frontiers | Research Topics

Traditional and up-to-date genomic insights into domestic animal diversity

Edited by
Johann Sölkner, Michael N. Romanov, Natalia A. Zinovieva, Steffen Weigend and Klaus Wimmers

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-83251-337-8
DOI 10.3389/978-2-83251-337-8

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

Traditional and up-to-date genomic insights into domestic animal diversity

Topic editors

Johann Sölkner — University of Natural Resources and Life Sciences Vienna, Austria
Michael N. Romanov — University of Kent, United Kingdom
Natalia A. Zinovieva — L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia
Steffen Weigend — Institute of Farm Animal Genetics, Friedrich Loeffler Institute (FLI), Germany
Klaus Wimmers — Research Institute for Farm Animal Biology (FBN), Germany

Citation

Sölkner, J., Romanov, M. N., Zinovieva, N. A., Weigend, S., Wimmers, K., eds. (2023). *Traditional and up-to-date genomic insights into domestic animal diversity*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-337-8

Table of contents

05 Editorial: Traditional and up-to-date genomic insights into domestic animal diversity
Michael N. Romanov, Johann Sölkner, Natalia A. Zinovieva, Klaus Wimmers and Steffen Weigend

12 Selection of Cashmere Fineness Functional Genes by Translatomics
Yu Zhang, Dongyun Zhang, Yanan Xu, Yuting Qin, Ming Gu, Weidong Cai, Zhixian Bai, Xinjiang Zhang, Rui Chen, Yingang Sun, Yanzhi Wu and Zeying Wang

29 Integrative Analysis of the lncRNA-Associated ceRNA Regulatory Network Response to Hypoxia in Alveolar Type II Epithelial Cells of Tibetan Pigs
Yanan Yang, Yongqing Li, Haonan Yuan, Xuanbo Liu, Yue Ren, Caixia Gao, Ting Jiao, Yuan Cai and Shengguo Zhao

40 Resequencing and Signatures of Selective Scans Point to Candidate Genetic Variants for Hair Length Traits in Long-Haired and Normal-Haired Tianzhu White Yak
Qi Bao, Xiaoming Ma, Congjun Jia, Xiaoyun Wu, Yi Wu, Guangyao Meng, Pengjia Bao, Min Chu, Xian Guo, Chunnian Liang and Ping Yan

51 Multi-Omics Reveals Different Strategies in the Immune and Metabolic Systems of High-Yielding Strains of Laying Hens
Muhammad Arsalan Iqbal, Henry Reyer, Michael Oster, Frieder Hadlich, Nares Trakooljul, Alvaro Perdomo-Sabogal, Sonja Schmucker, Volker Stefanski, Christoph Roth, Amélia Camarinha Silva, Korinna Huber, Vera Sommerfeld, Markus Rodehutscord, Klaus Wimmers and Siriluck Ponsuksili

71 Identification of *SSTR5* Gene Polymorphisms and Their Association With Growth Traits in Hulun Buir Sheep
Xue Li, Ning Ding, Zhichao Zhang, Dehong Tian, Buying Han, Dehui Liu, Sijia Liu, Fei Tian, Dejun Fu, Xiaoliang Song and Kai Zhao

80 Genome-Wide Identification and Characterization of Long Non-Coding RNAs in Longissimus dorsi Skeletal Muscle of Shandong Black Cattle and Luxi Cattle
Ruili Liu, Mingxuan Han, Xianxun Liu, Kun Yu, Xuejin Bai and Yajuan Dong

97 Genetics of Base Coat Colour Variations and Coat Colour-Patterns of the South African Nguni Cattle Investigated Using High-Density SNP Genotypes
Langelihle Mbali Kunene, Farai Catherine Muchadeyi, Khanyisile Hadebe, Gábor Mészáros, Johann Sölkner, Trevor Dugmore and Edgar Farai Dzomba

109 Genome-Wide Analysis of microRNAs Identifies the Lipid Metabolism Pathway to Be a Defining Factor in Adipose Tissue From Different Sheep
Tian-Yi Liu, Hui Feng, Salsabeel Yousuf, Ling-Li Xie and Xiang-Yang Miao

122 Opportunities of Genomics for the Use of Semen Cryo-Conserved in Gene Banks
J. Kor Oldenbroek and Jack J. Windig

132 High-Density Genomic Characterization of Native Croatian Sheep Breeds
Ivana Drzaic, Ino Curik, Boris Lukic, Mario Shihabi, Meng-Hua Li, Juha Kantanen, Salvatore Mastrangelo, Elena Ciani, Johannes A. Lenstra and Vlatka Cubric-Curik

144 Time-series transcriptome analysis identified differentially expressed genes in broiler chicken infected with mixed *Eimeria* species
Minjun Kim, Yoonji Chung, Prabuddha Manjula, Dongwon Seo, Sunghyun Cho, Eunjin Cho, Thisarani Kalhari Ediriweera, Myunghwan Yu, Sunju Nam and Jun Heon Lee

155 Conservation status and historical relatedness of South African communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker
T. C. Chokoe, K. Hadebe, F. C. Muchadeyi, K. A. Nephawe, E. F. Dzomba, T. D. Mphahlele, T. C. Matelele and B. J. Mtileni

171 Identification of potential miRNA-mRNA regulatory network and the key miRNAs in intramuscular and subcutaneous adipose
Hui Feng, Tianyi Liu, Salsabeel Yousuf, Xiuxiu Zhang, Wanlong Huang, Ai Li, Lingli Xie and Xiangyang Miao

184 Genetic diversity and population structure of Tongcheng pigs in China using whole-genome SNP chip
Jiao Yuan, Xiang Zhou, Guoqiang Xu, Sanping Xu and Bang Liu

195 The impact of using different ancestral reference populations in assessing crossbred population admixture and influence on performance
Mohd A. Jaafar, Bradley J. Heins, Chad Dechow and Heather J. Huson

211 An insight into the runs of homozygosity distribution and breed differentiation in Mangalitsa pigs
Sowah Addo and Lisa Jung

227 Application of next-generation sequencing for the high-resolution typing of MHC-*B* in Korean native chicken
Thisarani Kalhari Ediriweera, Prabuddha Manjula, Eunjin Cho, Minjun Kim and Jun Heon Lee

235 Selection signature analysis and genome-wide divergence of South African Merino breeds from their founders
E. F. Dzomba, M. A. Van Der Nest, J. N. T. Mthembu, P Soma, M. A. Snyman, M. Chimonyo and F. C. Muchadeyi

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1117708

OPEN ACCESS

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

*CORRESPONDENCE
Michael N. Romanov,
m.romanov@kent.ac.uk

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

RECEIVED 06 December 2022
ACCEPTED 16 December 2022
PUBLISHED 04 January 2023

CITATION
Romanov MN, Sölkner J, Zinovieva NA, Wimmers K and Weigend S (2023), Editorial: Traditional and up-to-date genomic insights into domestic animal diversity.
Front. Genet. 13:1117708.
doi: 10.3389/fgene.2022.1117708

Editorial: Traditional and up-to-date genomic insights into domestic animal diversity

Michael N. Romanov[1]*, Johann Sölkner[2], Natalia A. Zinovieva[3], Klaus Wimmers[4] and Steffen Weigend[5]

[1]School of Biosciences, University of Kent, Canterbury, United Kingdom, [2]Institute of Livestock Sciences (NUWI), University of Natural Resources and Life Sciences Vienna, Vienna, Austria, [3]L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Moscow, Russia, [4]Research Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, [5]Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut (FLI), Neustadt, Germany

KEYWORDS

domestic animal diversity, phenotypic traits, genomics, SNP genotyping, transcriptome, candidate genes, selective sweep, disease resistance

Editorial on the Research Topic
Traditional and up-to-date genomic insights into domestic animal diversity

Domesticated animals play a significant role in local, national, and international agricultural output as well as in daily human life and culture. Additionally, they make up a sizeable portion of the biodiversity of the planet, which is essential for producing food and other animal products for human consumption. The present *Frontiers in Genetics* Research Topic (Figure 1) is devoted to various issues pertinent diversity of farm animals. The latter is at serious risk today, which could result in a reduction in the resources available to produce breed-specific food products and other necessities of everyday living. Importantly, genetic diversity is necessary for future animal breeding to be flexible enough to adapt livestock populations to changing customer demands and climatic conditions. Continued efforts are required to protect biodiversity, stop the loss of animal breeds, and maintain genetic diversity and develop strategies to use resource population in regional (niche) production systems.

One of conventional ways to preserve the domestic animal diversity is the use of semen cryo-conserved in gene banks that also opens plentiful opportunities for genomics studies as reviewed by Oldenbroek and Windig. Gene banks were established not long after the beginning of implementation of cryo-conserved semen in the principal farm animal species. The fundamental goal was, and still is, preserving the genetic variation of agricultural animals for future use. DNA data from animals in living populations as well as from sires that have been cryopreserved is now accessible. Combining their DNA data opens up three potential avenues: 1) expanding the gene bank's Research Topic of genetic diversity; 2) tracing the

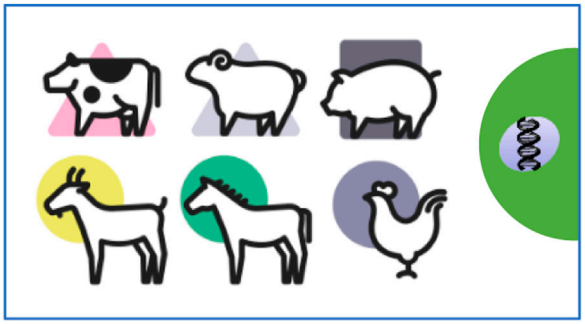


FIGURE 1
An artistic "logotype" for the *Frontiers in Genetics* Research Topic "Traditional and Up-to-date Genomic Insights into Domestic Animal Diversity".

evolution of the genetic diversity from living populations; and 3) enhancing the genetic diversity and performance of existing populations. These three possibilities for the use of gene bank sires in the genomic era have been detailed in many studies demonstrating the immense significance of a gene bank as a library of genetic variation.

For conservation and preservation measures, the identification and evaluation of important genetic resources is necessary. This requires a comprehensive characterization of populations at risk of loss. This includes recording the phenotypes and especially those traits for which superiority and potential usefulness of the populations is anticipated. Characterization also includes the determination of genetic variability and distance. Genetic markers are used for this purpose. Advances in genome sequencing, the availability of genotyping tools and techniques such as SNP microarrays and genotyping-by-sequencing make such analyses simple, fast and informative. Ongoing progress in genome annotation increasingly enables the assessment of the functional significance of polymorphisms. The use of holistic "omics" techniques, especially next-generation sequencing, also enables a comprehensive characterization of genetic resources along the genotype–phenotype map at the genome, epigenome, transcriptome, proteome and metabolome levels. Such analyses not only contribute to the comprehensive characterization of genetic resources, but also allow the use of selected populations with specific traits, e.g., adaptation to environmental and climatic stresses or pathogens, to facilitate the understanding of the expression of functional traits and adaptation mechanisms to abiotic and biotic stressors, as shown in several studies in this Research Topic. Old, local, site-adapted breeds can thus serve as valuable models. A number of good research publications as overviewed below by species have been provided within the present Research Topic that address conventional methodologies, in addition to genotypic and genomic data, and more recent developments in the study and conservation of domestic animal diversity.

Cattle

Kunene et al. used high-density single nucleotide polymorphism (SNP) genotypes to explore the genetics of base coat color variations and coat color patterns in South African Nguni cattle. The Nguni breed of cattle, which is similar to the Sanga, has mixed *B. taurus* and *B. indicus* lineage and has been shown to be resistant to ticks, illnesses, and other challenging environmental factors found in Africa. The leather business has developed a specialty market for the multicolored Nguni coats, which has prompted breeding goals for the propagation of such diversity. Limited research has been done on the genetic architecture that underlies coat color and pattern, which is a barrier to future breeding and development of that characteristic. Using Illumina Bovine HD (770K) genotypes and coat color phenotyped Nguni cattle, the authors examined genes underlying the base coat color, color-sidedness, and white forehead stripe in this cattle. Four indicative SNPs were identified on BTA18 as a result of the genome-wide association studies (GWAS) for base coat color (eumelanin vs. pheomelanin). A well-known gene, *MC1R*, was found to be located within 1 MB of the indicative SNPs, and it was discovered to be involved in both the mitogen-activated protein kinase (MAPK) signaling pathway and melanogenesis, the core pathway for the production of melanin. Four suggestive SNPs were discovered by GWAS for the color-sidedness gene, although none of them were located near the KIT candidate gene, which is linked to color-sidedness. Seventeen suggestive SNPs were found on BTA6 as a result of GWAS for the white forehead stripe. Four genes, including *MAPK10*, *EFNA5*, *PPP2R3C*, and *PAK1*, were discovered to be connected to the white forehead stripe and to the MAPK, adrenergic, and Wnt signaling pathways, which are mutually connected to melanin formation. These findings supported past theories about the function of *MC1R* in base coat colors in cattle and proposed a separate genetic mechanism for the phenotypes of Nguni cattle with forehead stripes.

Jaafar et al. evaluated the effects of employing several ancestral reference populations on the population admixture and performance of crossbred cattle. Animals from several breeds are bred together during the process of crossbreeding. The offspring show a combination of genetic improvements from the parental breeds that boost heterozygosity and counteract inbreeding depression, both additive and non-additive. However, because the advantages of heterosis rely on the type of crossbreeding systems utilized and the heritability of the traits, crossbreeding may also disrupt the special and frequently advantageous gene combinations in parental breeds, thus lowering performance potential. Regarding three-breed crossbreeding systems, it is yet unclear how crossbreeding

affects the genomic architecture in particular. In order to compare genomic ancestry estimates to pedigree-based estimates, this study connected breed composition with important production and health parameters among two rotational crossbred populations, ProCROSS and Grazecross. Rotational crossbreeding of the Viking Red (VKR; a marketing name for Swedish Red, Danish Red, and Finnish Ayrshire breeds), Holstein, and Montbeliarde led to the creation of ProCROSS. In contrast, Grazecross was made up of VKR, Normande, and Jersey. Both breeding initiatives attempted to maximize heterosis' beneficial effects. All genomic estimations considered, choosing the most appropriate and useful animals to use as the reference animals in admixture analysis is important when interpreting relationship and population structure results, but there is some uncertainty when determining how breed composition relates to phenotypic performance.

Long non-coding RNAs (lncRNAs) were identified and described by Liu et al. on a genome-wide scale in the *longissimus dorsi* skeletal muscle of Shandong Black and Luxi cattle. LncRNAs may play a regulatory role, which is becoming increasingly clear. The regulation of cell differentiation, fat synthesis, and embryonic development have been the main Research Topic of studies on cattle. However, there has not been much research on the potential function of lncRNAs in the skeletal muscle of domestic animals. Here, bioinformatics analysis was employed to build a network of lncRNAs, miRNAs, and mRNA interactions connected to muscle using the transcriptome numbers of distinct beef cattle, Shandong Black and Luxi. This can be utilized to advance animal husbandry, increase animal husbandry output, and elucidate the molecular basis of the growth of bovine muscle. A total of 1,415 transcripts (of which 480 were lncRNAs) were differently expressed in the two breeds. Furthermore, 1,164 protein-coding genes (*MYORG*, *Wnt4*, *PAK1*, *ADCY7*, etc.) were the targets of the most differentially expressed lncRNAs, which were located on chromosome 9. A probable trans-regulatory link between the differentially expressed lncRNAs and 43,844 mRNAs was also shown. The detected co-expressed mRNAs (*MYORG*, *Dll1*, *EFNB2*, *SOX6*, *MYOCD*, and *MYLK3*) are enriched in calcium and AMPK signaling, muscle cell and striated muscle tissue development, and strained muscle cell differentiation. A network of lncRNAs, miRNAs, and mRNA interactions was built as the putative foundation for biological control in the skeletal muscle of cattle based on this. The reported findings will theoretically support future research on lncRNA regulation and activity in various cattle breeds.

