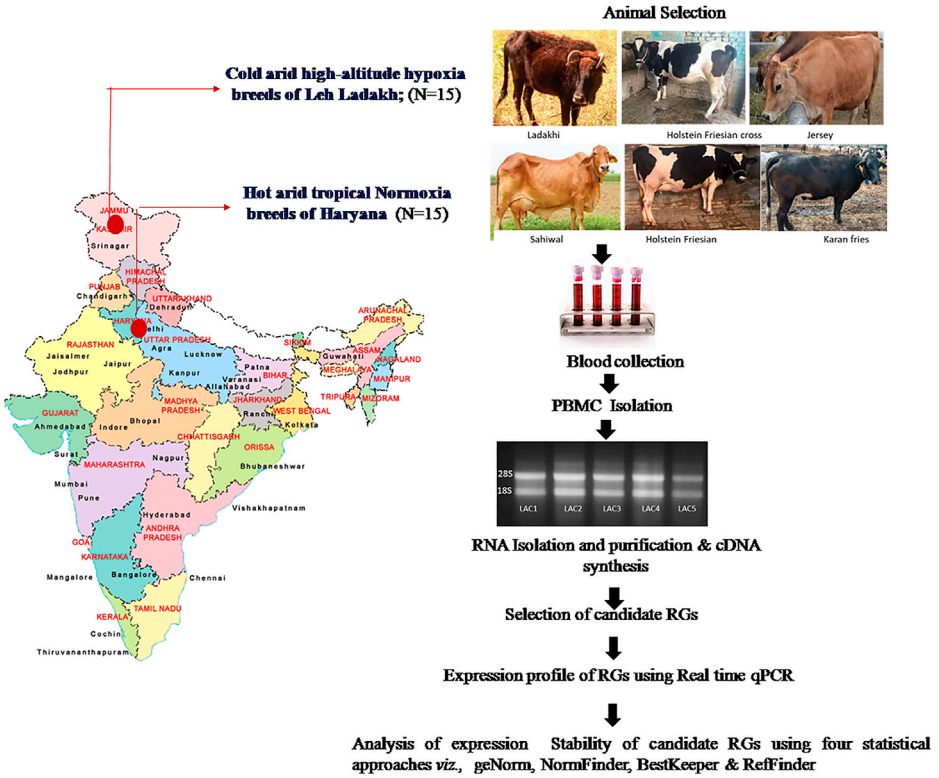

FIGURE 1 | The entire workflow of the present study conducted using PBMCs of cattle breeds from cold arid high-altitude hypoxia and hot arid normoxia groups. The map of India in **Figure 1** was created with the help of Smartdraw software- https://www.smartdraw.com).

In total, 30 blood samples were collected from adult heifers; five each of Sahiwal (SAC), Karan Fries (KFC), and Holstein Friesian (HFC) cows from the hot arid normoxia environment and 5 heifers each of Ladakhi (LAC), Holstein Frisian crosses (HFX), and Jersey (JYC) cows from the cold arid high-altitude hypoxia environment. The entire workflow of the qPCR experiment is depicted in **Figure 1**.

Peripheral Blood Mononuclear Cells Isolation, RNA Extraction, and cDNA Synthesis

Immediately after collection, the blood samples were processed to isolate peripheral blood mononuclear cells (PBMCs) using the density gradient centrifugation method as described by Verma et al. (2018b). In brief, blood was diluted in a 1:1 ratio with 1 × PBS (Ca^{2+} and Mg^{2+} free; Hyclone, Utah) and was gently over laid on a Histopaque-1077 (Sigma-Aldrich Inc., United States) followed by centrifugation at 4000 RPM for 30 min at RT. After removing the buffy coat in a separate 15 ml tube, cells were treated with 2 ml of chilled RBC lysis buffer for 5 min at RT and washed twice with 1 × PBS (Ca^{2+} and Mg^{2+} free; Hyclone, Utah). Finally, the cells were suspended in 1.0 ml of ice cold Trizol reagent (Invitrogen, Carlsbad, California), homogenized, and stored at −80°C. Total RNA was extracted using Trizol reagent according to the manufacturer's instructions (Invitrogen, Corp., CA, United States). The extracted RNA was further purified using RNeasy mini kit columns (Qiagen, Germany). To remove the traces of genomic DNA, rnase free DNase enzyme was used according to the manufacturer's instructions (Qiagen, Germany). RNA concentration and purity were checked using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, United States) and Experion Bio-analyzer (Bio-Rad, United States). The OD_{260}/OD_{280} absorption ratio for different samples varied from 1.92 to 2.10. The RNA integrity number (RIN) values for all purified RNA samples were in the satisfactory range (6.5–8.5). The RNA integrity of each sample was also confirmed by visualizing 28S and 18S ribosomal bands on 1.5% agarose gel. The cDNA was synthesized using a Revertaid First strand cDNA synthesis kit (Fermentas, United States) as described in our previous studies (Kishore et al., 2013; Verma et al., 2018b). Briefly, first strand cDNA was synthesized with 200 ng of purified RNA, oligo-dT (18) primer, dNTP mix, random primers, RiboLockRNase inhibitor, and M-MuLV reverse transcriptase supplied with RevertAid First Strand cDNA Synthesis (Thermo Scientific, CA, United States), using the program: 25°C for 5 min, 50°C for 60 min, and 70°C for 15 min. Before using them as templates for qPCR, each of the cDNA samples was diluted 1:4 (v:v) with DNase/RNase free water. The quality of 30 cDNAs was confirmed by amplifying the *GAPDH* gene using a similar protocol to that described for qPCR except for the final dissociation protocol. A small aliquot of amplified products for all the samples was run on 2.5% agarose gel to check the primer specificity and amplification quality.

TABLE 1 | Gene symbol, primer sequence, amplicon size, slope, PCR efficiency and (R^2) of RGs for each evaluated RG.

Gene Symbol	Accession no	Primer sequences	Annealing temp (°C)	Amplicon Size (bp)	Slope	PCR efficiency (%)	R^2
ACTB	NM_173979.3	F:5′GCGTGGCTACAGCTTCACC3′ R:3′TTGATGTCACGGACGATTTC5′	60	56	–3.14	108.20	0.996
GAPDH	NM_001034034.2	F:5′TGGAAAGGCCATCACCATCT3′ R:3′CCCACTTGATGTTGGCAG5′	60	60	–3.58	90.20	0.827
EEF1A1	NM_174535.2	F:5′CATCCCAGGCTGACTGTGC3′ R:3′TGTAAGCCAAAAGGGCATGC5′	60	101	–3.30	100.92	0.986
B2M	XM_002691119.4	F:5′CTGCTATGTGTATGGGTTCC3′ R:3′GGAGTGAACTCAGCGTG5′	60	101	–3.27	102.20	0.998
HMBS	NM_001046207.1	F:5′CTTTGGAGAGGAATGAAGTG3′ R:3′AATGGTGAAGCCAGGAGGA5′	60	80	–3.20	105.21	0.996
RPL4	NM_001014894.1	F:5′TTGGAAACATGTGTCGTGGG3′ R:3′GCAGATGGCGTATCGCTTCT5′	60	101	–3.32	101.2	0.922
RPS15	NM_001037443.2	F:5′GAATGGTGCGCATGAATGTC3′ R:3′GACTTTGGAGCACGGCCTAA5′	60	101	–3.54	91.6	0.989
RPS9	NM_001101152.2	F:5′CCTCGACCAAGAGCTGAAG3′ R:3′CCTCCAGACCTCACGTTTGTT5′	60	54	–3.34	99.25	0.941
UXT	NM_001037471.2	F:5′TGTGGCCCTTGGATATGGTT3′ R:3′GGTTGTCGCTGAGCTCTGTG5′	60	101	–3.22	104.4	0.997
HPRT11	NM_001034035.2	F:5′GAGAAGTCCGAGTTGAGTTT3′ R:3′GGCTCGTAGTGCAAATGAAG5′	60	101	–3.64	88.06	0.987
HIF1A	NM_174339.3	F:5′TGAAGGCACAGATGAATTGC3′ R:3′GTTCAAACTGAGTTAATCCC5′	60	129	–3.16	103	0.991
EPAS1	NM_174725.2	F:5′AGCAAGCCTTCCAAGACATGA3′ R: 3′GCTTGTCCGGCATCAAAGAG5′	60	90	–3.10	114	0.995
HSP70	JN604432.1	F:5′AACATGAAGAGCGCCGTGGAGG 3′ R:5′GTTACACACCTGCTCCAGCTCC3′	60	171	–2.90	120	0.990
HSP27	NM_001014911.1	F: 5′TACATTTCCCGTTGCTTCA3′ R: 3′GGACAGAGAGGAGGAGAC5′	60	78	–3.20	104	0.998

qPCR, efficiencies for each primer calculated pair-wise from a six-point standard curve using a five-fold dilution series of pooled DNA of Ladakhi and Sahiwal cow PBMCs
$R^{2:}$ *correlation coefficient of the slope of the standard curve.*

Selection of Candidate Reference Genes and Real-Time Quantitative PCR Primers

In the present study, 10 candidate genes belonging to different functional groups were selected for evaluation as suitable RGs (**Table 1**). The candidate genes included in the study were; glyceraldehyde 3-phosphate (*GAPDH*), ribosomal protein L-4 (*RPL4*), eukaryotic elongation factor 1 alpha (*EEF1A1*), ribosomal protein S9 (*RPS9*), hypoxanthine guanine phosphoribosyl transferase 1 (*HPRT1*), ubiquitin expressed transcript (*UXT*), hydroxyl methylbilane (*HMBS*), beta 2-microglobulin (*B2M*), ribosomal protein S15 (*RPS15*), and beta actin (*ACTB*). The primers specific for these 10 RGs were available in the laboratory and have been utilized successfully in several of our previous studies (Kapila et al., 2013; Jatav et al., 2016; Kaur et al., 2018). The information about sequences, amplicon length, and annealing temperature for each primer pair are summarized in **Table 1**.