Yak

Bao et al. used resequencing to examine the signals of selective scanning leading to potential genetic variations for hair length features in Tianzhu white yaks with long and normal hair. The Tianzhu white yak is a unique native yak breed in China that has an all-white coat. Breeders have recently identified long-haired individuals of the Tianzhu white yak, which are distinguished by long hair on the forehead. The length and density of the hair on these two sections of the body are also higher than that of the typical Tianzhu white yak. The authors re-sequenced the whole genomes of long-haired and normal Tianzhu White yaks to clarify the genetic basis of hair length in Tianzhu white yaks. Two hotspots were discovered on chromosome 6 that contain two (*FGF5* and *CFAP299*) and four (*ATP8A1*, *SLC30A9*, *SHISA3*, and *TMEM33*) genes, respectively. Ras, MAPK, PI3K-Akt and Rap1 signaling pathways were found to be involved in the process of hair length variation by function enrichment analysis of genes in two hotspots. In addition, four more genes (*ACOXL*, *PDPK1*, *MAGEL2*, and *CDH1*) were discovered as connected to the growth of hair follicles.

Pig

Using a whole-genome SNP chip, Yuan et al. investigated the genetic diversity and population dynamics of Tongcheng (TC) pigs. Indigenous to China, TC pigs are known for their high meat quality. Due to the introduction of global pig breeds and the African swine fever pathogen, the genetic resources of TC pigs are now seriously threatened. The current study used multiple SNP markers to analyze the genetic diversity and population structure of TC pigs in order to support their management and conservation. With an average linkage disequilibrium (LD) value of .15, LD and neutrality testing both showed a low selection of TC pigs. Estimates of minor allele frequency, observed heterozygosity (H_o), expected heterozygosity, and nucleotide diversity values pointed to the TC pigs' astonishingly great genetic diversity. Additionally, runs of homozygosity (ROHs) segments were examined in the whole genome of TC pigs. Based on ROHs, the average genomic inbreeding coefficient F_{ROH} was .04%. On nine separate autosomes, 14 ROH islands with 240 genes were discovered. Some of them, including *FFAR2*, *FFAR4*, *MAPK8*, *NPY5R*, and *KISS1*, overlapped with genes involved in immunological response, reproduction, muscle development, and fat deposition. These genes may be linked to qualities like meat quality and disease resistance in TC pigs. Genetic diversity and population structure data together revealed that the TC pig was a valuable genetic resource. To provide sufficient genetic diversity and prevent inbreeding depression, the TC pig breed conservation program needs to be further developed, and this research gives management and conservation methods for TC pigs a theoretical foundation.

Insights on the ROH distribution and breed differentiation in Mangalitsa pigs were provided by Addo and Jung. Mangalitsa pigs have three different color patterns on their coats, according

to which they can be classified as Red, Blond, or Swallow-bellied. The current work used studies of population structure, ROHs, and fixation index to examine genome-wide diversity and selection fingerprints in the three breeds genotyped using a modified ProcineSNP60 v2 Genotyping Bead Chip. Also provided as comparison outgroup data were 20 genotypes of the Hungarian Mangalitsa. For the Blond, Swallow-bellied, and Red Mangalitsa, respectively, estimates of observed heterozygosity were .27, .28, and .29, and estimates of inbreeding coefficients based on ROHs were 24.11%, 20.82%, and 16.34%. All breeds had ROH islands, but none of them were shared by any of the breeds. In a ROH island in the Swallow-bellied Mangalitsa, the *KIF16B* gene—previously known to be involved in synaptic signaling—was discovered. The same gene was discovered to contain a significantly different SNP (MARC0032380) when comparing Swallow-belied Mangalitsa to either Red or Blond. Some ROH islands in the Red Mangalitsa were connected to genes like *ABCA12*, *VIL1*, *PLSCR5*, and *USP37* that affect meat quality attributes. This research revealed that the variation and population structure of the three breeds were distinct, with the Red and Blond Mangalitsa pigs being the most closely related.

In a Chinese pig breed, Feng et al. discovered a putative miRNA-mRNA regulation network and the important miRNAs in intramuscular and subcutaneous adipose. Intramuscular fat (IMF) is a key metric for assessing the quality of meat. Breeds with high IMF content frequently also have high subcutaneous fat, which negatively impacts pigs' ability to produce meat. Important ramifications for pig breeding result from research into the processes of miRNAs involved in lipogenesis and lipid metabolism. Here, the patterns of lipogenesis in the Chinese breed of pig known as Laiwu were analyzed by creating two small RNA libraries from intramuscular and subcutaneous fat. Two types of adipose tissue were used to identify a total of 286 differentially expressed miRNAs (DEmiRNAs), comprising 193 known miRNA and 93 novel miRNA. Gene ontology (GO) and KEGG enrichment analysis for DEmiRNAs revealed that their target genes were involved in numerous biological processes and signaling pathways related to adipogenesis and lipid metabolism, including the Wnt, MAPK, Hippo and PI3K-Akt signaling pathways, melanogenesis, and signaling pathways controlling stem cell pluripotency, among others. After that, a network of interactions between miRNA and mRNA was built to determine which miRNAs were crucial for regulating the Wnt signaling pathway. MiR-331-3p, miR-339-5p, miR-874, and novel346 mature target *PPARD*, *WNT10B*, *RSPO3*, and *WNT2B* in this pathway. This research offers a theoretical foundation for future research into the post-transcriptional control mechanism of meat quality generation as well as disease diagnosis and management related to ectopic fat.

An integrative analysis of the alveolar type II epithelial (ATII) cells of Tibetan pigs and their response to hypoxia was conducted by Yang et al. with respect to the lncRNA-associated ceRNA regulation network. Understanding the regulatory mechanisms governing responses to hypoxia may help in alleviating harm brought on by hypoxia. ATII cells' ability to function is significantly hindered by oxygen deprivation. In this study, ATII cells were cultivated from Tibetan and Landrace pigs in hypoxic and normoxic conditions to search for differentially expressed (DE) lncRNAs and DEmiRNAs as well as build their related ceRNA regulation networks in response to hypoxia. Target genes of Tibetan and Landrace pig DElncRNAs were significantly enriched in the proteoglycans in cancer, renal cell carcinoma, and erbB signaling pathways between the normoxic and hypoxic groups, whereas DEmiRNAs' target genes were significantly enriched in the axon guidance, focal adhesion, and MAPK signaling pathways. Through the activation of the focal adhesion/PI3K-Akt/glycolysis pathway, hypoxia induction has been demonstrated to potentially encourage apoptosis. By controlling ATII cell autophagy in normoxic and hypoxic conditions, the ssc-miR-20b/MSTRG.57127.1/ssc-miR-7-5p axis may have significantly reduced hypoxia injury. The most impacted axis, MSTRG.14861.4-miR-11971-z-CCDC12, controlled a number of RNAs and may thus control the proliferation of ATII cells in Tibetan pigs under hypoxic settings. In Landrace pigs, the ACTA1/ssc-miR-30c-3p/MSTRG.23871.1 axis plays a critical role in reducing ATII cell damage and enhancing dysfunction and fibrosis brought on by oxidative stress. These findings give a better knowledge of how Tibetan pigs regulate their lncRNA, miRNA, and mRNA in hypoxic environments.

In Hulun Buir sheep, Li et al. discovered *SSTR5* gene polymorphisms and their correlation with growth traits. This investigation sought to identify *SSTR5* polymorphisms and assess their relationship to growth parameters in Hulun Buir sheep. Seven SNPs were found as a result of Sanger sequencing, showed moderate polymorphism (.25 < PIC <.5), and were then subject to association analysis in relation to the growth traits. At 9 months of age, cannon circumference was substantially related with SNP4 (rs605867745) and SNP3 (rs413380618). There was linkage disequilibrium among the five haplotypes and seven SNPs. These haplotypes were not connected to distinct ages of growth features, nevertheless. SNP1, SNP3, SNP4, and SNP7 may all function as molecular markers for the growth features of Hulun Buir sheep, to sum up.

A genome-wide divergence and selection signature investigation of South African Merino-derived breeds from their ancestors was carried out by Dzomba et al. Merino sheep are a preferred breed that are widely raised for their wool and mutton worth. Using the Illumina Ovine50K BeadChip, this study assessed genetic diversity, population structure, and breed divergence in the South African Merino (SAM), eight Merino-based sheep breeds, as well as non-Merino founding breeds (Damara, Ronderib Afrikaner, and Nguni). The Meatmaster, SAM and Dohne Merino (DM) showed the highest

genetic diversity levels, with H_o values of .37–.39. The degree of inbreeding varied from zero (DM) to .27 points (Nguni). High within population variance (80 + per cent) was observed across all population categories. Selection sweeps for the Afrino (12 sweeps), Meatmaster (four sweeps), and DM (29 sweeps) were identified. Such genes as *FGF12*, the metabolic genes *ICA1*, *NXPH1*, and *GPR171*, as well as the immune response genes *IL22*, *IL26*, *IFNAR1*, and *IL10RB*, have all been linked to hair and wool features. The DM vs. Merino, Meatmaster vs. Merino and Meatmaster vs. Nguni shared a selection sweep on chromosome 10 harboring the *RXFP2* gene for the polledness. Additionally, the DM vs. Merino and the Meatmaster vs. Merino shared a Rsb-based selection sweep on chromosome 1 connected to the *CAPN7* gene for calpain. Collectively, the analysis showed some genetic divergence caused by breed-specific selection objectives and some genetic similarity between the Merino and Merino-derived breeds originating from shared founder populations.

High-density genomic characterization of native Croatian sheep breeds was performed by Drzaic et al. Using 50K SNP profiles, this comprehensive genomic research has revealed that the regional Balkan sheep populations share a great deal of genetic variation with neighboring breeds, but they are also very distinct from them. Using the Ovine Infinium® HD SNP BeadChip, eight Croatian sheep breeds and mouflon were genotyped. Also, various Mediterranean sheep breeds and Balkan Pramenka, which are readily accessible, were added to the analysis. This research uncovered the intricate demographic structure of Croatian sheep breeds, as well as information on their geographic origins (island vs. mainland). The historical establishment of breeds and the routes of gene flow were confirmed by migration patterns. Between sheep populations, $F_{ROH>2\ Mb}$ coefficients ranged from .025 to .070, with Dalmatian Pramenka and Pag Island Sheep having lower inbreeding coefficients and Dubrovnik sheep having higher inbreeding. For the Krk Island Sheep and Dalmatian Pramenka, the estimated effective population size (N_e) varied from 61 to 1039, respectively. In order to retain genetic variation in particular breeds, there is a need for greater conservation management due to higher inbreeding rates and a smaller N_e. These findings will aid in breeding and developing conservation plans for Croatia's indigenous sheep breeds.

A genome-wide investigation of miRNAs by Liu et al. revealed that the lipid metabolism pathway is a key characteristic of adipose tissue from various sheep. Important non-coding RNAs known as miRNAs can take part in the control of biological processes. MiRNAs have been extensively explored in recent years, not just in humans and mice but also in animal husbandry. However, the investigation of miRNA in subcutaneous adipose tissue of sheep is not thorough compared to other livestock and poultry breeds. Using RNA-Seq technology, the transcriptomes of miRNAs in the subcutaneous adipose tissue of Duolang sheep and Small Tail Han sheep were analyzed to identify those that were expressed differently in these two breeds. As a result, 38 miRNAs were discovered that were differently expressed (nine novel miRNAs and 29 known miRNAs). Additionally, 854 target genes were predicted in total. The deposition of subcutaneous adipose tissue in Duolang and Small Tail Han sheep has been linked to the regulation of lipolysis in adipocytes. The genes involved in controlling lipolysis in adipocytes may be controlled by the miRNAs, which in turn may control fat accumulation. In particular, NC_ 040278.1 37602, oar-mir-493-3p, NC_ 040278.1 37521, and NC_ 040255.1 11627 may each target *PTGS2*, *AKT2*, *AKT3*, and *PIK3CA*, thus playing a crucial regulatory role. Overall, the findings establish the groundwork for further elucidating the mechanism underlying the deposition of subcutaneous adipose tissue in sheep, enhancing the performance of their ability to produce meat, and advancing the field of animal husbandry.

Goat

Using the genome-wide Illumina goat SNP50K BeadChip, Chokoe et al. clarified the conservation status and historical relatedness of communal indigenous goat populations in South Africa. Due to their tolerance to various production situations and support for communal farming, indigenous goats, which make up the bulk of populations in smallholder, low input, low output production systems, are regarded as an important genetic resource. In order to assist breeding strategies to utilize and conserve genetic resources, N_e, inbreeding rates, and ROHs are useful tools for examining genetic diversity and comprehending the demographic history. The historical N_e across populations indicates that the ancestor N_e has decreased in Free State (FS), North West (NW), Limpopo (LP), and Gauteng (GP), respectively, over the last 971 generations. The current N_e of GP was the lowest across populations. The Eastern Cape (EC) had the lowest levels of $F_{ROH>5\ Mb}$, and FS had the greatest levels. The FS, GP and NW populations had 871 ROH island areas that contain crucial environmental adaptation and thermo-tolerance genes such *IL10RB*, *IL23A*, *FGF9*, *IGF1*, *EGR1*, and *MAPK3*. Despite having a similar ancestor, the genomes of KZN and LP exhibit significant mixing from the EC and NW populations. Using genome-wide SNP markers, the results showed that the presence of high N_e and autozygosity differed significantly across communal indigenous goat populations at recent and ancient events. The migration of communal indigenous goat populations from the northern section (LP) of South Africa to the eastern regions (KZN) showed their historical kinship and coincided with the Bantu nation's migratory phases.

Translatomics can be used to select functional genes for cashmere fineness, as Zhang et al. demonstrated. An essential metric for assessing cashmere quality is cashmere fineness.

Although the Liaoning Cashmere Goat (LCG) breed studied produces a lot of cashmere and has long cashmere fiber, its fineness may be better. Thus, it is crucial to identify genes related to cashmere fineness that might be applied in next attempts to enhance this phenotype. Through high-throughput sequencing and genome-wide association analysis, the regulation of cashmere fineness has achieved unprecedented strides due to the ongoing improvement of technology. Translatomics has been demonstrated to be able to pinpoint genes linked to phenotypic features. The authors performed translatomic analysis after having sequenced the skin tissue of LCG sample groups with various cashmere fineness ranges by Ribo-seq. Differently expressed genes were found between the sample groups using these data. From these, 186 genes were downregulated and 343 genes were upregulated in the fine LCG group as compared to the coarse LCG group. The biological functions of differential genes were explored by GO enrichment analysis, with functional genes related to the extracellular area being predominant. The enrichment of the human papillomavirus infection pathway was particularly prominent in the KEGG enrichment study. The authors suggested that the *COL6A5* gene might impact the fineness of cashmere.