Real-Time Quantitative PCR Reference Genes Transcripts

Quantitative PCR was performed in a 10 μL reaction volume containing 4 μl of diluted cDNA and 6 μl of master mix composed of 5 μL of 2X LightCycler 480 SYBR Green (Roche Life Science, Germany), 0.4 μL each of 10 μM forward and reverse primers, and 0.2 μL of DNase/RNase free water. Each qRT-PCR reaction was performed in duplicate to check the quality by assessing intra-assay variation. The amplification was carried out in a 96-well block using a LightCycler 480-II real-time PCR instrument (Roche Life Science, Germany) with the following conditions; 2 min at 50°C, 10 min at 95°C, 40 cycles of 15 s at 95°C (denaturation), and 1 min at 60°C (annealing + extension). In order to evaluate the quality of qPCR reactions in terms of nonspecific amplification and primer-dimer formation, a dissociation curve for each gene was obtained by increasing the temperature from 60°C to 95°C. A six-point relative standard curve was prepared for each gene by using five-fold serial dilutions of pooled cDNA samples in duplicate. The amplification specificity for each primer was checked by the presence of a single band of expected size on 2.5% agarose gel (**Supplementary Figure S1**), and also by observing the single melt curve peak after completion of qPCR (**Supplementary Figure S2**). The qPCR data for each gene were extracted using the "second derivative maximum" method (Rasmussen 2001) as computed by Light Cycler software 3.5 for subsequent analysis.

Analysis of Expression Stability of Candidate Reference Genes

The qPCR data recorded for each gene were subsequently analysed to evaluate the expression stability. Four statistical approaches, *viz*., geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004), BestKeeper (Xu et al.,.

2016), and the RefFinder web tool (Xie et al., 2012) were used to calculate the expression stability and ranking of RGs in the high-altitude hypoxia cold arid group (LAC, JYC, and HFX) and normoxia hot arid group (SAC, KFC, and HFC). The C_q values of each RG were exported to an Excel work sheet and modified as per the requirement of the software. Like for creating an input file for geNorm and NormFinder analysis, the C_q values were first transformed into relative quantities by using formula $2-^{\Delta CT}$, in which ΔC_q = corresponding C_q value - minimum C_q value. geNorm calculated the expression stability of individual genes on the basis of the M value which indicates the stability in expression of a gene. The genes with smaller M values (<1.5) are considered to have higher expression stability (Xu et al., 2016). In addition, geNorm was also used to conduct pair-wise variation analysis (Vn/Vn+1) in order to select the optimal number of RGs to normalise the target gene expression data. The cut off value of Vn/Vn+1 < 0.15 was used to decide the optimal number of RGs to be employed for calculating the normalisation factor (Vandesompele et al., 2002). This analysis is based on the principal that the expression ratio of the two best RGs will always remain similar across samples. The NormFinder software calculated the stability values of each RG based on inter- and intra-group variations. Similar to geNorm, for Normfinder analysis, the C_T values were first transformed into relative quantities. However, for Bestkeeper analysis, C_T values were not transformed into relative quantities. The BestKeeper algorithm was used to calculate gene expression variation based on cycle threshold values (C_q), crossing point standard deviation [{SD (±CP)} <1], and coefficient of variance (CV [%CP]). In BestKeeper analysis, genes with low SD (<1), low CV, and high coefficient of correlation (r) are generally considered stably expressed and vice versa.

Finally, the RefFinder tool (https://www.heartcure.com.au/reffinder/) was also employed to estimate the overall ranking of the 10 RGs by assigning an appropriate weight to each gene (Xie et al., 2012). The RefFinder analysis integrates the outcome of geNorm, NormFinder, and BestKeeper tools to provide an overall final ranking of RGs.

RESULTS

Primer Specificity, Amplification Efficiency, and Descriptive Statistics

This study evaluated the expression stability of 10 RGs (*GAPDH*, *RPL4, EEF1A1, RPS9, HPRT1, UXT, HMBS, B2M, RPS15*, and *ACTB*) in PBMCs of cattle types from cold arid high-altitude hypoxia and hot arid normoxia environments. The specificity of each primer pair was ascertained by the presence of the single specific amplicon of expected size on 2.5% agarose gel (**Supplementary Figure S1**). The melt curve plot of each RG showed a single peak suggesting the highly specific nature of primer pairs used (**Supplementary Figure S2**). The amplification efficiencies estimated from the six-point standard curve (generated from five-fold serial dilution of pooled cDNA) ranged from 88 to 110%. The slope values of the standard curve for different RGs ranged between −3.05 and −3.5 which was within the acceptable limit (−2.96 to −3.6). Based on the overall evaluation of the melt curve, amplification efficiencies, and slope values, it could be safely assumed that RT-qPCR data for each primer pair were of high quality (**Table 1**).

The average raw Cq values of individual RGs across all PBMCs in cold arid and hot arid groups (combined analysis) are summarized in **Table 2**. The average Cq values of individual RGs were quite variable and ranged from 17.66 (*EEF1A1*) to 27.86 (*HMBS*). On the basis of their average C_t scores, the 10 RGs were classified into group-I (abundantly expressed), group-II (moderately expressed), and group-III (least expressed). Group-1 included *EEF1A1*, *B2M*, *RPS15,* and *RPL4* genes with a high expression level and average Cq scores of 17.66, 18.18, 19.35, and 19.68, respectively. Group-II included *RPS9*, *GAPDH,* and *ACTB* that displayed an intermediate expression level with average Cq scores of 20.33, 21.69, and 22.27, respectively. Group-III included *UXT*, *HPRT1,* and *HMBS* with the least expression and average Cq scores of 24.76, 26.18, and 27.86, respectively (**Table 2**). Considering the distribution of average raw Cq scores and interquartile range, *EEF1A1* exhibited the lowest coefficient of variations (lowest variability across samples). On the other hand, the *HMBS* gene showed the highest coefficient of variation. Based on this parameter, *EEF1A1*, *B2M*, *RPS15,* and *RPL4* RGs were most stable while *UXT*, *HPRT1,* and *HMBS* genes were the least stable across the combined dataset (**Figure 2**).

Similar to the combined dataset, in cold arid as well as hot arid normoxia groups, *EEF1A1* showed the highest maximum and *HMBS* showed the least expression level. The average Cq values of individual RGs in cold arid hypoxia and hot arid normoxia groups are summarized in **Supplementary Tables S1,S2** and **Supplementary Figures S3,S4** respectively.

Selection of Reference Genes by geNorm Analysis

The first analysis to determine the expression stability of 10 candidate RGs in PBMCs of heifer cows from the hypoxia cold arid group (LAC, JYC, and HFX) and normoxia hot arid group (SAC, KFC, and HFC) was based on the geNorm algorithm. The expression stability of individual RGs was evaluated first by combining RT-qPCR data for all 30 PBMCs samples (cold arid hypoxia and hot arid normoxia groups). All the RGs in combined group analysis showed stable expression with a stability index within the acceptable range (<M value < 1.5). In the combined analysis, *RPS9* and *RPS15* showed the highest expression stability ($M = 0.464$), followed by *RPL4* ($M = 0.527$) and *GAPDH* ($M = 0.539$), whereas *HPRT1* was the least stable gene ($M = 1.228$). On the basis of average expression stability measure, RGs were arranged from most stable (lowest M value) to the least stable (highest M value): *RPS9* = *RPS15* > *RPL4*> *GAPDH* > HMBS > EEF1A1> *B2M* > *UXT* > *ACTB* > *HPRT1* (**Figure 3A**). Since the samples were from two distinct altitudes, we tried to determine the optimal number of RGs on the basis of pair-wise variation (Vn/n+1). The V values were calculated in different combinations: V2/V3, V3/V4, V4/V5,

TABLE 2 | The average raw Ct values of individual RGs in cattle populations adapted to cold arid high-altitude hypoxia (LAC, HFX, JYC) and hot arid normoxia environments (SAC, KFC, HFC).

	S.no	Animal	*GAPDH*	*RPL4*	*EEF1A1*	*RPS9*	*HPRT11*	*UXT*	*HMBS*	*B2M*	*RPS15*	*ACTB*
Cold Arid Hypoxia group	1	LAC1	22.17	21.56	18.81	22.40	30.18	26.07	29.44	19.36	19.51	24.35
	2	LAC2	21.87	20.06	17.32	20.15	28.38	23.86	27.99	18.20	19.79	22.88
	3	LAC3	22.08	19.67	16.90	19.98	27.95	24.18	27.63	18.72	19.55	23.52
	4	LAC4	22.30	20.15	17.26	19.87	28.14	24.55	27.64	18.57	19.57	24.66
	5	LAC5	22.53	19.61	17.09	20.04	28.87	24.68	27.59	18.75	19.21	22.64
	6	LAC6	22.07	19.75	17.23	20.41	29.69	24.51	27.79	17.74	19.45	24.90
	7	HFX1	21.52	19.35	16.90	19.73	27.46	24.30	27.80	17.61	18.80	21.33
	8	HFX2	21.16	19.40	17.21	20.13	27.31	24.83	27.94	17.18	19.09	20.65
	9	HFX3	21.64	19.71	17.12	20.07	27.71	24.41	27.77	17.82	19.12	21.16
	10	HFX4	21.66	19.76	17.00	20.13	28.55	23.46	27.81	17.66	18.93	21.71
	11	HFX5	21.90	19.78	17.48	20.08	27.83	26.76	27.92	17.81	19.38	22.36
	12	JYC1	21.71	19.39	17.25	20.13	28.40	25.27	27.77	17.64	19.16	21.82
	13	JYC2	21.41	20.10	17.79	20.59	27.80	26.42	27.87	17.72	19.74	22.03
	14	JYC3	21.44	19.34	17.21	20.53	27.43	26.31	26.89	17.82	19.43	27.44
	15	JYC4	22.24	20.58	18.34	21.14	28.26	29.62	28.56	18.85	20.24	22.93
Hot Arid Normoxia group	16	SAC1	21.60	20.38	18.34	20.40	24.43	27.86	28.21	18.09	19.83	21.60
	17	SAC2	21.87	20.18	18.36	20.63	24.95	23.99	27.97	18.04	19.39	21.09
	18	SAC3	22.96	20.04	18.70	20.73	24.09	24.36	28.57	18.91	19.72	23.71
	19	SAC4	22.49	20.65	18.78	20.84	24.64	24.63	28.68	18.29	20.11	22.49
	20	SAC5	22.23	20.80	19.10	21.00	24.91	24.90	28.90	18.71	20.49	22.04
	21	KFC1	22.07	19.84	18.18	20.73	24.58	24.25	27.81	18.10	19.88	21.79
	22	KFC2	21.63	19.73	19.32	20.77	24.44	24.17	27.51	17.75	19.63	21.18
	23	KFC3	21.45	19.24	17.66	19.89	23.60	24.23	27.31	17.73	18.97	21.20
	24	KFC4	21.39	18.73	17.51	19.92	23.71	23.81	26.99	17.57	18.39	21.41
	25	KFC5	21.74	19.65	17.94	20.65	24.40	24.30	28.04	17.95	19.54	21.51
	26	HFC1	19.47	17.07	15.49	18.01	21.51	21.83	26.62	15.69	17.02	18.73
	27	HFC2	22.44	20.56	19.27	21.35	25.41	25.16	28.58	20.29	20.32	22.72
	28	HFC3	19.08	17.03	15.36	17.81	21.59	21.53	24.87	17.95	16.73	18.12
	29	HFC4	20.76	18.38	16.84	21.11	24.78	24.53	29.61	20.04	20.43	24.27
	30	HFC5	21.95	19.90	18.26	20.73	24.61	24.25	27.72	18.96	19.21	21.78
Average			21.7	19.7	17.7	20.3	26.2	24.8	27.9	18.2	19.4	22.3

LAC, ladakhi cattle; HFX, holstein friesian cross; JYC, jersey cattle; SAC, sahiwal cattle; KFC, karan fries cattle; HFC, holstein friesian cattle.

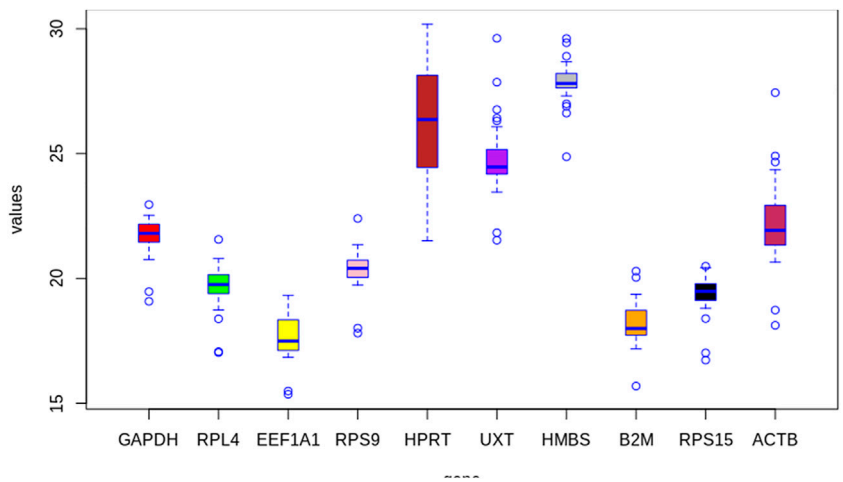

FIGURE 2 | Box whisker plot showing expression levels of individual candidate RGs in the combined group (cold arid and hot arid). The data are presented as quantification cycle (CP) values of each gene in the box-whisker diagram. The median is shown as a line across the box while whiskers indicate maximum and minimum values.

V5/V6 by adding the third, fourth, fifth, and sixth less stable genes, respectively. The V values for V2/V3 (0.169), V3/V4 (0.116), V4/V5 (0.101), V5/V6 (0.092), and V6/V7 (0.103) were either > or < the threshold of 0.15 (**Figure 3B**). It is generally considered that when Vn/n+1 is < 0.15 (threshold value), inclusion of an additional gene is not required for

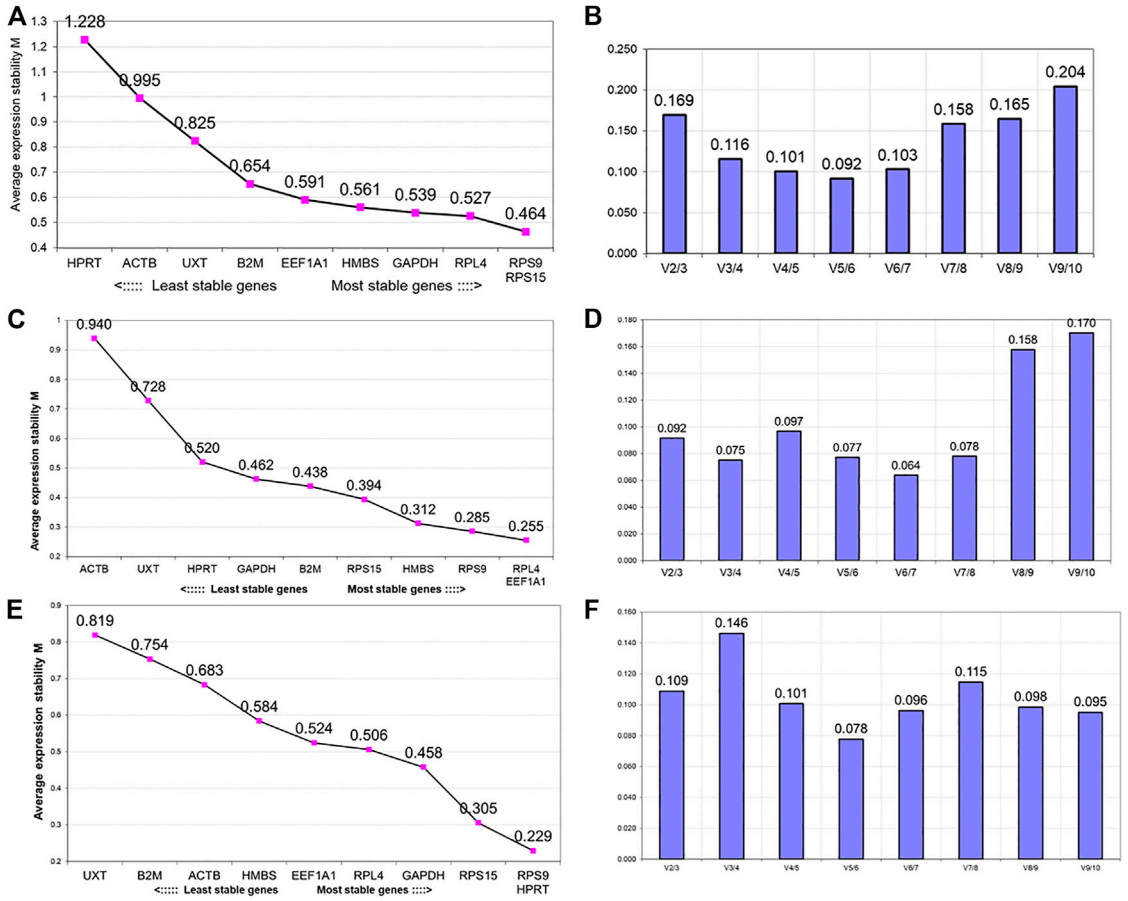

FIGURE 3 | GeNorm analysis for ranking of genes based on average expression stability measure (M value) and pair-wise variation (Vn/Vn+ 1) between the normalisation factors NFn and NFn + 1 to determine the optimal number of reference genes. Analysis in the combined group (cold arid and hot arid) (**A** and **B**, respectively), cold arid hypoxia group (**C** and **D**, respectively), and hot arid normoxia group (**E** and **F**, respectively).

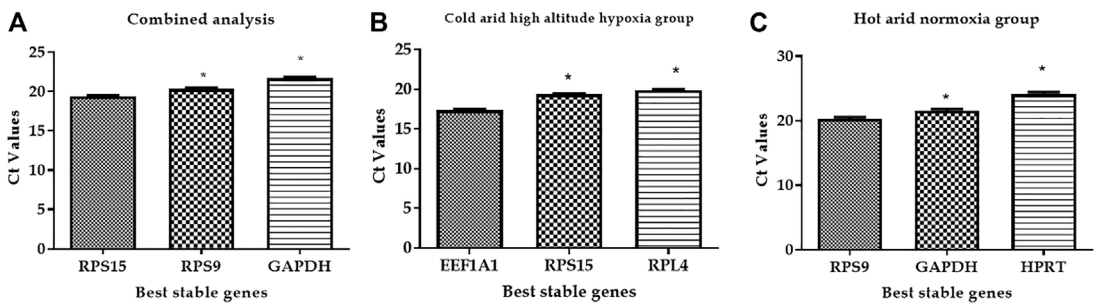

FIGURE 4 | Expression stability of the three best stable RGs. **(A)** Combined group (cold arid and hot arid). **(B)** cold arid high-altitude hypoxia group, and **(C)** hot arid normoxia group.

calculating the normalisation factor. In the combined dataset, combination V3/V4 resulted in a Vn/n+1 value < 0.116 which was less than the cut off value of 0.15 suggesting that the first three most stable RGs will be sufficient for accurate normalisation of RT-qPCR data. Therefore, based on M (stability) and V values (pair-wise variation) derived from geNorm analysis, the most stable RGs for the combined dataset were *RPS9, RPS15,* and *RPL4* (**Figure 4** and **Table 3**).

In addition to combined analysis, the geNorm tool was used to evaluate the expression stability of individual RGs separately in the cold arid hypoxia group (LAC, JYC, and HFX) and hot arid normoxia group (SAC, KFC, and HFC). The M values for all the 10

TABLE 3 | Overall ranking of the best suitable reference genes in combined (cold arid and hot arid groups), Cold arid hypoxia group, and hot arid normoxia group analysis.