Poultry

In highly productive lines of laying hens, Iqbal et al. investigated how multi-omics can disclose different strategies in the immune and metabolic systems. Due to their high egg production and outstanding commercial applicability, Lohmann Brown (LB) and Lohmann Selected Leghorn (LSL) are two major laying hen strains for the poultry industry. To further understand how the genetic makeup of the two strains effects their biological pathways, the genotype-phenotype map of the current study incorporated multiple data sets. The harvested data sets in the two strains were intestinal miRNA and mRNA transcriptome data, immune cells, inositol phosphate metabolites, minerals, and hormones from various organs. These were analyzed using the R package mixOmics. Among the most distinctive characteristics between the two strains, there were 20 miRNAs, 20 mRNAs, 16 immune cells, 10 microorganisms, 11 behavioral properties, and 16 metabolites. The enrichment of immune pathways in the LSL strain was correlated with the expression of particular miRNAs and the quantity of different immune cell types. On the other hand, more microbial taxa that were unique to the LB strain were discovered, and the prevalence of several bacteria showed a high correlation with transcripts enriched in immunological and metabolic pathways in the host gut. According to this research, both strains use unique innate mechanisms to develop and keep their immune and metabolic systems functioning well. The study also adds to our understanding of the role of host–gut interactions, including immune phenotype, microbiota, gut transcriptome, and metabolome, by offering a fresh perspective on the functional biodiversity that emerges during strain selection.

The utility of next-generation sequencing (NGS) for the high-resolution typing of major histocompatibility complex-B (MHC-B) in Korean native chickens (KNC) was proven by Ediriweera et al. The chicken MHC-B region plays a vital role in the development of the immune systems and is extremely variable across breeds, lines, and populations. It is crucial to examine this chromosomal region, particularly the class I and II genes, to ascertain the variation and diversity that eventually alter antigen presentation because it determines the resistance/susceptibility to a variety of infectious illnesses. Using NGS data, Geneious Prime-based assembly and variant calling with the Genome Analysis Toolkit (GATK) best practices pipeline, this study examined five KNC lines and the Ogye breed. For each line or breed of chicken, the consensus MHC-B (*BG1–BF2*) sequences were collected, and their variations were examined. Each of the KNC lines had an excessive number of mutations, including a sizable number of high-impact variants that revealed important details about altered MHC molecules. The study verified that MHC variations can be successfully detected using NGS techniques, and the KNC lines had a very diverse MHC-B region, indicating a significant divergence from the red junglefowl, the progenitor of domestic fowls.

Time-series transcriptome analysis revealed differently expressed genes in broiler chicken infected with mixed *Eimeria* species, as demonstrated by Kim et al. The coccidiosis disease brought on by the *Eimeria* species is quite harmful for the poultry production. RNA sequencing was employed to track the temporally-dependent host responses of chickens infected with *Eimeria* in order to investigate the genes and biological processes linked to parasite immunity. Four, seven, and 21 days post infection (dpi) were the times at which transcriptome analysis was carried out. Three categories of genes with differential expression were identified based on alterations in gene expression patterns. As a result, endoplasmic reticulum stress was document during the early stages of *Eimeria* infection. Furthermore, innate immune responses to the parasite were engaged at the time of the initial exposure and gradually returned to normal. Despite being considerably active at 4 dpi, the cytokine-cytokine receptor interaction pathway was downregulated, which had an anti-inflammatory effect. After *Eimeria* infection, the creation of gene co-expression networks also made it possible to identify key pattern recognition receptors and immunoregulation hub genes. These findings give a thorough insight of how chickens and *Eimeria* interact as hosts and pathogens. The gene clusters identified in this work can be used to enhance the coccidiosis resistance of chickens.

In conclusion, the evaluation of breeds by comprehensive phenotyping and the effective study of breed genetic diversity using appropriate polymorphic markers or genome-wide SNP genotyping are both possible today thanks to the combination of traditional techniques and approaches with cutting-edge

genetic and genomic methods. In this way, the genotype/genomic content is important and gains from the comparison or combination of conventional and genomic metrics. The papers in this Research Topic serve as excellent illustrations for how to create and contrast "portraits" of breeds at the level of complete genomic sequences and transcriptomes, as well as how to identify potential candidate genes for key features in particular breeds. The history of domestication and the development of breeds can be clarified using genomic data, and genomic regions with signs of artificial selection can be found. The formulation of breed conservation measures, their sustainable use, and marker-assisted breeding are all made possible with the help of comprehensive breed assessment using genetic, genomic and multi-omic approaches.

Author contributions

Writing—MR, JS, NZ, SW, and KW. Editing—MR, JS, NZ, SW, and KW.

Acknowledgments

The skilled technical assistance of Olga M. Romanova in preparing Figure 1 is kindly appreciated.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fgene.2021.775499

Selection of Cashmere Fineness Functional Genes by Translatomics

*Yu Zhang[1†], Dongyun Zhang[2†], Yanan Xu[1], Yuting Qin[1], Ming Gu[1], Weidong Cai[1], Zhixian Bai[1], Xinjiang Zhang[1], Rui Chen[1], Yingang Sun[1], Yanzhi Wu[1] and Zeying Wang[1]**

[1]College of Animal Science andVeterinary Medicine, Shenyang Agricultural University, Shenyang, China, [2]International Business School and International Economics and Trade, Shenyang Normal University, Shenyang, China

Cashmere fineness is an important index to evaluate cashmere quality. Liaoning Cashmere Goat (LCG) has a large cashmere production and long cashmere fiber, but its fineness is not ideal. Therefore, it is important to find genes involved in cashmere fineness that can be used in future endeavors aiming to improve this phenotype. With the continuous advancement of research, the regulation of cashmere fineness has made new developments through high-throughput sequencing and genome-wide association analysis. It has been found that translatomics can identify genes associated with phenotypic traits. Through translatomic analysis, the skin tissue of LCG sample groups differing in cashmere fineness was sequenced by Ribo-seq. With these data, we identified 529 differentially expressed genes between the sample groups among the 27197 expressed genes. From these, 343 genes were upregulated in the fine LCG group in relation to the coarse LCG group, and 186 were downregulated in the same relationship. Through GO enrichment analysis and KEGG enrichment analysis of differential genes, the biological functions and pathways of differential genes can be found. In the GO enrichment analysis, 491 genes were significantly enriched, and the functional region was mainly in the extracellular region. In the KEGG enrichment analysis, the enrichment of the human papillomavirus infection pathway was seen the most. We found that the *COL6A5* gene may affect cashmere fineness.

Keywords: cashmere fineness, translatomics, Ribo-seq, COL6A5, liaoning cashmere goat

OPEN ACCESS

Edited by:
Johann Sölkner,
University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Juliana Afonso,
Embrapa Pecuária Sudeste, Brazil
Ran Di,
Institute of Animal Sciences (CAAS), China

****Correspondence:***
Zeying Wang
wangzeying2012@syau.edu.cn

†These authors have contributed equally to this work

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

Received: *14 September 2021*
Accepted: *16 November 2021*
Published: *04 January 2022*

Citation:
Zhang Y, Zhang D, Xu Y, Qin Y, Gu M, Cai W, Bai Z, Zhang X, Chen R, Sun Y, Wu Y and Wang Z (2022) Selection of Cashmere Fineness Functional Genes by Translatomics. Front. Genet. 12:775499. doi: 10.3389/fgene.2021.775499

INTRODUCTION

Cashmere goat is a unique animal husbandry resource in China (Cai et al., 2020; Jin et al., 2020), which has made outstanding contributions to the needs of the livestock textile industry and human society (Yang et al., 2020). The characteristics of cashmere show significant differences due to different goat species and regions (Jin et al., 2017). Cashmere goat has a double coat, with the primary hair follicles producing coarse hair and the secondary hair follicles producing fine hair (Dai et al., 2019). The cashmere thickness produced by the secondary hair follicles varies among individuals. Cashmere fiber is famous for being slender and soft. To improve the economic value of cashmere, the fineness of its fibers needs to be reduced, making cultivating high-yield and high-quality cashmere goat varieties the core element to improve the economic value of cashmere fiber. Cashmere fineness is a quantitative characteristic, which is determined by micro effective genes. Molecular breeding research in domestic and foreign studies have proved that the single nucleotide polymorphisms (SNPs) of *KAP* (Jin et al., 2011), *RPL* (Mahata et al., 2012), *FGF* (Liu et al., 2009), *KRT* (Hui, 2020), *PROP* (Zeng et al., 2011), *MAF70* (Arranz

TABLE 1 | Quality control table of fine type and coarse type data of LCG.

Sample	Raw reads	Low quality	Clean reads	Total mapped	Uniquely mapped	Q20 (%)	rRNA reads	tRNA reads
FT LCG 1	50666392	99284 (0.20%)	50431480	7931912 (88.92%)	4235822 (47.49%)	99.10	38526496 (76.39%)	2434096 (4.83%)
FT LCG 2	50286756	87865 (0.17%)	50051017	8231481 (88.65%)	3621884 (39.01%)	99.12	37090706 (74.11%)	3282429 (6.56%)
FT LCG 3	52749309	129738 (0.25%)	51299342	6667995 (89%)	3922342 (52.35%)	98.97	40540273 (79.03%)	818312 (1.60%)
CT LCG 4	49894812	146884 (0.29%)	47332796	3722651 (71.99%)	1381041 (26.71%)	98.81	37754715 (79.76%)	210679 (0.45%)
CT LCG 5	51599006	90692 (0.18%)	51323088	7626758 (87.43%)	4235618 (48.55%)	99.13	37502937 (73.07%)	3121300 (6.08%)
CT LCG 6	59901380	100832 (0.17%)	59536550	7563025 (85.51%)	3718500 (42.04%)	99.15	45932536 (77.15%)	2854453 (4.79%)

FT LCG 1 is the fine Liaoning Cashmere Goat No. 1 sample, FT LCG 2 is the fine Liaoning Cashmere Goat No. 2 sample, FT LCG 3 is the fine Liaoning Cashmere Goat No. 3 sample, CT LCG 4 is the coarse Liaoning Cashmere Goat No. 4 sample, CT LCG 5 is the coarse Liaoning Cashmere Goat No. 5 sample, and CT LCG 6 is the coarse Liaoning Cashmere Goat No. 6 sample; Q20 is the percentage of bases with phred values greater than 20 in the total bases.

et al., 2001), and *KIFI* (Li, 2009) genes are correlated with cashmere fineness. Many studies have identified some genes related to goat wool fiber growth characteristics, such as *DSG1, IGF-IR, KRTAPs, ILK,* and *KRTAP* genes (Rufaut et al., 1999; Jia et al., 2006; Yu et al., 2011; Yang et al., 2012; Liu et al., 2015). Full transcriptomics also proved that *NFKBIA, TCHH, COL1A1, CXCL8,* and *LTBP2* genes are closely related to cashmere fineness, and researchers have been looking for ways to improve the quantity and quality of cashmere (Bai et al., 2016; Zhang et al., 2018), as well as analyzing the key genes, signal pathways, and expression regulation level under different cashmere state diameters (Fu et al., 2020). Despite the efforts, all the genes regulating cashmere fineness from the translation level remain unknown. Due to the gap between the transcriptome and proteome in omics, only 20–40% of proteins in mammalian cells are determined at the transcription level (Tian et al., 2004; Cox et al., 2005). Current translatomics can make up the gap between the two. In the study of cashmere fineness, translatomics can be used to identify genes that are not translated into proteins at the transcriptional level. Translatomics provides new insights into gene expression (Courtes et al., 2013).

Translatomics is the sequencing and analysis of translated RNA molecules, which can accurately quantify the genes being translated, and compare the translation amount of different samples of genes under different physiological and pathological conditions or different treatments. At the same time, translation control is the key determinant of protein abundance, which in turn determines cell state (Sendoel et al., 2017). The gene expression process reveals that there are specific phenotypic characteristics among species, but their evolutionary process is uncertain outside the transcriptome. There are studies on coevolution at the level of mammalian transcriptome and translatome. Ribosome analysis and RNA sequencing are used to analyze the three organs of five mammals (human, macaque, mouse, opossum, and platypus) and birds (chicken), coevolutionary analysis (brain, liver, and testis) shows that translation regulation widely exists in different organs, especially in spermatogenic cell types of testis, and some genes evolve faster at the translatome level (Wang et al., 2020). Ribosomal analysis can also evaluate translation efficiency on a genome-wide scale, which has been previously proved in yeast (McManus et al., 2014), nematodes (Stadler and Fire, 2013), primate (Wang et al., 2018) cells, and hybrid mouse cells, it was also found that translation efficiency was a momentous predictor of protein level in mouse fibroblasts (Hou et al., 2015). These studies provide preliminary insights into the coevolution model of the transcriptome and translatome. Therefore, protein abundance seems to be mainly regulated by ribosomes, highlighting the importance of translation control (Gebauer and Hentze, 2004; Sonenberg and Hinnebusch, 2009). Using microRNA (miR-430) in zebrafish to investigate its translational repression and mRNA decay, we found that translation repression occurs before mRNA decay, which is induced by reducing the translation initiation rate, and that mRNA decay is induced by deadenylation. Besides, microRNA has been proposed to affect protein translation by reducing the rate of translation initiation (Bazzini et al., 2012). On the other hand, because the translatome studies RNA molecules being translated, which includes the RNA molecules that are traditionally considered noncoding, it can provide direct translation evidence for the study of these new translatable molecules (Gerashchenko et al., 2012). In addition, the lack of a strict direct correlation between gene and protein levels limits translation studies by combining the transcriptome and proteome. Considering the high cost associated with protein synthesis, the dominant role of translation regulation is meaningful. Therefore, it promotes the progress of translatomics technology (King et al., 1996; Kirkpatrick et al., 2005).

According to Tian Wenliang, the standard of cashmere fineness in China is: the diameter of coarse hairs: 16.0–18.5 μm and the diameter of fine hairs: 15.5–16.0 μm. In 2010, the cashmere of LCG was the thickest among all cashmere goats, and its diameter was 20.32 μm (Tian, 2015). It can be found that the main disadvantage of cashmere in LCG is that the cashmere is thicker. In this study, we found the key genes regulating cashmere fineness by using translatomics, and understood the regulatory relationship of related genes. We used Ribo-seq to test the fine skin samples and coarse skin samples of LCG, and found the key genes, differential genes, and co-expression genes related to cashmere fineness through GO function enrichment analysis and KEGG pathway enrichment analysis. These findings pave the way for the study of the

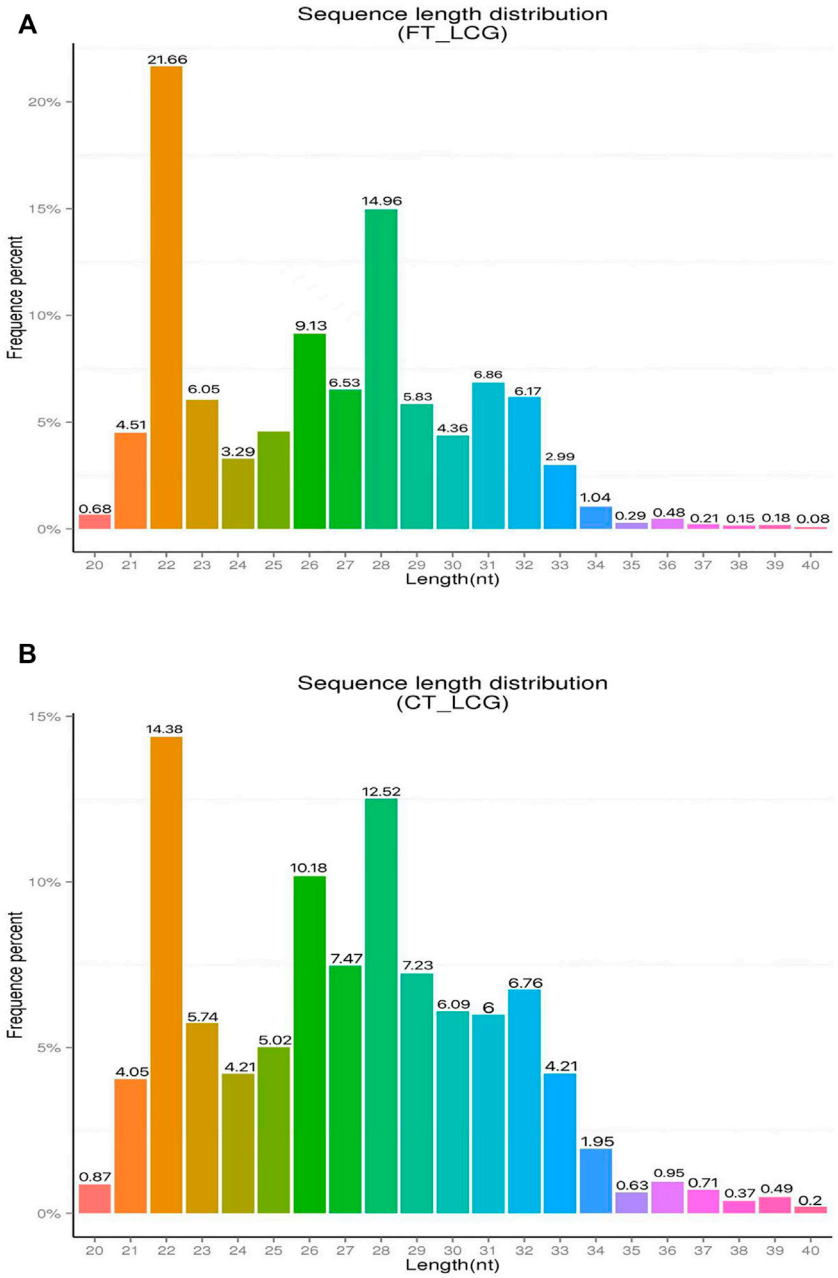


FIGURE 1 | Statistical chart of clean reads length enrichment of Ribo-seq in LCG. The abscissa is the enrichment length and the ordinate is the percentage of enrichment frequency. **(A)** Sample of fine LCG. **(B)** Sample of coarse LCG.

regulation mechanism of cashmere fineness and the protection and cultivation of cashmere varieties.