Group	Ranking	geNorm	NormFinder	BestKeeper		ReFinder
		M value	Stability value	STDEV	coff. of corr (r)	
Combined analysis	1	RPS9/RPS15 (0.464)	RPL4 (0.282)	GAPDH (0.52)	RPS9 (0.901)	RPS9 (1.41)
	2	RPL4 (0.527)	RPS9 (0.289)	HMBS (0.55)	RPL4 (0.894)	RPS15 (1.86)
	3	GAPDH (0.539)	RPS15 (0.292)	RPS15 (0.55)	RPS15 (0.890)	GAPDH (2.83)
	4	HMBS (0.561)	GAPDH (0.302)	RPS9 (0.58)	GAPDH (0.849)	RPL4 (3.41)
	5	EEF1A1 (0.591)	HMBS (0.382)	RPL4 (0.62)	HMBS (0.795)	HMBS (3.98)
	6	B2M (0.654)	B2M (0.432)	B2M (0.63)	ACTB (0.748)	EEF1A1 (6.48)
	7	UXT (0.825)	EEF1A1 (0.532)	EEF1A1 (0.77)	UXT (0.721)	B2M (6.48)
	8	ACTB (0.995)	UXT (0.605)	UXT (1.03)	EEF1A1 (0.704)	UXT (8)
	9	HPRT1 (1.228)	ACTB (0.737)	ACTB (1.29)	B2M (0.649)	ACTB (9)
	10	—	HPRT1 (1.288)	HPRT1 (2.08)	HPRT1 (0.640)	HPRT1 (10)
Cold arid hypoxia group	1	RPL4/EEF1A1 (0.255)	RPS15 (0.140)	RPS15 (0.28)	EEF1A1 (0.865)	EEF1A1 (1.41)
	2	RPS9 (0.285)	EEF1A1 (0.140)	GAPDH (0.32)	RPS9 (0.858)	RPL4 (2.11)
	3	HMBS (0.312)	RPS9 (0.155)	HMBS (0.32)	RPL4 (0.846)	RPS15 (3.16)
	4	RPS15 (0.394)	RPL4 (0.200)	EEF1A1 (0.38)	B2M (0.781)	RPS9 (3.57)
	5	B2M (0.438)	B2M (0.233)	RPL4 (0.41)	RPS15 (0.717)	HMBS (3.87)
	6	GAPDH (0.462)	GAPDH (0.259)	RPS9 (0.44)	UXT (0.639)	GAPDH (5.45)
	7	HPRT1 (0.520)	HMBS (0.320)	B2M (0.52)	HMBS (0.619)	B2M (5.96)
	8	UXT (0.728)	HPRT1 (0.389)	HPRT1 (0.60)	HPRT1 (0.612)	HPRT1 (8)
	9	ACTB (0.940)	UXT (0.685)	UXT (1.17)	GAPDH (0.571)	UXT (9)
	10	—	ACTB (0.808)	ACTB (1.34)	ACTB (0.571)	ACTB (10)
Hot arid normoxia group	1	HPRT1/RPS9 (0.229)	RPS15 (0.107)	GAPDH (0.74)	RPS9 (0.978)	RPS9 (1.32)
	2	RPS15 (0.305)	RPS9 (0.134)	B2M (0.74)	RPS15 (0.971)	HPRT1 (2.71)
	3	RPL4 (0.458)	HPRT1 (0.136)	RPS9 (0.75)	HPRT1 (0.962)	GAPDH (2.83)
	4	GAPDH (0.506)	HMBS (0.200)	HMBS (0.79)	RPL4 (0.905)	RPS15 (3.03)
	5	EEF1A1 (0.524)	GAPDH (0.225)	UXT (0.80)	GAPDH (0.904)	HMBS (5.38)
	6	HMBS (0.584)	RPL4 (0.225)	HPRT1 (0.81)	EEF1A1 (0.889)	RPL4 (6.06)
	7	ACTB (0.683)	EEF1A1 (0.238)	RPS15 (0.83)	HMBS (0.879)	B2M (6.18)
	8	B2M (0.754)	UXT (0.292)	EEF1A1 (0.91)	ACTB (0.879)	EEF1A1 (6.96)
	9	UXT (0.819)	ACTB (0.307)	RPL4 (0.93)	UXT (0.790)	UXT (8.41)
	10	—	B2M (0.363)	ACTB (1.04)	B2M (0.715)	ACTB (8.46)

RGs in the hypoxia cold arid group were <1.5. The ranking of RGs in order of expression stability within this group were: *RPL4* = *EEF1A1* >*RPS9*> *HMBS* > *RPS15* > *B2M* > *GAPDH* > *HPRT11* > *UXT* > *ACTB* (**Figure 3C**; **Table 3**). *RPL4* and *EEF1A1* showed the highest expression stability (M = 0.255) followed by *RPS9* (M = 0.285) and *HMBS* (M = 0.312), whereas *ACTB* was least stable (M = 0.940). The pair-wise variation analysis for different combinations of RGs was well below the threshold value of 0.15. The results obtained for V2/3 (0.092), V3/4 (0.075), V4/5 (0.097), and V5/6 (0.077) combinations indicated that *RPL4* and *EEF1A1* would provide the most accurate normalisation (**Figure 3D**). The expression stability of the three best RGs in combined analysis is shown in **Figure 4B**.

In the hot arid normoxia group (SAC, KFC, and HFC), the ranking of RGs in order of expression stability were: *HPRT1* = *RPS9*> *RPS15* > *GAPDH* > *RPL4*> *EEF1A1*> *HMBS* > *ACTB* > *B2M* > *UXT* (**Figure 3E**; **Table 3**). The M values for all the 10 RGs were <1.5. *HPRT1* (M = 0.229) and *RPS9* (M = 0.229) showed the highest stability followed by *RSP15* (M = 0.305) whereas *UXT* was least stable (M = 0.819). The results for V2/3 (0.109), indicated that *RPS9* and *HPRT1* would provide reliable normalisation in PBMCs of the hot arid normoxia group (**Figure 3F**). The results of geNorm analysis for the three groups are summarized in **Table 3**. Overall, combined and group-wise analysis implied that *RPL4, RPS9*, *RPS15,* and *EEF1A1* were the most stably expressed RGs (**Figure 4C**).

Selection of Reference Genes by NormFinder Analysis

In the NormFinder-based intergroup (combined) analysis covering both the conditions and all samples (cold arid hypoxia and hot arid normoxia), *RPL4, RPS9,* and *RPS15* were found to be most stable with stability values of 0.282, 0.289, and 0.292 respectively. On the other hand, *UXT*, *ACTB,* and *HPRT1* were least stable with stability values of 0.605, 0.737, and 1.288, respectively (**Table 3**). The graph showing intragroup variation of RGs in the combined dataset is shown in **Figure 5A**. Based on stability values, the RGs were ranked as *RPL4*> *RPS9*> *RPS15* > *GAPDH* > *HMBS* > *B2M* > *EEF1A1*> *UXT* > *ACTB* > *HPRT1*. Within the cold arid hypoxia group, *RPS15*, *EEF1A1,* and *RPS9* were the most stable RGs with stability values of 0.140, 0.140, and 0.155, respectively, whereas, *ACTB* was the least stable gene with the highest variability value of 0.808 (**Table 3**). The graph showing intragroup variation analysis of RGs is shown in **Figure 5B**. The RGs were ranked as *RPS15*> *EEF1A1*> *RPS9*> *RPL4*> *B2M* > *GAPDH* > *HMBS* > *HPRT1*>*UXT* > *ACTB*.

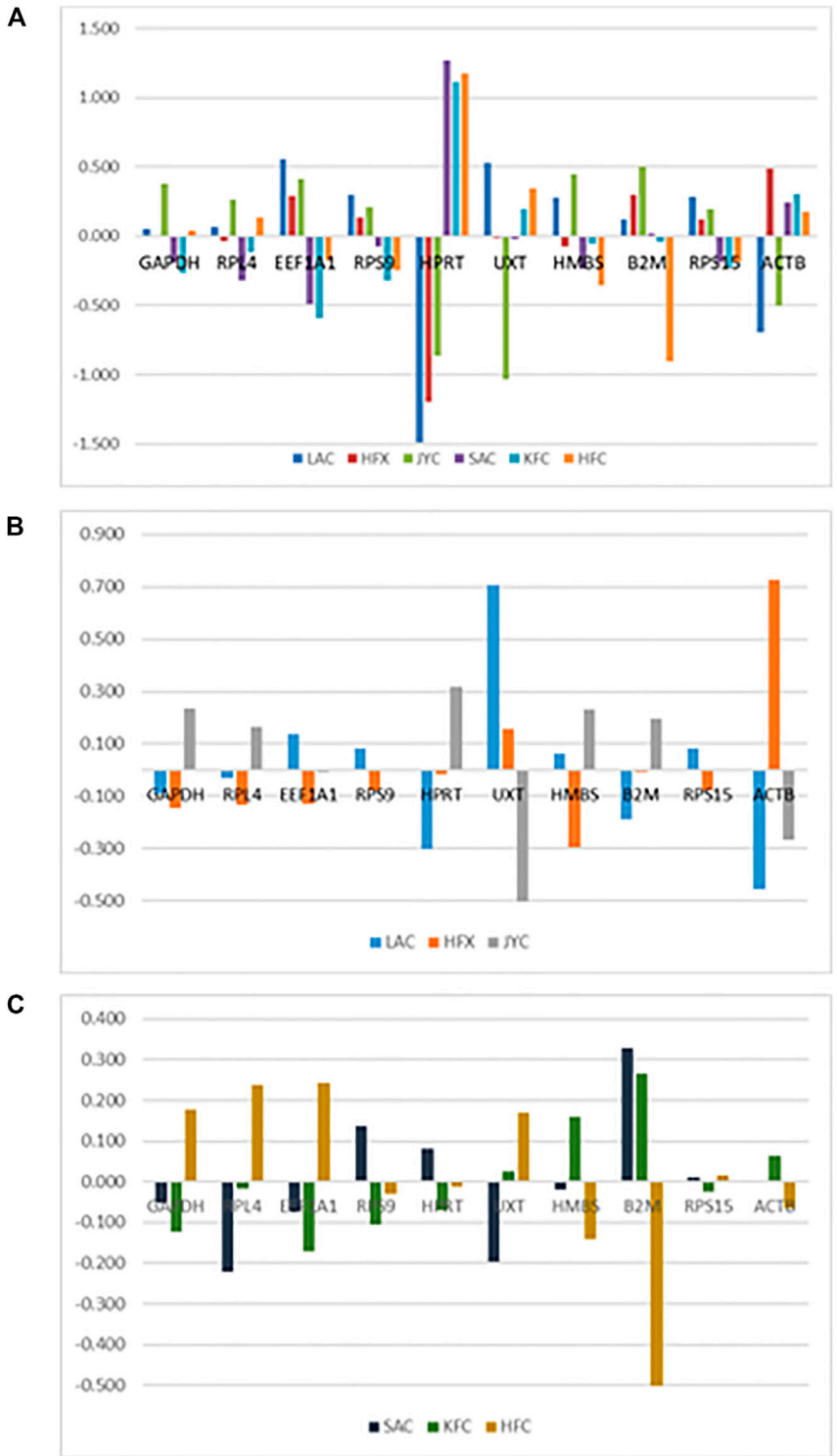

FIGURE 5 | Inter-group variation analysis of RGs **(A)** Combined group (cold arid and hot arid). **(B)** cold arid hypoxia group, and **(C)** hot arid normoxia group.