MATERIALS AND METHODS

Ethics Statement

The whole process of experiments was based on guidelines from the Animal Experimental Committee of Shenyang Agricultural University (Shenyang, China, 201906099).

Sample Preparation

The sample collection was a crucial step for Ribo-seq since it was the starting point of library construction. Two groups of three Liaoning Cashmere Goats (LCGs) differing in cashmere fineness were used here, including three fine LCGs (14.32, 14.69, and 14.77 μm) and three coarse LCGs (17.23, 17.63, and 17.91 μm). These 2-year-old adult female Liaoning Cashmere Goats were reared under the same (sheep house, environment, management, nutrition) conditions. Skin samples were collected from the front edge of the left scapula, with a

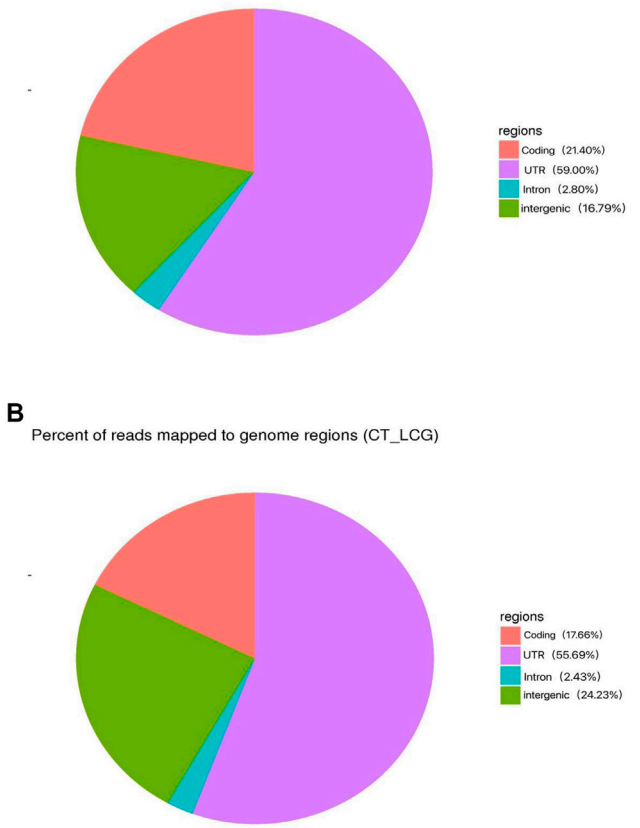


FIGURE 2 | Distribution of reads in different regions of the reference genome. Red indicates the coding area. Purple indicates the UTR area. Green indicates the intergenic area. Blue indicates the intron area. **(A)** Sample of fine LCG. **(B)** Sample of coarse LCG.

sample size of 2 cm^2. Anesthetics were used to relieve the pain of goats.

Library Construction for Ribo-Seq

To digest RNA other than RPFs, cell or tissue lysate was treated with unspecific endoribonuclease RNase I. Isolation of monosomes was performed by size-exclusion chromatography with MicroSpin S-400HR columns. The RNA samples were then treated with an rRNA depletion kit to deplete the samples of as much rRNA contamination as possible before PAGE purification of the relatively short (20–38 nt) RPFs. Following PAGE purification, both ends of the RPF were phosphorylated and ligated with 5′ and 3′ adapters, respectively. Then the fragments were reversely transcribed to the cDNAs and amplified by PCR (Aeschimann et al., 2015).

After library construction, the concentration of the library was measured by The Qubit® 2.0 Fluorometer and adjusted to 1 ng/uL. An Agilent 2100 Bioanalyzer was deployed to examine the insert size of the acquired library. At last, the accurate concentration of the cDNA library was again examined using qPCR. Once the insert size and concentration of the library became identical, the samples could then be subjected for sequencing.

Sequencing

After library preparation and pooling of different samples, the samples were subjected to Illumina sequencing. Commonly, the Ribo-seq uses PE150 (paired-end 150 nt) sequencing for 15 G raw data.

Quality Control for Raw Data

Firstly, the initial data (in the format of FASTQ) and the adapter were processed to delete the 3′ ends sequence and obtain the clean data of Q20. The following analysis was based on the clean data.

Mapping

Ribo-seq used TopHat2 for genome mapping. TopHat2 is an enhanced version of TopHat, using short read aligner Bowtie to align the RNA-seq reads to mammalian-sized genomes and analyzing the mapping result to identify splice junctions. TopHat2 allows variable-length indels with respect to the reference genome, which give it the ability to accurately align the transcriptomes in the presence of insertions, deletions, and gene fusions (Kim et al., 2013).

Quantification of Gene Expression Level

Quantification of mapped results to gene level was carried out using HTSeq. HTSeq is a Python package that calculates the number of mapped reads to each gene (Anders et al., 2015). RPKM values were generated to represent the gene expression level of each specific gene. RPKM is the abbreviation of "Reads Per kilobase of transcript, per Million mapped reads," which normalizes both sequencing depth and gene length (Gross et al., 2013).

Differential Expression Analysis

For samples with biological replicates, the DESeq2 R package (1.14.1) was used for differential expression analysis. DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on negative binomial distribution (Wang et al., 2010). The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with $p < 0.05$ found by DESeq2 were assigned as differentally expressed (Tang et al., 2007).

GO and KEGG Enrichment Analysis

Gene Ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species. GO covers three domains: cellular component, molecular function, and biological process. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances (Kanehisa et al., 2008). In the KEGG pathway database, the wiring diagram database, is the core of the KEGG resource. It is a collection of pathway maps integrating many entities including genes, proteins, RNAs, chemical

TABLE 2 | Quantitative analysis table of the first 30 gene expression levels of each sample.

Gene name	FT LCG 1	FT LCG 2	FT LCG 3	CT LCG 4	CT LCG 5	CT LCG 6
LOC108638395	445319	336977	1344217	574326	742089	514676
MIR148A	243740	359838	356349	26236	222275	342894
COL1A1	22968	46391	2220	4944	23410	30471
COL1A2	14548	28783	1640	2639	14371	17909
EEF1A1	451	716	208	105	395	337
MIR10B	243316	172218	204214	10272	113416	170024
FTH1	383	727	49	99	442	291
COL3A1	7095	19128	1036	1499	7544	10158
MIRLET7I	97833	118750	141902	44311	158213	56831
LOC108635080	157034	98967	135504	102991	47625	62559
LOC102184404	5989	7402	1133	392	6438	12665
KRT5	6949	13975	3529	1867	4549	7209
MIR99A	124457	66498	46951	10679	69591	180316
TCHH	4054	6942	1962	504	1893	4115
KRT14	4814	8188	1586	843	2640	4205
FASN	32449	4204	423	1209	9573	6267
LOC102185436	1517	1742	510	383	906	1625
LOC102177231	6496	10901	2395	1128	4154	5305
MIR26A	92770	73542	122982	7897	34599	79360
LOC102184223	2180	2251	646	404	1482	2195
KRT25	3583	6185	1085	662	2790	3315
RPL4	167	175	30	17	142	136
RPS8	141	219	24	22	150	100
MIR126	75155	59465	76507	8566	59254	55597
TPT1	73	72	42	52	74	71
SPARC	5707	10185	875	1213	4087	5190
KRT10	1490	3452	173	46	359	628
DSP	2491	4369	739	282	1181	2158
RPLP0	299	323	113	67	271	266
COL6A5	50	682	805	39	28	37

compounds, glycans, and chemical reactions, as well as disease genes and drug targets, which are stored as individual entries in the other databases of KEGG.

P- Site Analysis

Identifying the A- and P-site locations on ribosome-protected mRNA fragments from Ribo-Seq experiments was a fundamental step in the quantitative analysis of transcriptome-wide translation properties at the codon level. The P-site (for peptidyl) is the second binding site for tRNA in the ribosome. During protein translation, the P-site holds the tRNA which is linked to the growing polypeptide chain. When a stop codon is reached, the peptidyl-tRNA bond of the tRNA located in the P-site is cleaved releasing the newly synthesized protein. Since translation occurs at the A- and P-sites, the identification of these sites was critical to address translation-related questions (Ahmed et al., 2019). Novogene used the Ribocode package to analyze the P-site using Ribo-seq data (Xiao et al., 2018).

uORF Analysis

An upstream Open Reading Frame (uORF) is an Open Reading Frame (ORF) within the 5′ untranslated region (5′UTR) of an mRNA. uORFs can regulate eukaryotic gene expression. Associated with mRNA-seq, all identified ORFs by Ribo-seq were classified. Ribotape was then used to analyze the motif of translated/untranslated uORFs, which can be used to study the base composition bias of uORF sequences.

RESULT

Quality Control of Sequencing Data of Six LCGs

Through high-throughput sequencing, the raw reads sequences of fine LCG samples accounted to 50666392, 50286756, and 52749309. Low quality data accounted for about 0.21%. The rest were clean reads. The percentage of sequencing sequences that could be located on the genome was about 88.86%. The average percentage of sequencing sequences with unique alignment positions on the reference sequence was about 46.28%. About 99.06% of the base group had a mass value greater than 20. The average proportion of filtered rRNA to total clean reads was about 76.51%. The average proportion of filtered tRNA in the total number of clean reads was about 4.33%.

The raw reads sequences of coarse LCG samples accounted to 49894812, 51599006, and 59901380. Low quality data accounted for about 0.21%. The rest were clean reads. The percentage of sequencing sequences that could be located on the genome was about 81.64%. The average percentage of sequencing sequences with unique alignment positions on the reference sequence was about 39.10%. About 99.03% of the base group had a mass value greater than 20. The average proportion of filtered rRNA to total clean reads was about 76.66%. The average proportion of filtered tRNA in the total number of clean reads was about 3.77% (**Table 1**).

The effect of experimental enrichment can be evaluated by counting the length of ribosome protected RNA fragments

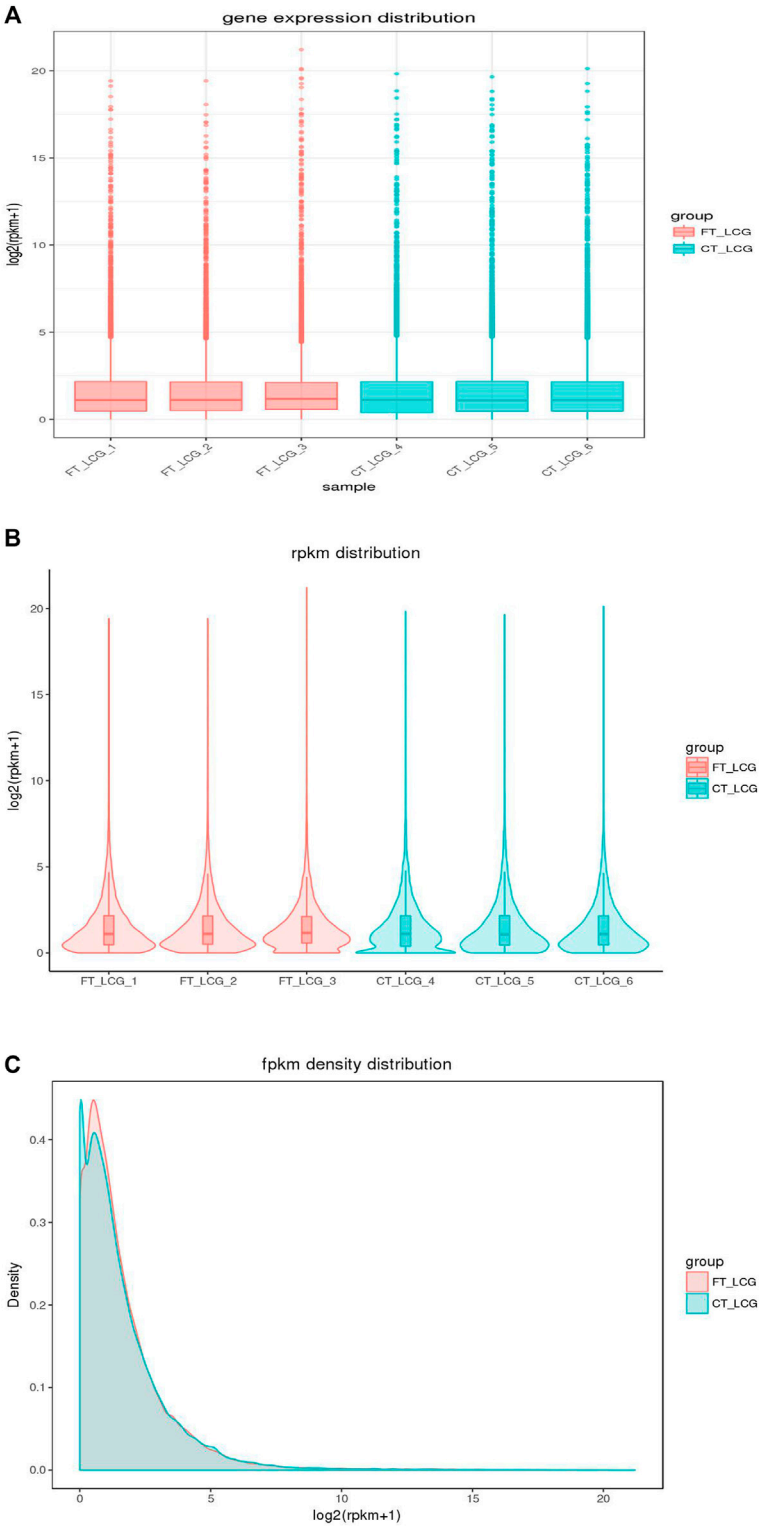


FIGURE 3 | Distribution of gene expression in fine and coarse type LCGs. The abscissa in the graph is the sample name, and the ordinate is log2 (RPKM+1). The fine LCG sample is shown in pink, and the coarse LCG sample is shown in blue. The box graph of each region has five statistics (the maximum, upper quartile, median, lower quartile, and minimum from top to bottom).

TABLE 3 | Partial up-down regulation difference gene table between fine type and coarse type LCGs.

Gene name	FT LCG	CT LCG	P-value	Up/down
COL6A5	513.0210584	35.10807307	1.09E-06	Up
ITIH4	13.30490937	0.652476505	8.14E-06	Up
LOC102173569	19.27928687	2.77555756	1.16E-05	Up
LOC102173761	21.50257982	2.956095588	1.34E-05	Up
CX3CR1	11.42687495	0.597894581	1.38E-05	Up
LOC102183314	14.40120352	0.597421578	2.00E-05	Up
LOC108638245	13.94060036	0.324997368	2.54E-05	Up
SCIN	54.00576352	12.50050163	3.37E-05	Up
MIR223	526.9564919	77.06293267	6.86E-05	Up
KRT2	80.95366623	6.940780142	7.45E-05	Up
LOC102181854	11.9493693	0.668851595	0.000179452	Up
SPI1	7.221664344	2.77555756	0.000211171	Up
TNC	180.2254246	12.67495436	0.000214188	Up
LTF	49.98989748	6.732485479	0.000223432	Up
SHANK3	5.344749546	2.77555756	0.000234583	Up
LOC102181202	95.51666376	24.59851436	0.000284781	Up
AGP	8.736937811	2.77555756	0.000318954	Up
LOC102185525	92.67098586	15.49360806	0.00039787	Up
LOC102191415	34.15538801	3.040163575	0.000861835	Up
BCL2A1	8.331100162	0.977471879	0.001012572	Up
ZNF550	3.50939654215679	17.2539274369117	0.000425155705185925	Down
MIOX	0.178346971505201	6.58571066698514	0.000550580885740381	Down
LOC108636746	20.2230313525199	71.2478009465496	0.000575441389747231	Down
LOC102175702	2.77555756156289E-17	4.81857339537376	0.00092949091732039	Down
LOC102168522	0.942200471766223	7.50055476502101	0.00126514633141028	Down
MIR9	9.82710757202138	80.2714587053271	0.00139800098028482	Down
MIR671	5.33927118533789	56.4684485860653	0.00147473232541579	Down
SDC3	17.3545579173651	81.4319352084731	0.00162620276767528	Down
LHX1	1.34518325391645	8.62420593115666	0.00167178598923491	Down
ABHD15	5.12668482061614	49.4793752852943	0.00203138254923832	Down
LOC102173583	1.26347479869531	8.90431324840639	0.00210487578853974	Down
LOC108634715	0.573483698205509	7.36640964136039	0.00238509701852834	Down
LOC108634352	0.563608783979002	5.76016723909547	0.00260838978865612	Down
LOC102190399	0.178346971505201	4.31371001374939	0.00265357125604965	Down
LOC102188887	4.46437256485464	16.6419755873072	0.00330633387847916	Down
CHD3	40.3297349602123	203.461408220537	0.00342774246368716	Down
LOC108638293	9.2957498124219	30.6384923277523	0.00358064723751997	Down
RARRES1	43.7517203353152	121.773870596007	0.00366888299547616	Down
CUX2	0.383840891809286	6.11577320580318	0.00368401375810571	Down
SLC25A11	12.9565771393252	35.6648973518473	0.00369472188184243	Down

(RPFs). The length statistics of Ribo-seq clean reads of fine and coarse samples of LCG showed that when the enrichment length was 22 nt, the enrichment frequency was the highest, 21.66 and 14.38%, respectively **(Figures 1A,B)**. The genomic region sequencing distribution is shown in the figure: the coding region of fine Liaoning cashmere goat accounted for 21.40%, the UTR region accounted for 59.00%, the intron region accounted for 2.80%, and the intergenic region accounted for 16.79% (**Figure 2A**). The coding region of coarse Liaoning cashmere goat accounted for 17.66%, the UTR region accounted for 55.69%, the intron region accounted for 2.43%, and the intergenic region accounted for 24.23% (**Figure 2B**).

Quantitative Analysis and Distribution of Gene Expression in Six LCGs

The translation level of a gene protein is directly reflected by the abundance of ribosome binding on its corresponding transcript. The higher the binding abundance is, the higher the level of gene translation is. We quantitatively analyzed the gene expression level of the top 30 genes in each sample, as shown in **Table 2**.

In addition to the true translation level, the reads count was positively correlated with the sequencing depth. Generally, the gene expression value is not expressed by reading count, but by RPKM, which corrects the sequencing depth and gene length successively. After calculating all gene expression values (RPKM) of each sample, we showed the distribution of gene expression levels of different samples by box graph. From the box graph, we can see that there were differences in the expression levels of all the genes detected among the six samples, and the box graph of sample 3 was smaller than that of other samples, so we can see that the differences between the genes detected in sample 3 and other samples were more obvious. At the same time, obvious different genes were found between the two groups. The coincidence of the two groups was the co-expressed genes, and the noncoincidence was the differentially expressed genes (**Figures 3A–C**).

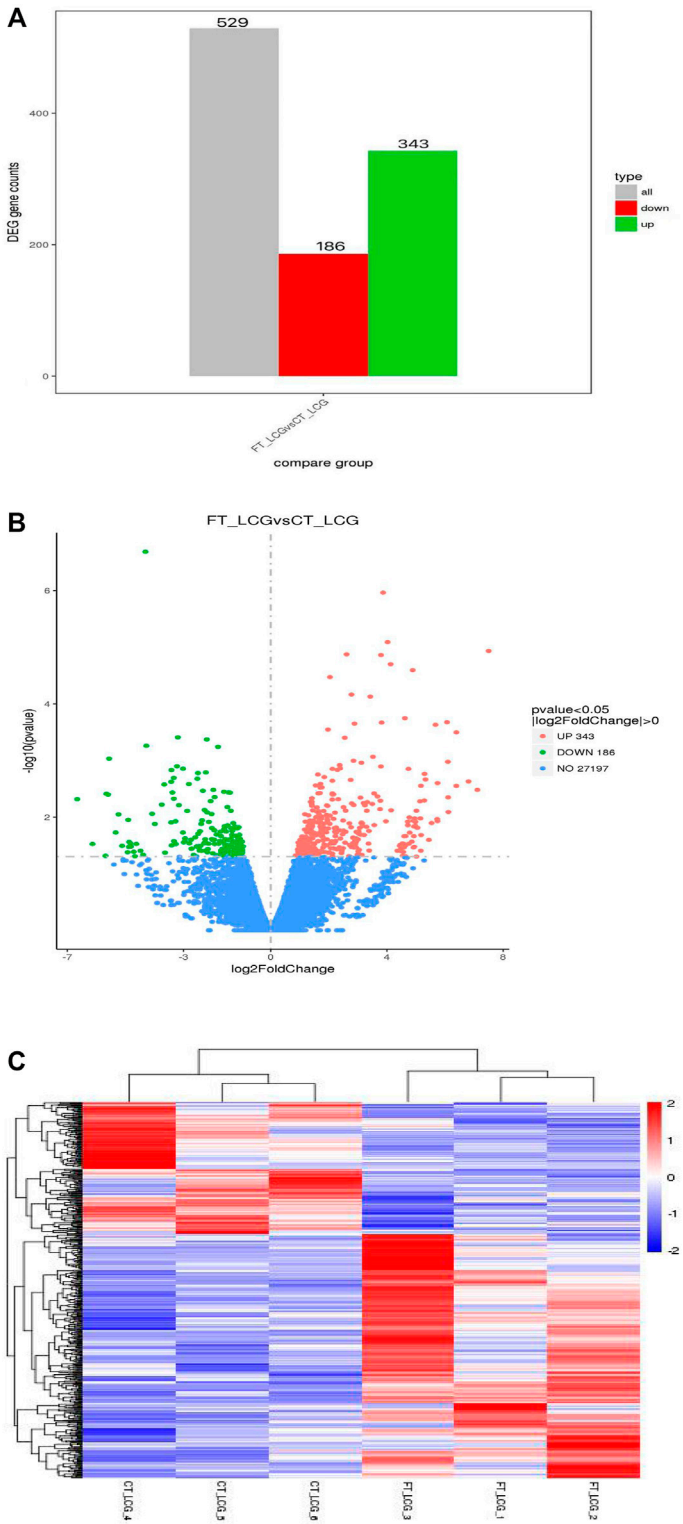


FIGURE 4 | Statistical chart of the number of different genes between fine type and coarse type LCG. **(A)** The abscissa of the histogram is divided into two groups: fine type and coarse type, the ordinate is the gene counts. **(B)** In the volcanic map, the abscissa represents the log2foldchange of gene expression in the treatment and control groups, and the ordinate represents the significant level of gene expression difference between the treatment and control groups, in which the upregulated genes are indicated by red dots and the downregulated genes are indicated by green dots. **(C)** Gene expression cluster map of fine and coarse type LCG. In the graph, red indicates high gene expression and blue indicates low gene expression. The abscissa is the sample name, and the ordinate is the normalized value of RPKM.

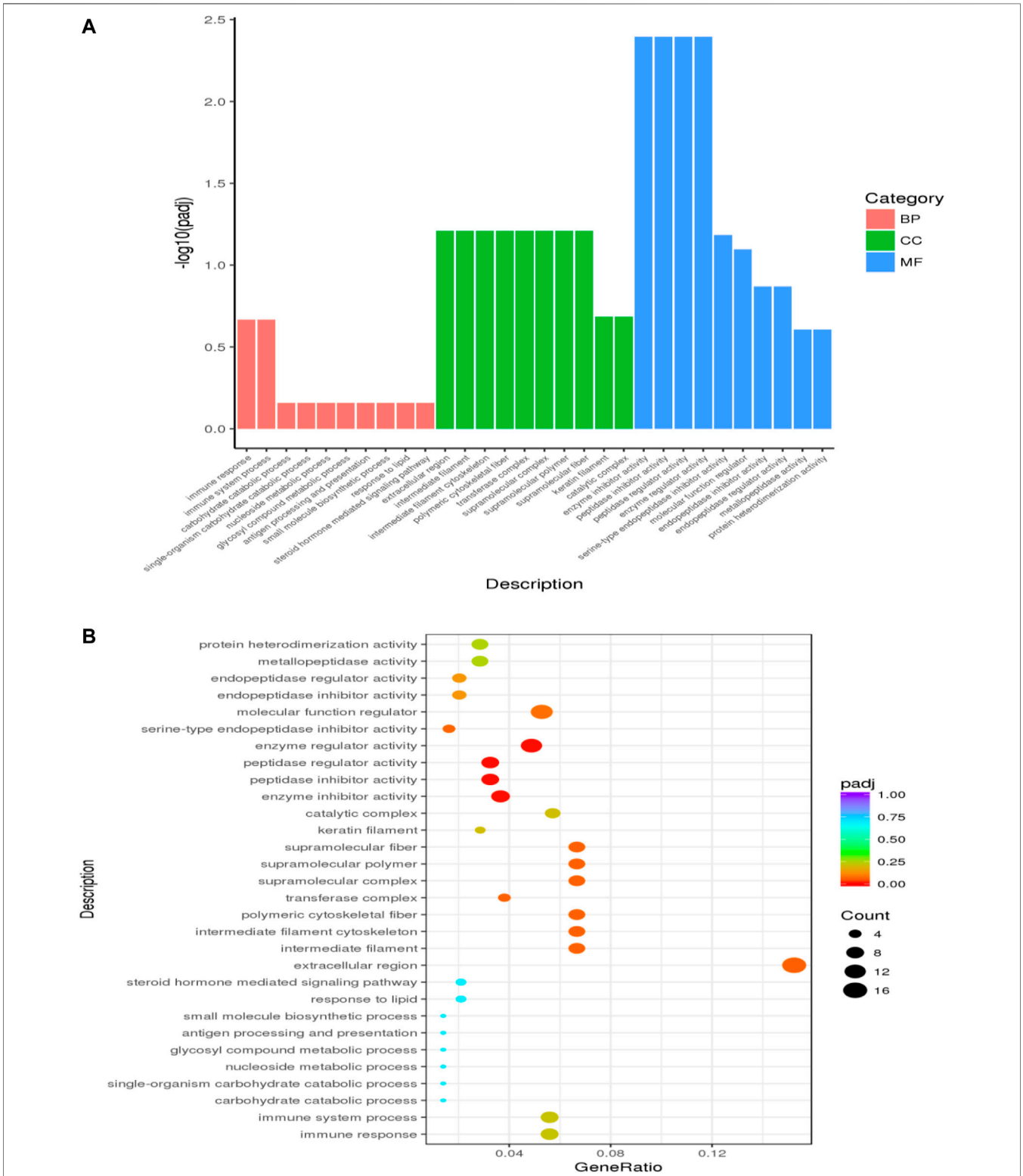
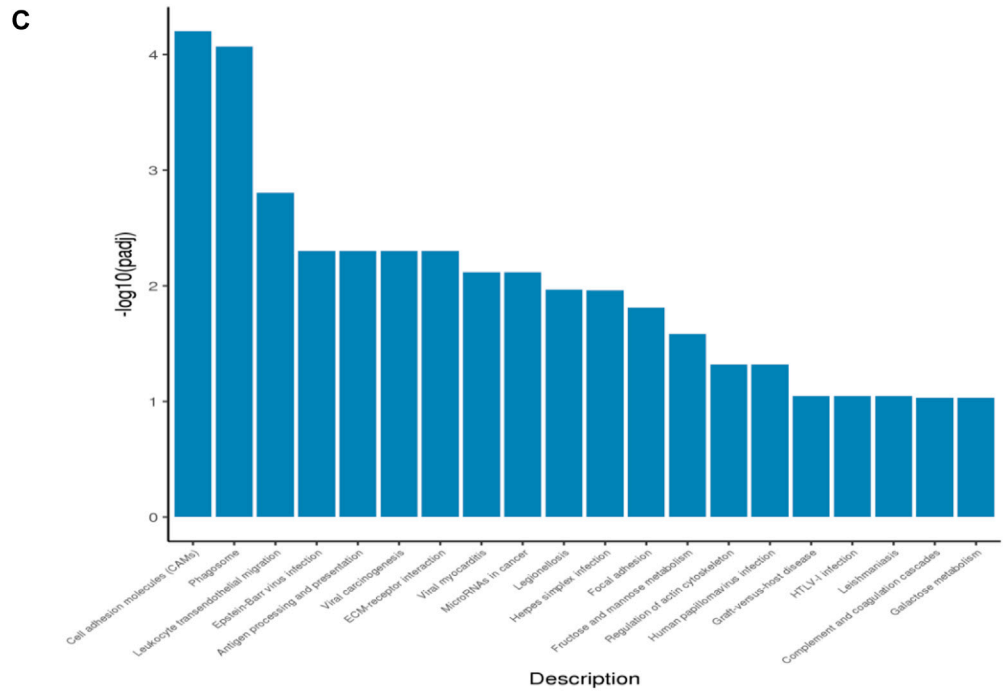


FIGURE 5 | Enrichment analysis in LCG. **(A)** GO enrichment analysis histogram, the abscissa is the GO term, and the ordinate is the significance level of GO term enrichment. Height was positively correlated with the significance. Different colors represent BP, CC, and MF. **(B)** Scatter plot of enrichment analysis, the horizontal coordinate in the Panel is the ratio of the number of different genes annotated to the GO term and the total number of different genes. The ordinate is the GO term, the size of points represents the number of genes annotated to the GO term, and the level of enrichment varies from purple to red. **(C)** KEGG enrichment analysis histogram, the horizontal coordinate in the panel is the KEGG pathway, and the longitudinal coordinate is the significance level of channel enrichment, and the height of the histogram indicates the degree of enrichment. **(D)** Scatter plot of KEGG enrichment analysis, the abscissa is the ratio of the number of differential genes annotated to the KEGG pathway to the total number of differential genes. The ordinate is the KEGG pathway. The size of the dot represents the number of genes annotated to the KEGG pathway. And the level of enrichment varies from purple to red.