Within the hot arid normoxia group, NormFinder analysis identified *RPS15*, *RPS9*, and *HPRT1* as the three most stable RGs with stability values of 0.107, 0.134, and 0.136, respectively (**Table 3**). The graph showing intragroup variation analysis of RGs in the combined dataset is shown in **Figure 5C**. Based on stability values, the RGs were ranked as *RPS15* > *RPS9*> *HPRT1*> *HMBS* > *GAPDH* > *RPL4*> *EEF1A1*> *UXT* > *ACTB* > *B2M*. Overall, there was a good agreement in

TABLE 4 | Combined (cold arid and hot arid groups) analysis of parameters-based quantitative cycling points (CP) for 10 candidate RGs

	GAPDH	*RPL4*	*EEF1A1*	*RPS9*	*HPRT1*	*UXT*	*HMBS*	*B2M*	*RPS15*	*ACTB*
N	30	30	30	30	30	30	30	30	30	30
geo Mean [CP]	21.68	19.66	17.64	20.31	26.09	24.72	27.85	18.16	19.34	22.2
ar Mean [CP]	21.69	19.68	17.67	20.33	26.19	24.77	27.86	18.18	19.35	22.27
min [CP]	19.08	17.03	15.36	17.81	21.51	21.53	24.87	15.69	16.73	18.12
max [CP]	22.96	21.56	19.32	22.4	30.18	29.62	29.61	20.29	20.49	27.44
std dev [± CP]	0.52	0.62	0.77	0.58	2.08	1.03	0.55	0.63	0.55	1.29
CV [% CP]	2.4	3.15	4.34	2.85	7.93	4.18	1.97	3.46	2.85	5.81
min [x-fold]	−6.06	−6.18	−4.86	−5.67	−23.84	−9.14	−7.87	−5.56	−6.09	−16.89
max [x-fold]	2.43	3.74	3.2	4.25	17.08	29.82	3.39	4.37	2.23	37.85
std dev [± x-fold]	1.44	1.54	1.7	1.5	4.22	2.05	1.46	1.55	1.47	2.45

N, number of samples; geo Mean [CP], geometric mean of CP; ar Mean [CP], arithmetic mean of CP; min [CP] and max [CP], extreme values of CP; Std dev [±CP], standard deviation of the CP; CV [%CP], coefficient of variation expressed as a percentage on the CP, values; min [x-fold] and max [x-fold], extreme values of expression levels expressed as absolute x-fold over or under the coefficient; std dev[±x-fold], standard deviation of the absolute regulation coefficients.

TABLE 5 | Combined (cold arid and hot arid groups) analysis of repeated pair-wise correlation amongst genes with the BestKeeper index.

	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
RPL4	0.861	—	—	—	—	—	—	—	—	—
p-value	0.001	—	—	—	—	—	—	—	—	—
EEF1A1	0.725	0.817	—	—	—	—	—	—	—	—
p-value	0.001	0.001	—	—	—	—	—	—	—	—
RPS9	0.722	0.834	0.822	—	—	—	—	—	—	—
p-value	0.001	0.001	0.001	—	—	—	—	—	—	—
HPRT1	0.489	0.523	0.041	0.39	—	—	—	—	—	—
p-value	0.006	0.003	0.829	0.033	—	—	—	—	—	—
UXT	0.469	0.614	0.468	0.586	0.457	—	—	—	—	—
p-value	0.009	0.001	0.009	0.001	0.011	—	—	—	—	—
HMBS	0.639	0.714	0.636	0.847	0.351	0.5	—	—	—	—
p-value	0.001	0.001	0.001	0.001	0.057	0.005	—	—	—	—
B2M	0.509	0.494	0.491	0.66	0.178	0.307	0.589	—	—	—
p-value	0.004	0.006	0.006	0.001	0.348	0.099	0.001	—	—	—
RPS15	0.769	0.787	0.741	0.849	0.366	0.615	0.831	0.643	—	—
p-value	0.001	0.001	0.001	0.001	0.047	0.001	0.001	0.001	—	—
ACTB	0.569	0.478	0.270	0.584	0.57	0.458	0.458	0.483	0.609	—
p-value	0.001	0.008	0.149	0.001	0.001	0.011	0.011	—	—	—
BestKeeper vs.	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
coeff. of corr. [r]	0.849	0.894	0.704	0.901	0.640	0.721	0.795	0.649	0.890	0.748
p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

geNorm and NormFinder outcomes for all the datasets (**Table 3**).

Selection of Reference Genes by BestKeeper Analysis

The descriptive statistics for BestKeeper analysis of the combined dataset are shown in **Table 4**. The analysis suggested stable expression (SD < 1) for the majority of genes: *GAPDH* (CV ± SD = 2.40 ± 0.52), *HMBS* (CV ± SD = 1.97 ± 0.55), *RPS15* (CV ± SD = 2.85 ± 0.55), *RPS9* (CV ± SD = 2.85 ± 0.58), *RPL4* (CV ± SD = 3.15 ± 0.62), *B2M* (CV ± SD = 3.46 ± 0.63), and *EEF1A1* (CV ± SD = 4.35 ± 0.77). Only three genes, *UXT* (CV ± SD = 4.18 ± 1.03), *ACTB* (CV ± SD = 5.81 ± 1.29), and *HPRT1* (CV ± SD = 7.93 ± 2.07) showed SD > 1 and hence were considered unacceptable as RGs. Additionally, the BestKeeper software was used to measure the inter-gene relation amongst the 10 RGs in the combined dataset by calculating Pearson correlation coefficient (*r*) values. Several RG pairs *GAPDH/RPL4* (r = 0.861) *RPS9/RPS15* (r = 0.849), *RPS9/HMBS* (r = 0.847), *RPL4/RPS9* (r = 0.834), and *HMBS/RPS5* (r = 0.832) showed high correlation coefficients (**Table 5**). The RG pairs with high *r* values suggested that their expression pattern in PBMCs of cattle populations from two distinct climates and altitudes are more or less similar to each other. Further, the BestKeeper index (BI) calculated for each gene was used to estimate the correlation value between BI and each RG. As shown in **Table 5**, the best correlation values were obtained for *RPS9* (r = 0.901), *RPL4* (r = 0.894), *RPS15* (r = 0.890), and *GAPDH* (r = 0.849). The high correlation values strongly suggested their suitability as reliable RGs in the experimental set up of the study.

The BestKeeper analysis for the cold arid hypoxia group and hot arid normoxia group was also conducted separately. In the cold arid hypoxia group, the following 8 RGs showed lower a coefficient of variation (CV) and standard deviation (SD): *RPS15* (CV ± SD = 1.44 ± 0.28), *HMBS* (CV ± SD = 1.14 ± 0.32), *GAPDH* (CV ± SD = 1.46 ± 0.32), *EEF1A1* (CV±SD = 2.18 ± 0.38), *RPL4* (CV ± SD =

TABLE 6 | Parameters-based quantitative cycling points (CP) for 10 RGs in PBMCs of cattle populations adapted to cold arid high-altitude hypoxia conditions.

	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
N	15	15	15	15	15	15	15	15	15	15
geo Mean [CP]	21.84	19.87	17.39	20.35	28.25	25.24	27.89	18.09	19.39	22.9
ar Mean [CP]	21.85	19.88	17.39	20.36	28.26	25.28	27.89	18.1	19.4	22.96
min [CP]	21.16	19.34	16.9	19.73	27.31	23.46	26.89	17.18	18.8	20.65
max [CP]	22.53	21.56	18.81	22.4	30.18	29.62	29.44	19.36	20.24	27.44
std dev [+/- CP]	0.32	0.41	0.38	0.44	0.6	1.17	0.32	0.52	0.28	1.34
CV [% CP]	1.46	2.04	2.18	2.15	2.12	4.63	1.14	2.85	1.44	5.85
min [x-fold]	–1.61	–1.45	–1.4	–1.54	–1.92	–3.43	–2	–1.88	–1.51	–4.75
max [x-fold]	1.61	3.22	2.68	4.14	3.8	20.83	2.93	2.42	1.8	23.31

N, number of samples, geo Mean [CP], geometric mean of CP; ar Mean [CP], arithmetic mean of CP; min [CP] and max [CP], extreme values of CP; Std dev [±CP], standard deviation of the CP; CV [%CP], coefficient of variation expressed as a percentage on the CP, values; min [x-fold] and max [x-fold], extreme values of expression levels expressed as absolute x-fold over or under the coefficient; std dev[±x-fold], standard deviation of the absolute regulation coefficients.

TABLE 7 | Repeated pair-wise correlation amongst genes with the BestKeeper index in cattle populations adapted to cold arid high-altitude hypoxia conditions.

	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
RPL4	0.458	—	—	—	—	—	—	—	—	—
p-value	0.086	—	—	—	—	—	—	—	—	—
EEF1A1	0.261	0.896	—	—	—	—	—	—	—	—
p-value	0.348	0.001	—	—	—	—	—	—	—	—
RPS9	0.213	0.856	0.925	—	—	—	—	—	—	—
p-value	0.446	0.001	0.001	—	—	—	—	—	—	—
HPRT1	0.617	0.635	0.489	0.612	—	—	—	—	—	—
p-value	0.014	0.011	0.065	0.015	—	—	—	—	—	—
UXT	0.128	0.401	0.702	0.531	–0.031	—	—	—	—	—
p-value	0.65	0.139	0.004	0.042	0.913	—	—	—	—	—
HMBS	0.285	0.866	0.831	0.78	0.589	0.339	—	—	—	—
p-value	0.303	0.001	0.001	0.001	0.021	0.216	—	—	—	—
B2M	0.819	0.739	0.566	0.568	0.551	0.283	0.524	—	—	—
p-value	0.001	0.002	0.028	0.027	0.033	0.306	0.045	—	—	—
RPS15	0.401	0.554	0.6	0.443	0.175	0.647	0.302	0.537	—	—
p-value	0.138	0.032	0.018	0.098	0.532	0.009	0.274	0.039	—	—
ACTB	0.311	0.201	0.177	0.324	0.286	0.182	–0.187	0.378	0.384	—
p-value	0.259	0.473	0.529	0.239	0.302	0.516	0.504	0.165	0.158	—
BestKeeper vs.	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
coeff. of corr. [r]	0.569	0.846	0.866	0.858	0.613	0.639	0.618	0.782	0.715	0.571
p-value	0.027	0.001	0.001	0.001	0.015	0.01	0.014	0.001	0.003	0.026

2.04 ± 0.41), *RPS9* (CV ± SD = 2.15 ± 0.44), *B2M* (CV ± SD = 2.85 ± 0.52), and *HPRT1* (CV ± SD = 2.12 ± 0.60) (**Table 6**). The remaining two RGs showed high SD and CV values: *ACTB* (CV ± SD = 5.85 ± 1.34) and *UXT* (CV ± SD = 4.62 ± 1.17) and were unstable. The inter-gene relation showed very high correlation coefficient values (*r*) for *EEF1A1*/*RPS9* (*r* = 0.93), *EEF1A1*/*RPL4* (*r* = 0.90), and *RPL4*/*RPS9* (*r* = 0.86) (**Table 7**). Further, the coefficient of correlation analysis (*r*) of individual RGs with the BestKeeper index resulted in high values for *EEF1A1* (*r* = 0.865), *RPS9* (*r* = 0.858), and *RPL4* (*r* = 0.846). These results strongly suggest that *EEF1A1*, *RPS9,* and *RPL4* with low CV, low SD, and high coefficient of correlation to the BestKeeper index should be the ideal panel for the cold arid high-altitude hypoxia group.