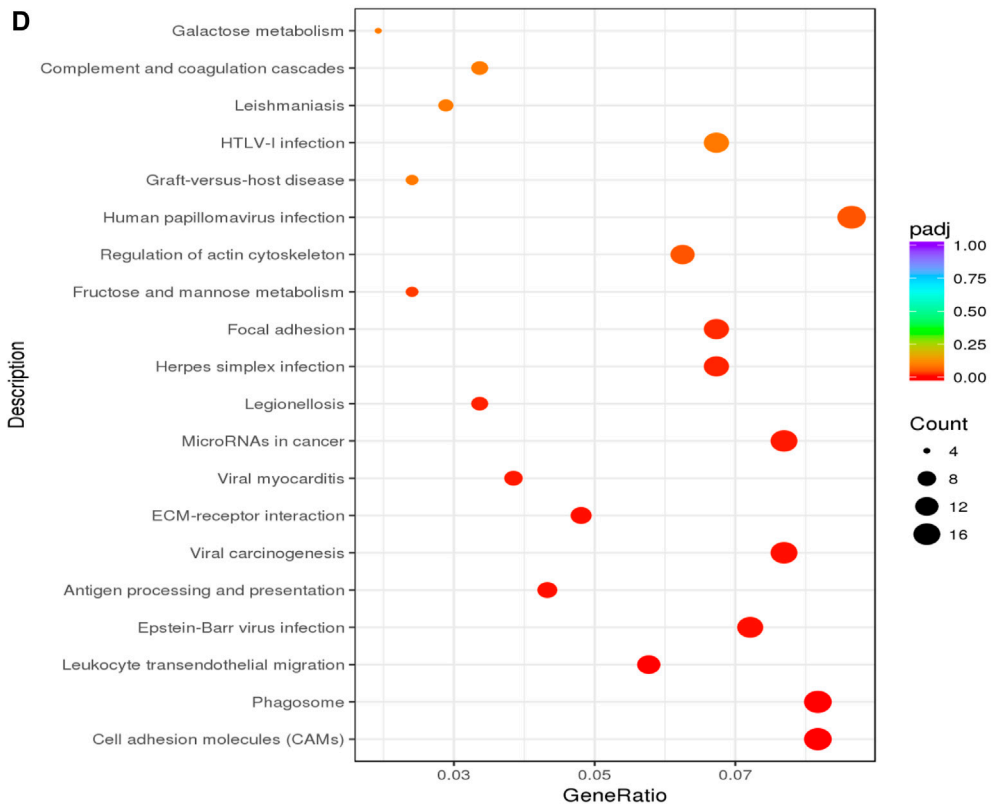


FIGURE 5 | (continued)

The Statistics of Differential Genes Were Carried out, and the Volcanic Map and Cluster Map Were Drawn

After the quantitative analysis of gene expression, it is necessary to conduct statistical analysis on the expression data, and screen the genes with significantly different expression levels in different states. Some differentially expressed genes are shown in **Table 3**.

The fine type and coarse type LCGs were compared by histogram and volcanic map, we can see that there were 27197 coexpressed genes and 529 differential genes, of which 343 were upregulated and 186 were downregulated. The clustering graph compared the gene expression of three samples of coarse LCG and three samples of fine LCG. Therefore, it can be seen that the overall gene expression trend of the three samples of coarse LCG was significantly different from that of the three samples of fine LCG. The genes with high expression in fine LCG had lower expression in coarse LCG, and the genes with low expression in fine LCG had higher expression in coarse LCG (**Figures 4A–C**).

GO Enrichment Analysis and KEGG Enrichment Analysis of Differential Genes

We Used ClusterProfiler Software (Yu et al., 2012) for analysis, through GO enrichment analysis, we selected the first 30 for analysis, and found 491 significantly enriched genes, including biological process (BP, 257), cell composition (CC, 64), and molecular function (MF, 170). The function was mainly in the extracellular region (**Figures 5A,B**). From the results of KEGG enrichment, the most significant 20 KEGG pathways were selected, and the most enriched pathway was the human papillomavirus infection pathway (**Figures 5C,D**).

P-Site Analysis of Six LCGs

During translation, ribosomes move in the unit of codon length (3 nt) relative to RNA. Therefore, based on P-site, RPF fragments from normal translation should be periodically distributed on RNA. The translation starts at 12 nt upstream of the initiation codon, and the distance from the termination codon 15 nt disappears gradually. This is direct evidence of whether an RNA is translated. The length of RNA fragments protected by the ribosome from the initial codon to the termination codon of three samples of LCG was 26, 27, and 28 nt, respectively (**Figure 6**).

ORF Analysis of Six Samples of LCGs

In the samples of fine LCG, 9925 genes were transformed into protein during the translation process, accounting for 95.5% of the total. Among them, 47 uORF (0.5%) and 9 dORF (0.1%) could be translated into protein. There were 69 overlapping genes (0.7%), 286 novel-PCGs (2.8%), and 59 novel-NonPCGs (0.6%) (**Figure 7A**). There were 8540 genes (96.3%) transformed into protein in coarse LCG samples during translation, among which 23 genes (0.3%) were uORF and 6 genes (0.1%) were dORF. There were 35 overlapping genes (0.4%), 229 novel-PCGs (2.6%), and 36 novel-NonPCGs (0.4%) (**Figure 7B**).

DISCUSSION

LCG is unique in China, which produces a large number of high-quality cashmere fibers. Moreover, China is one of the largest cashmere producing countries in the world, which has made great contributions to the fiber industry and textile industry, and plays an indispensable role in the global cashmere industry (Zheng et al., 2020). But at present, the pursuit of cashmere fineness is increasing, and the cashmere fineness of LCG is still showing a relatively coarse trend, and quality cashmere products are still insufficient. Reducing cashmere fineness is an important issue (Zheng et al., 2019). Ribosome profiling is a mature method to identify translation regions by high-throughput sequencing, which fills the gap between RNA sequencing and proteomics, and has become a powerful tool for gene expression (Calviello and Ohler, 2017). Ribo-seq can not only measure the translation efficiency according to the relative abundance of ribosomes on transcripts, but also reveal the dynamic and local regulation of different translation stages according to the location information of footprints on transcripts (Li et al., 2020).

However, little is known about the issue of cashmere fineness of LCG by Ribo-seq sequencing in translatomics. In this study, we selected six adult female Liaoning Cashmere Goats (LCGs) with different cashmere fineness (divided into two groups). The coarse Liaoning Cashmere Goat sample group was the reference group, and the fine Liaoning Cashmere Goat sample group was the experimental group. The classification of groups was based on the phenotypic determination of cashmere fineness (cashmere fineness analyzer, sirolan technology, Australia). A total of 529 differentially expressed genes were identified by Ribo-seq sequencing, of which 343 were upregulated and 186 were downregulated. And the enrichment length of mRNA fragments was 22 nt.

COL6A5 (formerly known as *COL29A1*) is a member of the collagen superfamily. The gene is located on the long arm of chromosome 3 (Strafella et al., 2019), with a length of 8742. It is a protein-coding gene and is considered to be an extracellular matrix protein, accounting for 30% of the total mammalian protein (Haq et al., 2019). *COL6A5* was found in the epithelial tissues of lung and gastrointestinal tract (Fitzgerald et al., 2008), but the highest expression was found below the dermal epidermal junction and around the reticular dermal vessels (Sabatelli et al., 2011). It was found that *COL6A5* fibroblasts existed in atopic dermatitis skin, but not in healthy tissues (He et al., 2020). The *COL6A5* gene is not only associated with skin inflammation, but also with cancer. It has been confirmed that the *COL6A5* gene is significantly associated with the overall survival rate of patients with esophageal squamous cell carcinoma (ESCC). The overall survival rate of ESCC patients with low expression of the *COL6A5* gene is poor, which may become a diagnostic marker of ESCC as a collagen gene (Li et al., 2019). Abundant type ⅵcollagen in lung tissue α5 (*COL6A5*), rs13062453, rs1497305, and rs77123808 of *COL6A5* polymorphism are associated with lung cancer risk in Chinese Han population, and the overall survival rate of patients with low expression of the *COL6A5* gene is poor (Duan et al., 2020). These studies can infer that the *COL6A5* gene may play a role in changing the hair follicle and cashmere fineness of LCG.

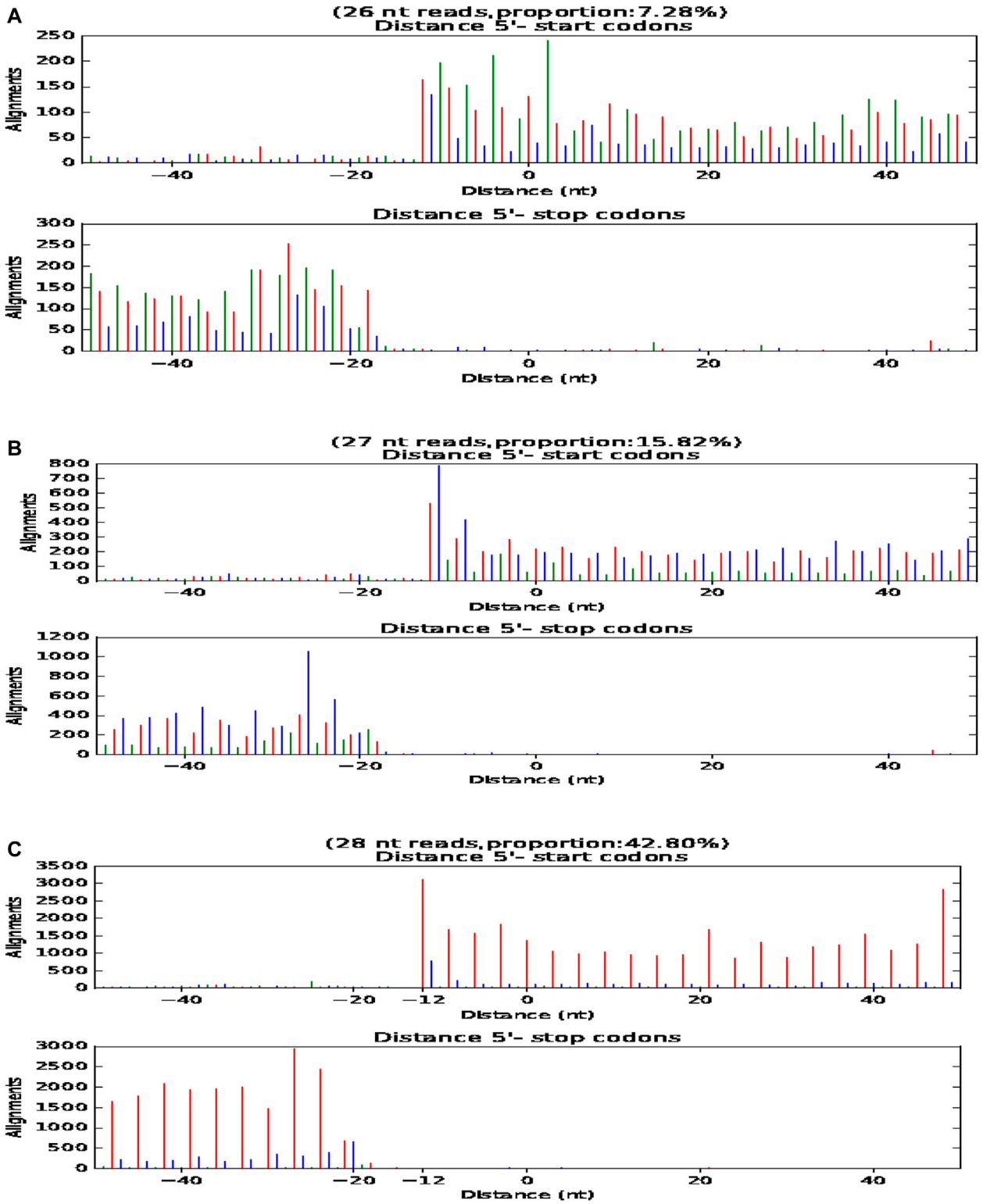


FIGURE 6 | P-site map of fine type and coarse type LCG. The abscissa is the distance from the start code or stop code, and the ordinate is the reads on the comparison. **(A–C)** P-site map of three fine type samples of LCG. **(D–F)** P-site map of three coarse samples of LCG.

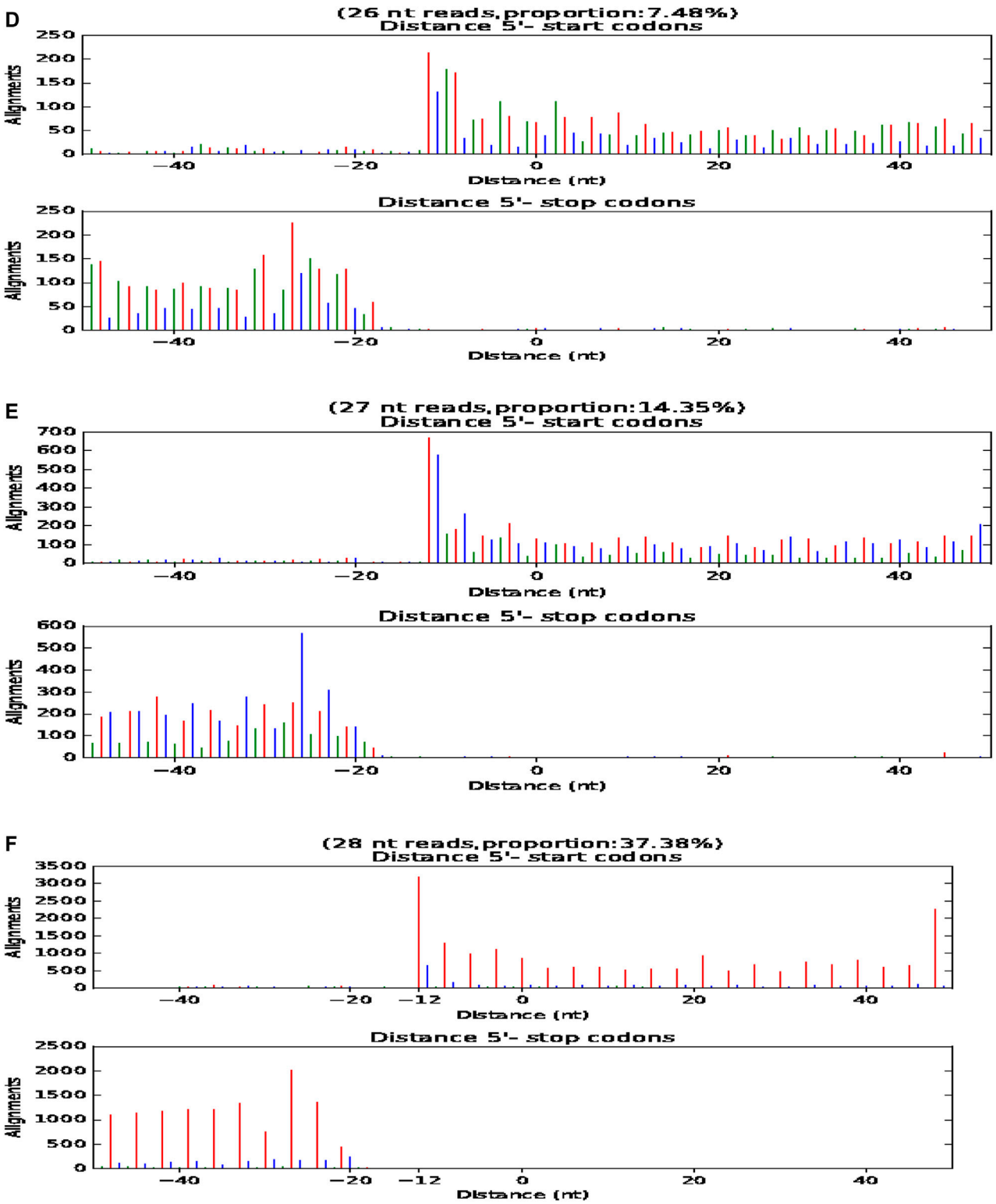


FIGURE 6 | (continued)

Because of that and because another gene from the same family, COL1A1, is known to have an impact on changing cashmere fineness (Wang et al., 2021), we hypothesize that COL6A5 is a candidate gene for future studies regarding cashmere fineness. Studies have shown that *COL6A5* is associated with familial chronic neurotrophic itching (Martinelli-Boneschi et al., 2017). It

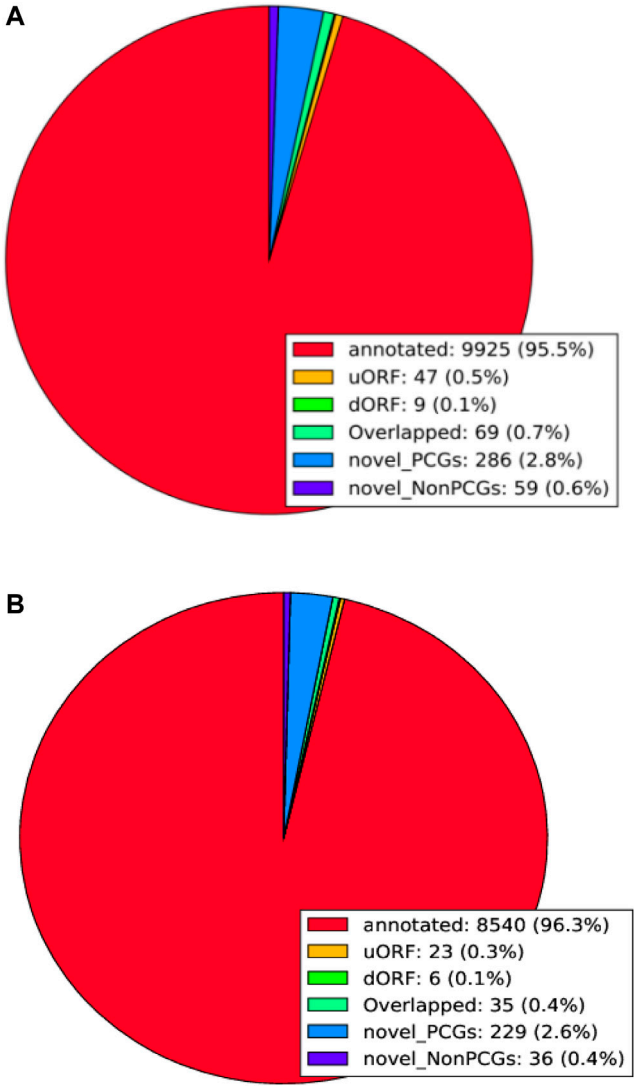


FIGURE 7 | ORF analysis of LCG. **(A)** ORF analysis of fine LCG. **(B)** ORF analysis of coarse LCG.

has been found that *COL6A5* expression in the papillary dermis and the surrounding dermis of the patients is reduced (Weisshaar and Dalgard, 2009; Ständer et al., 2010), and the incidence rate increases with age. This is the first time that the link between the *COL6A5* gene and chronic pruritus has been revealed. Some studies have found a link between *COL6A5* variants, reduced bone mass, dyspnea, and giant cell arteritis (Wang et al., 2016). In psoriasis (Ps) and psoriatic arthritis (PsA), bioinformatics analysis revealed that *COL6A5* and *COL8A1* participate in the altered proliferation and angiogenesis pathways in Ps/PsA, participate in inflammatory response together with *miR-146a,* and participate in the common and different biological pathways of Ps and PsA (Caputo et al., 2020).

The collagen gene may be closely related to the PI3K/Akt/mTOR pathway, p53 pathway, apoptosis, and cell cycle. *COL1A1* and *FGF10* genes are also enriched in the PI3K/Akt/mTOR pathway. *COL1A1* can regulate the growth of alpaca wool fiber, and *FGF10* can prolong the growth period of mouse hair follicles and promote hair growth (Maiese, 2015; Huang et al., 2017; Mendoza et al., 2019). Therefore, we take *COL6A5* gene as a candidate gene for cashmere fineness research in the future.

In living organisms, ribosomes synthesize proteins in the process of translation, and translation regulation itself goes beyond the three processes of transcription, mRNA degradation, and protein degradation. Like other omics, translatomics analyzes all components in the translation process, and also includes the study of mRNAs, ribosomes, tRNAs, regulatory RNAs, and newborn peptide chains (Zhao et al., 2019). Meanwhile, the correlation between the transcriptome and proteome is usually very poor in omics data, because the phenomenon of post transcriptional regulation, translation control, and other factors such as frameset translation is common. Translational sequencing can accurately quantify the genes being translated, and indirectly detect the protein expression from the genomic level, indicating the real situation of gene transcription and expression in biological samples. In addition, by comparing the gene translation differences between different samples, we can reveal the molecular mechanism of related physiological and pathological differences. Translatomics is a bridge between transcriptomics and proteomics. The multi omics analysis of translatomics can better study the mechanism of translation regulation. By analyzing the correlation between translatomics and transcriptomics, we can study the change of translation rate within genes, compare the relationship between gene transcription and translation, and study the difference of gene translation efficiency under different physiological and pathological conditions by calculating gene translation efficiency, to explain its molecular mechanism. The association analysis of translatomics and proteomics can study the relationship between transient translation and protein accumulation, assist proteome identification, provide evidence of gene translation, and indirectly determine the proteins expressed in biological samples.

It is critical to accurately investigate the underlying mechanisms of miRNA translation inhibition, and analyze the effect of post transcriptional regulation and RNA modification on gene translation. It is generally believed that the gene that can encode a protein with a length of more than 100 amino acids is a protein coding gene, while other gene sequences are noncoding sequences. However, in recent years, more studies have shown that some RNA regions (including lncRNA, 5′UTR, 3′UTR, circRNA, etc.) that are traditionally considered not to encode proteins can translate some peptides with a length of less than 100 amino acids. These peptides, of less than 100 amino acids in length, also play a variety of important roles in organisms, including ontogeny, muscle contraction, and DNA repair. Because translatomics is used to sequence and quantify the RNA molecules being translated, it can provide direct translation evidence for these noncoding sequences, and help to find and identify new unknown peptides.

CONCLUSION

In conclusion, this study analyzed the process of screening cashmere fineness functional genes by translatomics through Ribo-seq sequencing, found that the *COL6A5* gene may play

an important role in cashmere fineness regulation, and provided some theoretical basis for future research on this gene in the field of cashmere fineness regulation.

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: GEO Database, GSE186959.

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Experimental Committee of Shenyang Agricultural University.

AUTHOR CONTRIBUTIONS

YZ and DZ carried out data analysis and writing of the original draft. ZB, WC, and XZ were in charge of experiments. YX, MG, YQ, RC, YS, and YW were responsible for the collection of samples. ZW reviewed and edited the manuscript.

FUNDING

Our scientific research was financially aided by three projects:1. Postdoctoral Science Foundation of China: Genetic trajectory and expression localization of key genes of cashmere fineness by multi-omics, Project No. 2021M693859. 2. 2021 Liaoning Province "the open competition mechanism to select the best candidates" Science and Technology Research Project: Selection and breeding of special advantageous livestock and poultry breeds in Liaoning and key technology of whole industry chain production, Project No. 2021JH1/10400033. 3. National modern agricultural industrial technology system, project No: cars-39-27.

ACKNOWLEDGMENTS

We are sincerely grateful to the coworkers who work inside our laboratory at Shenyang Agricultural University and Liaoning Cashmere Breeding base for their help in the collection of samples from the feedlot of Liaoning cashmere goats and in the massive analysis of the samples. We would also like to express our gratitude to Novogene Biotechnology Co., Ltd for providing valuable help in sequencing and bioinformatics analysis.

REFERENCES

Aeschimann, F., Xiong, J., Arnold, A., Dieterich, C., and Großhans, H. (2015). Transcriptome-Wide Measurement of Ribosomal Occupancy by Ribosome Profiling. *Methods* 85, 75–89. doi:10.1016/j.ymeth.2015.06.013

Ahmed, N., Sormanni, P., Ciryam, P., Vendruscolo, M., Dobson, C. M., and O'Brien, E. P. (2019). Identifying A- and P-Site Locations on Ribosome-Protected mRNA Fragments Using Integer Programming. *Sci. Rep.* 9 (1), 6256. doi:10.1038/s41598-019-42348-x

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* 31 (2), 166–169. doi:10.1093/bioinformatics/btu638

Arranz, J. J., Bayón, Y., and Primitivo, F. S. (2001). Genetic Variation at Microsatellite Loci in Spanish Sheep. *Small Ruminant Res.* 39 (1), 3–10. doi:10.1016/s0921-4488(00)00164-4

Bai, W. L., Dang, Y. L., Wang, J. J., Yin, R. H., Wang, Z. Y., Zhu, Y. B., et al. (2016). Molecular Characterization, Expression and Methylation Status Analysis of BMP4 Gene in Skin Tissue of Liaoning cashmere Goat during Hair Follicle Cycle. *Genetica* 144 (4), 457–467. doi:10.1007/s10709-016-9914-1

Bazzini, A. A., Lee, M. T., and Giraldez, A. J. (2012). Ribosome Profiling Shows that miR-430 Reduces Translation before Causing mRNA Decay in Zebrafish. *Science* 336 (6078), 233–237. doi:10.1126/science.1215704

Cai, Y., Fu, W., Cai, D., Heller, R., Zheng, Z., Wen, J., et al. (2020). Ancient Genomes Reveal the Evolutionary History and Origin of Cashmere-Producing Goats in China. *Mol. Biol. Evol.* 37 (7), 2099–2109. doi:10.1093/molbev/msaa103

Calviello, L., and Ohler, U. (2017). Beyond Read-Counts: Ribo-Seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet.* 33 (10), 728–744. doi:10.1016/j.tig.2017.08.003

Caputo, V., Strafella, C., Termine, A., Campione, E., Bianchi, L., Novelli, G., et al. (2020). RNAseq-Based Prioritization Revealed COL6A5, COL8A1, COL10A1 and MIR146A as Common and Differential Susceptibility Biomarkers for Psoriasis and Psoriatic Arthritis: Confirmation from Genotyping Analysis of 1417 Italian Subjects. *Int. J. Mol. Sci.* 21 (8), 2740. doi:10.3390/ijms21082740

Courtes, F. C., Lin, J., Lim, H. L., Ng, S. W., Wong, N. S. C., Koh, G., et al. (2013). Translatome Analysis of CHO Cells to Identify Key Growth Genes. *J. Biotechnol.* 167 (3), 215–224. doi:10.1016/j.jbiotec.2013.07.010

Cox, B., Kislinger, T., and Emili, A. (2005). Integrating Gene and Protein Expression Data: Pattern Analysis and Profile Mining. *Methods* 35 (3), 303–314. doi:10.1016/j.ymeth.2004.08.021

Dai, B., Zhang, M., Yuan, J.-L., Ren, L.-Q., Han, X.-Y., and Liu, D.-J. (2019). Integrative Analysis of Methylation and Transcriptional Profiles to Reveal the Genetic Stability of Cashmere Traits in the Tβ4 Overexpression of Cashmere Goats. *Animals* 9 (12), 1002. doi:10.3390/ani9121002

Duan, Y., Liu, G., Sun, Y., Wu, J., Xiong, Z., Jin, T., et al. (2020). Collagen Type VI α5 Gene Variations May Predict the Risk of Lung Cancer Development in Chinese Han Population. *Sci. Rep.* 10 (1), 5010. doi:10.1038/s41598-020-61614-x

Fitzgerald, J., Rich, C., Zhou, F. H., and Hansen, U. (2008). Three Novel Collagen VI Chains, α4(VI), α5(VI), and α6(VI). *J. Biol. Chem.* 283 (29), 20170–20180. doi:10.1074/jbc.m710139200

Fu, X., Zhao, B., Tian, K., Wu, Y., Suo, L., Ba, G., et al. (2020). Integrated Analysis of lncRNA and mRNA Reveals Novel Insights into cashmere Fineness in Tibetan cashmere Goats. *PeerJ* 8, e10217. doi:10.7717/peerj.10217

Gebauer, F., and Hentze, M. W. (2004). Molecular Mechanisms of Translational Control. *Nat. Rev. Mol. Cel Biol* 5 (10), 827–835. doi:10.1038/nrm1488

Gerashchenko, M. V., Lobanov, A. V., and Gladyshev, V. N. (2012). Genome-Wide Ribosome Profiling Reveals Complex Translational Regulation in Response to Oxidative Stress. *Proc. Natl. Acad. Sci.* 109 (43), 17394–17399. doi:10.1073/pnas.1120799109

Gross, S. M., Martin, J. A., Simpson, J., Abraham-Juarez, M. J., Wang, Z., and Visel, A. (2013). De Novo transcriptome Assembly of Drought Tolerant CAM Plants, Agave Deserti and Agave Tequilana. *BMC Genomics* 14, 563. doi:10.1186/1471-2164-14-563

Haq, F., Ahmed, N., and Qasim, M. (2019). Comparative Genomic Analysis of Collagen Gene Diversity. *3 Biotech.* 9 (3), 83. doi:10.1007/s13205-019-1616-9

He, H., Suryawanshi, H., Morozov, P., Gay-Mimbrera, J., Del Duca, E., Kim, H. J., et al. (2020). Single-Cell Transcriptome Analysis of Human Skin Identifies Novel Fibroblast Subpopulation and Enrichment of Immune Subsets in Atopic Dermatitis. *J. Allergy Clin. Immunol.* 145 (6), 1615–1628. doi:10.1016/j.jaci.2020.01.042

Hou, J., Wang, X., McShane, E., Zauber, H., Sun, W., Selbach, M., et al. (2015). Extensive Allele-Specific Translational Regulation in Hybrid Mice. *Mol. Syst. Biol.* 11 (8), 825. doi:10.15252/msb.156240

Huang, Z., Xu, A., and Cheung, B. M. Y. (2017). The Potential Role of Fibroblast Growth Factor 21 in Lipid Metabolism and Hypertension. *Curr. Hypertens. Rep.* 19 (4), 28. doi:10.1007/s11906-017-0730-5

Hui, T. Y. (2020). *Correlation Analysis of cashmere Goat Skin ceRNA Regulatory Network and KRT26 and cashmere Fineness [D]*. China: Shenyang Agricultural Uinversity. doi:10.27327/d.cnki.gshnu.2020.000662

Jia, B., Xi, J. F., Zhang, S. Y., Zhao, Z. S., Zhao, R. Q., and Chen, J. (2006). The Developmental Patterns of GH-R, IGF-1 and IGF-IR Gene Expression in Sheep Skin. *Yi Chuan* 28 (9), 1078–1082. Chinese. doi:10.1360/yc-006-1078

Jin, M., Wang, L., Li, S., Xing, M. X., and Zhang, X. (2011). Characterization and Expression Analysis of KAP7.1, KAP8.2 Gene in Liaoning New-Breeding cashmere Goat Hair Follicle. *Mol. Biol. Rep.* 38 (5), 3023–3028. doi:10.1007/s11033-010-9968-6

Jin, M., Cao, Q., Wang, R., Piao, J., Zhao, F., and Piao, J. (2017). Molecular Characterization and Expression Pattern of a Novel Keratin-Associated Protein 11.1 Gene in the Liaoning cashmere Goat (*Capra Hircus*). *Asian-australas J. Anim. Sci.* 30 (3), 328–337. doi:10.5713/ajas.16.0078

Jin, M., Lu, J., Fei, X., Lu, Z., Quan, K., Liu, Y., et al. (2020). Genetic Signatures of Selection for Cashmere Traits in Chinese Goats. *Animals* 10 (10), 1905. doi:10.3390/ani10101905

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for Linking Genomes to Life and the Environment. *Nucleic Acids Res.* 36 (Database issue), D480–D484. doi:10.1093/nar/gkm882

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biol.* 14 (4), R36. doi:10.1186/gb-2013-14-4-r36

King, R. W., Deshaies, R. J., Peters, J.-M., and Kirschner, M. W. (1996). How Proteolysis Drives the Cell Cycle. *Science* 274 (5293), 1652–1659. doi:10.1126/science.274.5293.1652

Kirkpatrick, D. S., Denison, C., and Gygi, S. P. (2005). Weighing in on Ubiquitin: the Expanding Role of Mass-Spectrometry-Based Proteomics. *Nat. Cel Biol* 7 (8), 750–757. doi:10.1038/ncb0805-750

Li, J., Wang, X., Zheng, K., Liu, Y., Li, J., Wang, S., et al. (2019). The Clinical Significance of Collagen Family Gene Expression in Esophageal Squamous Cell Carcinoma. *PeerJ* 7, e7705. doi:10.7717/peerj.7705

Li, K., Hope, C. M., Wang, X. A., and Wang, J.-P. (2020). RiboDiPA: A Novel Tool for Differential Pattern Analysis in Ribo-Seq Data. *Nucleic Acids Res.* 48 (21), 12016–12029. doi:10.1093/nar/gkaa1049

Li, L. (2009). *Correlation between KIFI Gene Polymorphism and cashmere Traits of Three cashmere Goat Breeds [D]*. Gansu Agricultural University. In Chinese.

Liu, H., Yang, G., Zhang, W., Zhu, X., and Jia, Z. (2009). Effects of FGF5 Gene on Cashmere Traits of Inner Mongolia Cashmere. *Genetics* 31 (02), 175–179. doi:10.3724/sp.j.1005.2009.00175

Liu, G., Liu, R., Tang, X., Cao, J., Zhao, S., and Yu, M. (2015). Expression Profiling Reveals Genes Involved in the Regulation of Wool Follicle Bulb Regression and Regeneration in Sheep. *Int. J. Mol. Sci.* 16 (5), 9152–9166. doi:10.3390/ijms16059152

Mahata, B., Sundqvist, A., and Xirodimas, D. P. (2012). Recruitment of RPL11 at Promoter Sites of P53-Regulated Genes upon Nucleolar Stress through NEDD8 and in an Mdm2-Dependent Manner. *Oncogene* 31 (25), 3060–3071. doi:10.1038/onc.2011.482

Maiese, K. (2015). Novel Applications of Trophic Factors, Wnt and WISP for Neuronal Repair and Regeneration in Metabolic Disease. *Neural Regen. Res.* 10 (4), 518–528. doi:10.4103/1673-5374.155427

Martinelli-Boneschi, F., Colombi, M., Castori, M., Devigili, G., Eleopra, R., Malik, R. A., et al. (2017). COL6A5 Variants in Familial Neuropathic Chronic Itch. *Brain* 140 (3), 555–567. doi:10.1093/brain/aww343

McManus, C. J., May, G. E., Spealman, P., and Shteyman, A. (2014). Ribosome Profiling Reveals Post-Transcriptional Buffering of Divergent Gene Expression in Yeast. *Genome Res.* 24 (3), 422–430. doi:10.1101/gr.164996.113

Mendoza, M. N., Raudsepp, T., Alshanbari, F., Gutiérrez, G., and Ponce de León, F. A. (2019). Chromosomal Localization of Candidate Genes for Fiber Growth and Color in Alpaca (*Vicugna Pacos*). *Front. Genet.* 10, 583. doi:10.3389/fgene.2019.00583

Rufaut, N. W., Pearson, A. J., Nixon, A. J., Wheeler, T. T., and Wilkins, R. J. (1999). Identification of Differentially Expressed Genes during a Wool Follicle Growth Cycle Induced by Prolactin. *J. Invest. Dermatol.* 113 (6), 865–872. doi:10.1046/j.1523-1747.1999.00775.x

Sabatelli, P., Gara, S. K., Grumati, P., Urciuolo, A., Gualandi, F., Curci, R., et al. (2011). Expression of the Collagen VI α5 and α6 Chains in Normal Human Skin and in Skin of Patients with Collagen VI-Related Myopathies. *J. Invest. Dermatol.* 131 (1), 99–107. doi:10.1038/jid.2010.284

Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., et al. (2017). Translation from Unconventional 5′ Start Sites Drives Tumour Initiation. *Nature* 541 (7638), 494–499. doi:10.1038/nature21036

Sonenberg, N., and Hinnebusch, A. G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* 136 (4), 731–745. doi:10.1016/j.cell.2009.01.042

Stadler, M., and Fire, A. (2013). Conserved Translatome Remodeling in Nematode Species Executing a Shared Developmental Transition. *Plos Genet.* 9 (10), e1003739. doi:10.1371/journal.pgen.1003739

Ständer, S., Schäfer, I., Phan, N. Q., Blome, C., Herberger, K., Heigel, H., et al. (2010). Prevalence of Chronic Pruritus in Germany: Results of a Cross-Sectional Study in a Sample Working Population of 11,730. *Dermatology* 221 (3), 229–235. doi:10.1159/000319862

Strafella, C., Caputo, V., Minozzi, G., Milano, F., Arcangeli, M., Sobhy, N., et al. (2019). Atopic Eczema: Genetic Analysis of COL6A5, COL8A1, and COL10A1 in Mediterranean Populations. *Biomed. Res. Int.* 2019, 3457898. doi:10.1155/2019/3457898

Tang, Y., Ghosal, S., and Roy, A. (2007). Nonparametric Bayesian Estimation of Positive False Discovery Rates. *Biometrics* 63 (4), 1126–1134. doi:10.1111/j.1541-0420.2007.00819.x

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., et al. (2004). Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Mol. Cell Proteomics* 3 (10), 960–969. doi:10.1074/mcp.m400055-mcp200

Tian, W. (2015). Quality Analysis of Chinese Cashmere. *China Fiber Inspect.* 2 (03), 22–25. doi:10.14162/j.cnki.11-4772/t.2015.03.011

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: An R Package for Identifying Differentially Expressed Genes from RNA-Seq Data. *Bioinformatics* 26 (1), 136–138. doi:10.1093/bioinformatics/btp612

Wang, X., Pandey, A. K., Mulligan, M. K., Williams, E. G., Mozhui, K., Li, Z., et al. (2016). Joint Mouse-Human Phenome-Wide Association to Test Gene Function and Disease Risk. *Nat. Commun.* 7, 10464. doi:10.1038/ncomms10464

Wang, S. H., Hsiao, C. J., Khan, Z., and Pritchard, J. K. (2018). Post-Translational Buffering Leads to Convergent Protein Expression Levels between Primates. *Genome Biol.* 19 (1), 83. doi:10.1186/s13059-018-1451-z

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., et al. (2020). Transcriptome and Translatome Co-Evolution in Mammals. *Nature* 588 (7839), 642–647. doi:10.1038/s41586-020-2899-z

Wang, Y., Li, G., Zhang, X., Zheng, Y., Guo, S., Guo, D., et al. (2021). Analysis of m6A Methylation in Skin Tissues of Different Sex Liaoning Cashmere Goats. *Anim. Biotechnol.* 25, 1–11. doi:10.1080/10495398.2021.1962897

Weisshaar, E., and Dalgard, F. (2009). Epidemiology of Itch: Adding to the Burden of Skin Morbidity. *Acta Derm Venerol* 89 (4), 339–350. doi:10.2340/00015555-0662

Xiao, Z., Huang, R., Xing, X., Chen, Y., Deng, H., and Yang, X. (2018). De Novo Annotation and Characterization of the Translatome with Ribosome Profiling Data. *Nucleic Acids Res.* 46 (10), e61. doi:10.1093/nar/gky179

Yang, J.-B., Gan, S.-Q., Yang, Y.-L., Zhang, H.-L., Song, T.-Z., Feng, J., et al. (2012). Cloning and Expression in Follicle Anagen of ILK Gene in Sheep. *Hereditas (Beijing)* 34 (6), 719–726. Chinese. doi:10.3724/sp.j.1005.2012.00719

Yang, F., Liu, Z., Zhao, M., Mu, Q., Che, T., Xie, Y., et al. (2020). Skin Transcriptome Reveals the Periodic Changes in Genes Underlying cashmere (Ground Hair) Follicle Transition in cashmere Goats. *BMC Genomics* 21 (1), 392. doi:10.1186/s12864-020-06779-5

Yu, Z., Wildermoth, J. E., Wallace, O. A. M., Gordon, S. W., Maqbool, N. J., Maclean, P. H., et al. (2011). Annotation of Sheep Keratin Intermediate Filament Genes and Their Patterns of Expression. *Exp. Dermatol.* 20 (7), 582–588. doi:10.1111/j.1600-0625.2011.01274.x

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. (clusterfile). doi:10.1089/omi.2011.0118

Zeng, X.-C., Chen, H.-Y., Jia, B., Zhao, Z.-S., Hui, W.-Q., Wang, Z.-B., et al. (2011). Identification of SNPs within the Sheep PROP1 Gene and Their Effects on Wool Traits. *Mol. Biol. Rep.* 38 (4), 2723–2728. doi:10.1007/s11033-010-0416-4

Zhang, B., Chang, L., Lan, X., Asif, N., Guan, F., Fu, D., et al. (2018). Genome-Wide Definition of Selective Sweeps Reveals Molecular Evidence of Trait-Driven Domestication Among Elite Goat (Capra Species) Breeds for the Production of Dairy, cashmere, and Meat. *Gigascience* 7 (12), giy105. doi:10.1093/gigascience/giy105

Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T., and Zhang, G. (2019). Translatomics: The Global View of Translation. *Int. J. Mol. Sci.* 20 (1), 212. doi:10.3390/ijms20010212

Zheng, Y. Y., Sheng, S. D., Hui, T. Y., Yue, C., Sun, J. M., Guo, D., et al. (2019). An Integrated Analysis of Cashmere Fineness lncRNAs in Cashmere Goats. *Genes* 10 (4), 266. doi:10.3390/genes10040266

Zheng, Y., Hui, T., Yue, C., Sun, J., Guo, D., Guo, S., et al. (2020). Comprehensive Analysis of circRNAs from cashmere Goat Skin by Next Generation RNA Sequencing (RNA-Seq). *Sci. Rep.* 10 (1), 516. doi:10.1038/s41598-019-57404-9

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 Veterinary Science

ORIGINAL RESEARCH
published: 08 February 2022
doi: 10.3389/fvets.2022.834566

Integrative Analysis of the lncRNA-Associated ceRNA Regulatory Network Response to Hypoxia in Alveolar Type II Epithelial Cells of Tibetan Pigs

*Yanan Yang[1], Yongqing Li[2], Haonan Yuan[1], Xuanbo Liu[1], Yue Ren[3], Caixia Gao[4], Ting Jiao[1,5], Yuan Cai[1] and Shengguo Zhao[1]**

[1] College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China, [2] Xinjiang Academy of Animal Sciences, Urumqi, China, [3] Institute of Animal Husbandry and Veterinary Medicine, Academy of Agriculture and Animal Husbandry Sciences, Lhasa, China, [4] State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China, [5] College of Grassland Science, Gansu Agricultural University, Lanzhou, China*

OPEN ACCESS

Edited by:
Klaus Wimmers,
Leibniz Institute for Farm Animal Biology (FBN), Germany

Reviewed by:
Kun Li,
Nanjing Agricultural University, China
Ran Di,
Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Shengguo Zhao
zhaoshengguo0628@hotmail.com

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

Received: 13 December 2021
Accepted: 14 January 2022
Published: 08 February 2022

Citation:
Yang Y, Li Y, Yuan H, Liu X, Ren Y, Gao C, Jiao T, Cai Y and Zhao S (2022) Integrative Analysis of the lncRNA-Associated ceRNA Regulatory Network Response to Hypoxia in Alveolar Type II Epithelial Cells of Tibetan Pigs. Front. Vet. Sci. 9:834566. doi: 10.3389/fvets.2022.834566

The function of alveolar type II epithelial (ATII) cells is severely hampered by oxygen deficiency, and understanding the regulatory mechanisms controlling responses to hypoxia may assist in relieving injury induced by hypoxia. In this study, we cultured ATII cells from Tibetan pigs and Landrace pigs under hypoxic and normoxic environments to screen for differentially expressed (DE) lncRNAs, DEmiRNAs, and construct their associated ceRNA regulatory networks in response to hypoxia. Enrichment analysis revealed that target genes of DElncRNAs of Tibetan pigs and Landrace pig between the normoxic (TN, LN) and hypoxic (TL, LL) groups significantly enriched in the proteoglycans in cancer, renal cell carcinoma, and erbB signaling pathways, while the target genes of DEmiRNAs were significantly enriched in the axon guidance, focal adhesion, and mitogen-activated protein kinase (MAPK) signaling pathways. Hypoxia induction was shown to potentially promote apoptosis by activating the focal adhesion/PI3K-Akt/glycolysis pathway. The ssc-miR-20b/MSTRG.57127.1/ssc-miR-7-5p axis potentially played a vital role in alleviating hypoxic injury by regulating ATII cell autophagy under normoxic and hypoxic conditions. MSTRG.14861.4-miR-11971-z-CCDC12, the most affected axis, regulated numerous RNAs and may thus regulate ATII cell growth in Tibetan pigs under hypoxic conditions. The ACTA1/ssc-miR-30c-3p/MSTRG.23871.1 axis is key for limiting ATII cell injury and improving dysfunction and fibrosis mediated by oxidative stress in Landrace pigs. Our findings provide a deeper understanding of the lncRNA/miRNA/mRNA regulatory mechanisms of Tibetan pigs under hypoxic conditions.

Keywords: hypoxia, ATII cells, swine, ceRNA network, ssc-miR-20b/MSTRG.57127.1/ssc-miR-7-5p axis

INTRODUCTION

The breeding of Tibetan pigs by humans can be traced back to the seventh century, and these animals adapted to hypoxic and low-pressure environments due to long-term natural selection on the Tibetan Plateau (1–3). Tibetan pigs harbor a distinct fungal community (4) and immunomodulatory function (5) and have developed organs that are adaptable to hypoxia induced by high altitudes to facilitate oxygen delivery under hypobaric hypoxic conditions. Alveolar type II (ATII) cells produce, secrete, and recycle a variety of pulmonary surfactant proteins and chemokines to recruit macrophages, eliminate inflammatory processes, and