The BestKeeper analysis for the hot arid normoxia group also resulted in several RGs with SD < 1 (**Table 8**). The following nine RGs: *GAPDH* (CV ± SD = 3.44 ± 0.74), *B2M* (CV ± SD = 4.07 ± 0.74), *HMBS* (CV ± SD = 3.47 ± 0.96), *RPS 9* (CV ± SD = 2.85 ± 0.79), *UXT* (CV ± SD = 3.28 ± 0.80), *HPRT1* (CV ± SD = 3.35 ± 0.81), *RPS15* (4.30 ± 0.83), *EEF1A1* (CV ± SD = 5.09 ± 0.91), and *RPL4* (CV ± SD = 4.75 ± 0.93) showed low SD and CV values. Whereas *ACTB* (CV ± SD = 54.82 ± 1.04) showed SD > 1 and hence was considered unstable.

The inter-gene relation showed a strong correlation coefficient (*r*) for *GAPDH*/*RPL4* (*r* = 0.930), *EEF1A1*/RPL4 (*r* = 0.951), *RPS9*/*HPRT11* (*r* = 0.981), *RPS15*/*HMBS* (*r* = 0.923), *HMBS*/*ACTB* (*r* = 0.908), *HPRT11*/*RPL4* (*r* = 874), *HPRT11*/*EEF1A1* (*r* = 861), *HMBS*/*RPS9* (*r* = 0.874), and *RPS9*/*RPS15* (*r* = 0.965) (**Table 9**). The highest correlation coefficient of individual RGs with the BestKeeper index was observed for *RPS9* (*r* = 0.978), *RPS15* (0.971), *HPRT1* (*r* = 0.962), *RPL4* (0.905), and *GAPDH* (*r* = 0.904) (**Table 9**). The high correlation coefficient of these genes suggested their suitability as RGs in the hot arid normoxia group.

Selection of Reference Genes by RefFinder Analysis

Additionally, the RefFinder algorithm was used to evaluate the comprehensive ranking of individual RGs in combined, cold arid

TABLE 8 | Parameters-based quantitative cycling points (CP) for 10 RGs in cattle populations adapted to hot arid normoxia conditions.

	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
N	15	15	15	15	15	15	15	15	15	15
geo Mean [CP]	21.52	19.44	17.9	20.28	24.08	24.22	27.8	18.24	19.28	21.52
ar Mean [CP]	21.54	19.48	17.94	20.3	24.11	24.25	27.83	18.27	19.31	21.58
min [CP]	19.08	17.03	15.36	17.81	21.51	21.53	24.87	15.69	16.73	18.12
max [CP]	22.96	20.8	19.32	21.35	25.41	27.86	29.61	20.29	20.49	24.27
std dev [+/- CP]	0.74	0.93	0.91	0.75	0.81	0.8	0.79	0.74	0.83	1.04
CV [% CP]	3.44	4.75	5.09	3.67	3.35	3.28	2.85	4.07	4.3	4.82
min [x-fold]	–5.41	–5.33	–5.81	–5.53	–5.95	–6.43	–7.64	–5.86	–5.85	–10.56
max [x-fold]	2.72	2.56	2.68	2.1	2.51	12.51	3.5	4.14	2.32	6.73
std dev [+/- x-fold]	1.67	1.9	1.88	1.68	1.75	1.74	1.73	1.67	1.78	2.05

N, number of samples; geo Mean [CP], geometric mean of CP; ar Mean [CP], arithmetic mean of CP; min [CP] and max [CP], extreme values of CP; Std dev [±CP], standard deviation of the CP; CV [%CP], coefficient of variation expressed as a percentage on the CP, values; min [x-fold] and max [x-fold], extreme values of expression levels expressed as absolute x-fold over or under the coefficient; std dev[±x-fold], standard deviation of the absolute regulation coefficients.

TABLE 9 | Repeated pair-wise correlation amongst genes with the BestKeeper index of cattle populations adapted to hot arid normoxia conditions.

	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
RPL4	0.93	—	—	—	—	—	—	—	—	—
p-value	0.001	—	—	—	—	—	—	—	—	—
EEF1A1	0.923	0.951	—	—	—	—	—	—	—	—
p-value	0.001	0.001	—	—	—	—	—	—	—	—
RPS9	0.869	0.852	0.865	—	—	—	—	—	—	—
p-value	0.001	0.001	0.001	—	—	—	—	—	—	—
HPRT1	0.841	0.874	0.861	0.981	—	—	—	—	—	—
p-value	0.001	0.001	0.001	0.001	—	—	—	—	—	—
UXT	0.64	0.758	0.678	0.706	0.726	—	—	—	—	—
p-value	0.01	0.001	0.005	0.003	0.002	—	—	—	—	—
HMBS	0.711	0.689	0.644	0.874	0.828	0.674	—	—	—	—
p-value	0.003	0.004	0.009	0.001	0.001	0.006	—	—	—	—
B2M	0.495	0.475	0.468	0.703	0.698	0.447	0.618	—	—	—
p-value	0.061	0.073	0.079	0.003	0.004	0.095	0.014	—	—	—
RPS15	0.819	0.842	0.822	0.965	0.943	0.762	0.923	0.69	—	—
p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.004	—	—
ACTB	0.75	0.623	0.627	0.877	0.802	0.617	0.908	0.749	0.875	—
p-value	0.001	0.013	0.012	0.001	0.001	0.014	0.001	0.001	0.001	—
BestKeeper vs.	GAPDH	RPL4	EEF1A1	RPS9	HPRT1	UXT	HMBS	B2M	RPS15	ACTB
coeff. of corr. [r]	0.904	0.905	0.889	0.978	0.962	0.79	0.879	0.715	0.971	0.879
p-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.003	0.001	0.001

hypoxia, and hot arid normoxia group datasets (Xie et al., 2012). In the combined dataset, RefFinder ranked *RPS9* (1.41), *RPS15* (1.86), and *GAPDH* (2.83) as the three most stable RGs, followed by *RPL4* (3.41), *HMBS* (3.98), *EEF1A1* (6.48), and *B2M* (6.48), whereas *UXT* (8.00), *ACTB* (9.00), and *HPRT1* (10.00) ranked as the three least stable RGs (**Table 3**). In the cold arid hypoxia group, RefFinder analysis ranked *EEF1A1* (1.41), *RPL4* (2.11), and *RPS15* (3.16) as the most stable RGs; while *HPRT11* (8.00), *UXT* (9.00), and *ACTB* (10.00) as the most unstable RGs. In the hot arid normoxia group, the analysis displayed *RPS9* (1.32), *HPRT1* (2.71), and *GAPDH* (2.83) as most stable while *EEF1A1* (6.96), *UXT* (8.41), and *ACTB* (8.46) as unstable RGs.

Validation of Selected Reference Genes

To evaluate the reliability of the best suitable and worst panel of RGs, a validation qPCR experiment was performed using some of the known candidate target genes associated with high-altitude hypoxia and heat stress response. The qPCR data for two target genes associated with hypoxia and high altitude such as HIF-alpha and *EPAS1* and two target genes associated with heat stress response such as *HSP70* and *HSP27* were generated in PBMC samples of high-altitude-adapted and tropically adapted cattle populations. As shown in **Figure 6A**, it is quite evident that the panel of best reference genes (*RPS19*, *RPS15,* and *GAPDH*) normalised the target gene data more accurately. We expected higher expression of high-altitude-associated genes in PBMCs of high-altitude cattle populations. On the other hand, genes related to heat stress response such as HSPs should have higher expression in PBMCs of cattle populations from the tropical region. As shown in **Figure 6A**, the expression of target genes such as HIF-*alpha* and *EPAS1* was higher in high-altitude (HA)-adapted cattle while genes like *HSP70* and *HSP27* were more expressed in low-altitude (LA) cattle breeds. Further, the relative expression, standard deviation (SD), and standard error (SE) of the four target genes also supported the high quality of the panel of reference genes used (**Supplementary Table S3**).

On the other hand, normalisation of same target genes with least stable genes (*HPRT* and *ACTB*) resulted in an unexpected

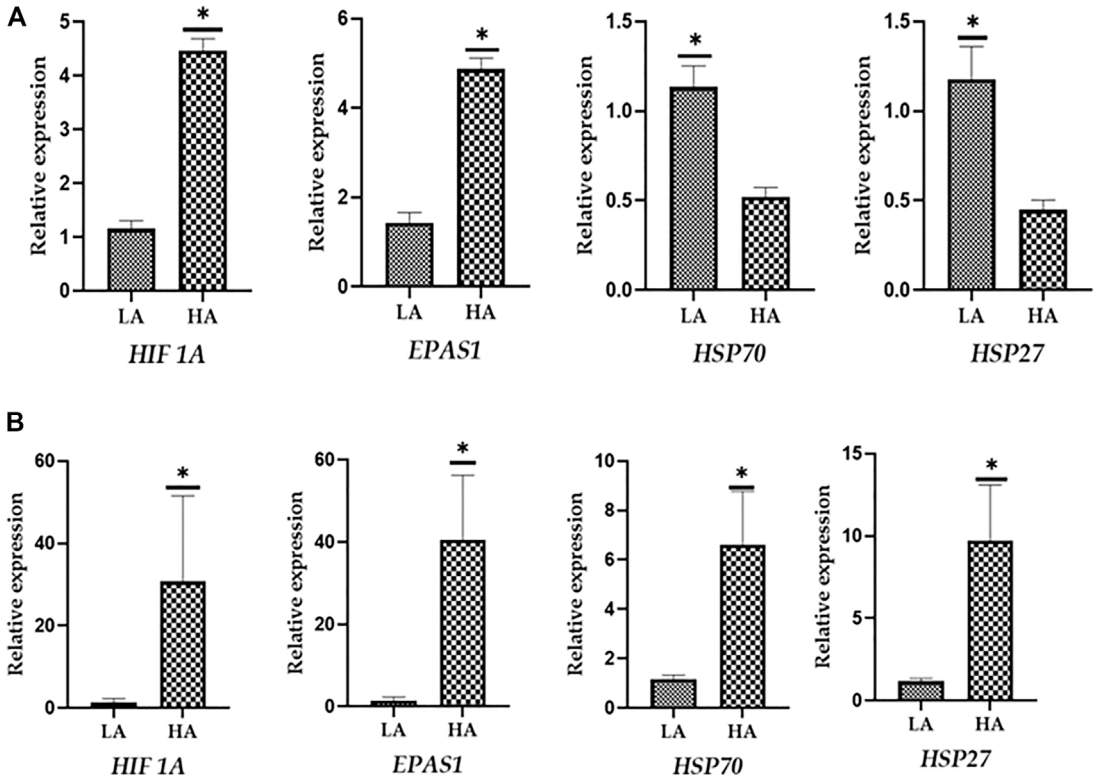

FIGURE 6 | (A) Evaluation of best reference genes (*RPS19, RPS15,* and *GAPDH*) in normalising the target genes. **(B)** Evaluation of least stable reference genes (*HPRT* and *ACTB*) in normalising the target genes.

pattern of expression. As shown in **Figure 6B**, HSPs showed higher expression in PBMCs of high-altitude cattle populations. Further, the use of least stable reference genes resulted in higher SD and SE values (**Supplementary Table S4**). Considering the above facts, it could be stated that the panel of RGs identified in the present study will be applicable for accurate normalisation of target genes in studies involving high and low-altitude cattle populations.

DISCUSSION

In an era of high-throughput platforms, quantitative PCR (qRT-PCR) is being employed as a most preferred tool to validate gene expression data. Even though qRT-PCR is the most sensitive technique, it suffers from several analytical variations like differences in the amount of starting material, RNA extraction, and efficiency of the reverse transcription process. To a great extent the effect of these non-biological variables can be nullified by normalising gene expression data by a panel of stable reference genes (RGs). As suggested in the "Minimum information for publication of Quantitative Real-time PCR Experiments" (MIQE) guideline (Bustin et al., 2009), accuracy in gene expression is largely governed by the availability of reliable RGs. At present there is no consensus for the set of reference gene(s) that can be used universally for the normalisation purposes. In the past, several studies have utilized RGs without proper validation or arbitrarily selected commonly used reference genes such as *ACTB* and *GAPDH*. Unfortunately, the use of such empirical RGs might not provide accurate normalisation and create doubt on the reliable estimation of gene expression. Moreover, many reports showed the variable expression of commonly used reference genes in different cells, tissue, and conditions (Kim and Yun 2011; Zhao et al., 2012; Thomas et al., 2014).

The major challenge in any biological experiment involving different tissues, cell types, disease state, and physiological and or developmental stages is the knowledge about appropriate RGs whose expression remains constant without any observable variations across samples (Bustin 2002; Radonić et al., 2004). It has been seen that a particular RG appropriate in one condition might have variable expression in another set of conditions. Thus, identification and validation of proper RGs is the prerequisite for any specific experimental condition. Our group has successfully identified a panel of appropriate RGs for various cellular types and experimental conditions involving zebu cattle and riverine buffaloes (Kishore et al., 2013; Sood et al., 2017; Lagah et al., 2019). The present study was also conducted on similar lines to identify suitable reference genes for cattle populations from diverse environmental conditions. The 10 RGs (RGs selected for the present work were part of our earlier studies, wherein these genes were evaluated for their suitability as normalisers in different cell types, cattle breeds, and experimental conditions (Kishore et al., 2013; Jatav et al., 2016; Kaur et al., 2018)).

Multiple RGs are generally preferred over a single RG to reduce the experimental variation for more effective normalisation (Vandesompele et al., 2002; Huang et al., 2014; Engdahl et al., 2016; Xu et al., 2016). Hence, in the present study, an optimal number of genes to be used as normalisers was identified. The data were analysed to identify the most appropriate panel of RGs for the combined dataset (cold arid and hot arid) and separately for the cold arid high-altitude hypoxia group and the hot arid normoxia group. Based on overall analysis, *RPS9*, *RPS15*, and *GAPDH* were marked as the most stable RGs for combined data. In the cold arid hypoxia group; *RPL4*, *EEF1A1,* and *RPS15* were the most stable RGs, while in the hot arid normoxia cattle group, *RPS9*, *HPRT1*, and *GAPDH* were identified as the most stable RGs.

In the present analysis, *RPS9, RPS15,* and, to a certain extent, *RPL4* genes were identified as stable RGs in the three groups (combined dataset, cold arid hypoxia group, and hot arid normoxia group). Both *RPS9* and *RPS15* genes are part of the ribosomal component of the small 40S subunit while *RPL4* is a 60S ribosomal protein L4. In the past, these ribosomal genes were also identified as ideal reference genes in many studies involving the mammary gland of dairy cows (Bionaz and Loor 2007), mammary epithelial cells of native cows (Pradeep et al., 2014), different tissues of riverine buffaloes (Kaur et al., 2018), heat-stressed MECs of riverine buffaloes (Kapila et al., 2013), and PBMCs of native cows and riverine buffaloes (Kishore et al., 2013). It is known that these ribosomal proteins are highly conserved (Hsiao et al., 2001) and required by all life forms to synthesise new ribosomes. The unchanged expression of ribosomal proteins has also been reported in the red flour beetle (*T. castaneum*) during a fungal infection (Lu et al., 2018b). Similarly, stability of different ribosomal gene expressions has also been validated in several studies involving abiotic and biotic challenges (Ma et al., 2016; Yan et al., 2016; Lu et al., 2018a; Luo et al., 2018). In our study, the genes encoding ribosomal proteins also showed stabilised expression in PBMCs derived from different cattle types maintained at different altitudes. Based on current findings as well as previous studies, it could be concluded that ribosomal proteins exhibit higher expression stability and are good candidates as reference genes. *GAPDH*, the other stable RG in our study, has also been identified as an appropriate reference gene in several other studies (Lisowski et al., 2008; Yang et al., 2014; Engdahl et al., 2016; Jatav et al., 2016). Along with *GAPDH*, *EEF1A1* and *HPRT1* were the other genes that ranked amongst the first three most stable RGs in cold arid hypoxia and hot arid normoxia groups, respectively. *EEF1A1* regulates the enzymatic delivery of aminoacyl tRNAs to the ribosome while *HPRT11* catalyses the conversion of hypoxanthine to inosine monophosphate and guanine to guanosine monophosphate, and the regulates generation of purine nucleotides through the purine salvage pathway. To the best of our knowledge, this is the first study to assess the expression stability of 10 candidate RGs in cattle populations adapted to distinct altitudes.

CONCLUSION

This study has identified an altitude-specific panel of RGs in PBMCs of cattle populations from hot arid normoxia and cold arid high-altitude hypoxic environments. In the cold arid hypoxia group, *RPL4*, *EEF1A1,* and *RPS15* RGs were the most stable, while in the hot arid normoxia group, *RPS9*, *HPRT1*, and *GAPDH* were identified as the most stable RGs. Further, the combined analysis resulted in identification of a panel of *RPS9, RPS15, and GAPDH* RGs that could act as a useful resource to unravel the accurate transcriptional profile of peripheral blood mononuclear cells of cattle populations adapted to tropical and high-altitude conditions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Institute Animal Ethical Committee, ICAR-NBAGR, Karnal.

AUTHOR CONTRIBUTIONS

MM, MS, and RK conceived and designed the study. VB, PD, SL, VS, NM, VP, and SN collected blood samples. PV, AS, and MT performed qRT-PCR experiments. PV, AS, MT, and MM conducted bioinformatics and data analysis. MS and MM drafted the manuscript. PD, SL, and VB supported in animal selection and data analysis.

FUNDING

The work was supported by the Indian Council of Agricultural Research, New Delhi, under the National Fellow Scheme (Grant number: 27 (3)/2010-HRD).

ACKNOWLEDGMENTS

The authors duly acknowledge the Director of NBAGR for providing research facilities for this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.730599/full#supplementary-material

REFERENCES

Aggarwal, J., Sharma, A., Kishore, A., Mishra, B. P., Yadav, A., Mohanty, A., et al. (2013). Identification of Suitable Housekeeping Genes for Normalization of Quantitative Real-Time PCR Data during Different Physiological Stages of Mammary Gland in Riverine Buffaloes (Bubalus Bubalis). *J. Anim. Physiol. Anim. Nutr.* 97 (6), 1132–1141. doi:10.1111/jpn.12027

Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: a Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and colon Cancer Data Sets. *Cancer Res.* 64 (15), 5245–5250. doi:10.1158/0008-5472.CAN-04-0496

Bionaz, M., and Loor, J. J. (2007). Identification of Reference Genes for Quantitative Real-Time PCR in the Bovine Mammary Gland during the Lactation Cycle. *Physiol. Genomics* 29 (3), 312–319. doi:10.1152/physiolgenomics.00223.2006

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* 55, 611–622. doi:10.1373/clinchem.2008.112797

Bustin, S. A. (2010). Why the Need for qPCR Publication Guidelines?-The Case for MIQE. *Methods* 50 (4), 217–226. doi:10.1016/j.ymeth.2009.12.006

Bustin, S. (2002). Quantification of mRNA Using Real-Time Reverse Transcription PCR (RT-PCR): Trends and Problems. *J. Mol. Endocrinol.* 29, 23–39. doi:10.1677/jme.0.0290023

Castigliego, L., Armani, A., Li, X., Grifoni, G., Gianfaldoni, D., and Guidi, A. (2010). Selecting Reference Genes in the white Blood Cells of Buffalos Treated with Recombinant Growth Hormone. *Anal. Biochem.* 403 (1-2), 120–122. doi:10.1016/j.ab.2010.04.001

Crookenden, M. A., Walker, C. G., Kuhn-Sherlock, B., Murray, A., Dukkipati, V. S. R., Heiser, A., et al. (2017). Technical Note: Evaluation of Endogenous Control Gene Expression in Bovine Neutrophils by Reverse-Transcription Quantitative PCR Using Microfluidics Gene Expression Arrays. *J. Dairy Sci.* 100 (8), 6763–6771. doi:10.3168/jds.2016-12460

Die, J. V., Baldwin, R. L., Rowland, L. J., Li, R., Oh, S., Li, C., et al. (2017). Selection of Internal Reference Genes for Normalization of Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR) Analysis in the Rumen Epithelium. *Plos one* 12 (2), e0172674. doi:10.1371/journal.pone.0172674

Engdahl, E., Dunn, N., and Fogdell-Hahn, A. (2016). Investigation of Reference Gene Expression during Human Herpesvirus 6B Infection Indicates Peptidylprolyl Isomerase A as a Stable Reference Gene and TATA Box Binding Protein as a Gene Up-Regulated by This Virus. *J. Virol. Methods* 227, 47–49. doi:10.1016/j.jviromet.2015.10.011

Hsiao, L.-L., Dangond, F., Yoshida, T., Hong, R., Jensen, R. V., Misra, J., et al. (2001). A Compendium of Gene Expression in normal Human Tissues. *Physiol. Genomics* 7, 97–104. doi:10.1152/physiolgenomics.00040.2001

Huang, L., Yan, H., Jiang, X., Yin, G., Zhang, X., Qi, X., et al. (2014). Identification of Candidate Reference Genes in Perennial Ryegrass for Quantitative RT-PCR under Various Abiotic Stress Conditions. *PloS one* 9 (4), e93724. doi:10.1371/journal.pone.0093724

Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR Normalisation; Strategies and Considerations. *Genes Immun.* 6 (4), 279–284. doi:10.1038/sj.gene.6364190

Jatav, P., Sodhi, M., Sharma, A., Mann, S., Kishore, A., Shandilya, U. K., et al. (2016). Identification of Internal Control Genes in Milk-Derived Mammary Epithelial Cells during Lactation Cycle of Indian Zebu Cow. *Anim. Sci. J.* 87 (3), 344–353. doi:10.1111/asj.12384

Kapila, N., Kishore, A., Sodhi, M., Sharma, A., Kumar, P., Mohanty, A. K., et al. (2013). Identification of Appropriate Reference Genes for qRT-PCR Analysis of Heat-Stressed Mammary Epithelial Cells in Riverine Buffaloes (*Bubalus Bubalis*). *ISRN Biotechnol.* 2013, 1–9. doi:10.5402/2013/735053

Kaur, R., Sodhi, M., Sharma, A., Sharma, V. L., Verma, P., Swami, S. K., et al. (2018). Selection of Suitable Reference Genes for Normalization of Quantitative RT-PCR (RT-qPCR) Expression Data across Twelve Tissues of Riverine Buffaloes (*Bubalus Bubalis*). *PloS one* 13 (3), e0191558. doi:10.1371/journal.pone.0191558

Kim, H. K., and Yun, S. H. (2011). Evaluation of Potential Reference Genes for Quantitative RT-PCR Analysis in Fusarium Graminearum under Different Culture Conditions. *Plant Pathol. J.* 27 (301), 9. doi:10.5423/ppj.2011.27.4.301

Kishore, A., Sodhi, M., Khate, K., Kapila, N., Kumari, P., and Mukesh, M. (2013). Selection of Stable Reference Genes in Heat Stressed Peripheral Blood Mononuclear Cells of Tropically Adapted Indian Cattle and Buffaloes. *Mol. Cell Probes* 27 (3-4), 140–144. doi:10.1016/j.mcp.2013.02.003

Kishore, A., Sodhi, M., Kumari, P., Mohanty, A. K., Sadana, D. K., Kapila, N., et al. (2014). Peripheral Blood Mononuclear Cells: a Potential Cellular System to Understand Differential Heat Shock Response across Native Cattle (*Bos indicus*), Exotic Cattle (*Bos taurus*), and Riverine Buffaloes (*Bubalus Bubalis*) of India. *Cell Stress and Chaperones* 19 (5), 613–621. doi:10.1007/s12192-013-0486-z

Lagah, S. V., Sood, T. J., Palta, P., Mukesh, M., Chauhan, M. S., Manik, R. S., et al. (2019). Selection of Reference miRNAs for Relative Quantification in Buffalo (*Bubalus Bubalis*) Blastocysts Produced by Hand-Made Cloning and *In Vitro* Fertilization. *Cell reprogramming* 21 (4), 200–209. doi:10.1089/cell.2019.0022

Li, J., Jia, H., Han, X., Zhang, J., Sun, P., Lu, M., et al. (2016). Selection of Reliable Reference Genes for Gene Expression Analysis under Abiotic Stresses in the Desert Biomass Willow, Salix Psammophila. *Front. Plant Sci.* 7, 1505. doi:10.3389/fpls.2016.01505

Lisowski, P., Pierzchała, M., Gościk, J., Pareek, C. S., and Zwierzchowski, L. (2008). Evaluation of Reference Genes for Studies of Gene Expression in the Bovine Liver, Kidney, Pituitary, and Thyroid. *J. Appl. Genet.* 49 (4), 367–372. doi:10.1007/BF03195635

Lü, J., Yang, C., Zhang, Y., and Pan, H. (2018b). Selection of Reference Genes for the Normalization of RT-qPCR Data in Gene Expression Studies in Insects: a Systematic Review. *Front. Physiol.* 9, 1560. doi:10.3389/fphys.2018.01560

Lu, P., Magwanga, R., Lu, H., Kirungu, J., Wei, Y., Dong, Q., et al. (2018a). A Novel G-Protein-Coupled Receptors Gene from upland Cotton Enhances Salt Stress Tolerance in Transgenic Arabidopsis. *Genes* 9 (4), 209. doi:10.3390/genes9040209

Luo, M., Gao, Z., Li, H., Li, Q., Zhang, C., Xu, W., et al. (2018). Selection of Reference Genes for miRNA qRT-PCR under Abiotic Stress in grapevine. *Sci. Rep.* 8 (1), 1–11. doi:10.1038/s41598-018-22743-6

Ma, R., Xu, S., Zhao, Y., Xia, B., and Wang, R. (2016). Selection and Validation of Appropriate Reference Genes for Quantitative Real-Time PCR Analysis of Gene Expression in Lycoris Aurea. *Front. Plant Sci.* 7, 536. doi:10.3389/fpls.2016.00536

Pradeep, J., Monika, S., Ankita, S., Umesh, K. S., Amit, K., Ashok, M., et al. (2014). Expression Analysis of Solute Carrier (SLC2A) Genes in Milk Derived Mammary Epithelial Cells during Different Stages of Lactation in Sahiwal (*Bos indicus*) Cows. Research Article. *J. Adv. Dairy Res.* 2 (117), 2.

Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., and Nitsche, A. (2004). Guideline to Reference Gene Selection for Quantitative Real-Time PCR. *Biochem. Biophys. Res. Commun.* 313 (4), 856–862. doi:10.1016/j.bbrc.2003.11.177

Rasmussen, R. (2001). "Quantification on the LightCycler," in *Rapid Cycle Real-Time PCR* (Berlin, Heidelberg: Springer), 21-34. doi:10.1007/978-3-642-59524-0_3

Sang, J., Wang, Z., Li, M., Cao, J., Niu, G., Xia, L., et al. (2018). ICG: a Wiki-Driven Knowledgebase of Internal Control Genes for RT-qPCR Normalization. *Nucleic Acids Res.* 46 (D1), D121–D126. doi:10.1093/nar/gkx875

Sharma, P., Sharma, A., Sodhi, M., Verma, P., Parvesh, K., Swami, S. K., et al. (2019). Characterizing Binding Sites of Heat Responsive microRNAs and Their Expression Pattern in Heat Stressed PBMCs of Native Cattle, Exotic Cattle and Riverine Buffaloes. *Mol. Biol. Rep.* 46 (6), 6513–6524. doi:10.1007/s11033-019-05097-8

Sood, T. J., Lagah, S. V., Sharma, A., Singla, S. K., Mukesh, M., Chauhan, M. S., et al. (2017). Selection of Suitable Internal Control Genes for Accurate Normalization of Real-Time Quantitative PCR Data of buffalo (Bubalus Bubalis) Blastocysts Produced by SCNT and IVF. *Cell Reprogramming* 19 (5), 302–310. doi:10.1089/cell.2017.0010

Thomas, K. C., Zheng, X. F., Garces Suarez, F., Raftery, J. M., Quinlan, K. G. R., Yang, N., et al. (2014). Evidence Based Selection of Commonly Used RT-qPCR Reference Genes for the Analysis of Mouse Skeletal Muscle. *PloS one* 9 (2), e88653. doi:10.1371/journal.pone.0088653

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes. *Genome Biol.* 3 (7), research0034–1. doi:10.1186/gb-2002-3-7-research0034

Verma, P., Sharma, A., Sodhi, M., Thakur, K., Bharti, V. K., Kumar, P., et al. (2018b). Overexpression of Genes Associated with Hypoxia in Cattle Adapted to Trans Himalayan Region of Ladakh. *Cell. Biol. Int.* 42 (9), 1141–1148. doi:10.1002/cbin.10981

Verma, P., Sharma, A., Sodhi, M., Thakur, K., Kataria, R. S., Niranjan, S. K., et al. (2018a). Transcriptome Analysis of Circulating PBMCs to Understand Mechanism of High Altitude Adaptation in Native Cattle of Ladakh Region. *Sci. Rep.* 8 (1), 1–15. doi:10.1038/s41598-018-25736-7

Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: a miRNA Analysis Tool for Deep Sequencing of Plant Small RNAs. *Plant Mol. Biol.* 80 (1), 75–84. doi:10.1007/s11103-012-9885-2

Xu, H., Li, C., Zeng, Q., Agrawal, I., Zhu, X., and Gong, Z. (2016). Genome-wide Identification of Suitable Zebrafish *Danio rerio* Reference Genes for Normalization of Gene Expression Data by RT-qPCR. *J. Fish. Biol.* 88 (6), 2095–2110. doi:10.1111/jfb.12915

Yan, Q., Cui, X., Lin, S., Gan, S., Xing, H., and Dou, D. (2016). GmCYP82A3, a Soybean Cytochrome P450 Family Gene Involved in the Jasmonic Acid and Ethylene Signaling Pathway, Enhances Plant Resistance to Biotic and Abiotic Stresses. *PloS one* 11 (9), e0162253. doi:10.1371/journal.pone.0162253

Yang, Q., Yin, J., Li, G., Qi, L., Yang, F., Wang, R., et al. (2014). Reference Gene Selection for qRT-PCR in Caragana Korshinskii Kom. Under Different Stress Conditions. *Mol. Biol. Rep.* 41 (4), 2325–2334. doi:10.1007/s11033-014-3086-9

Zhao, D.-J., Guo, K., and Kang, L. (2012). Identification of Condition-specific Reference Genes from Microarray Data for Locusts Exposed to Hypobaric Hypoxia. *FEBS open bio* 2, 235–240. doi:10.1016/j.fob.2012.08.001

Zhu, W., Lin, Y., Liao, H., and Wang, Y. (2015). Selection of Reference Genes for Gene Expression Studies Related to Intramuscular Fat Deposition in *Capra hircus* Skeletal Muscle. *PloS one* 10 (3), e0121280. doi:10.1371/journal.pone.0121280

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership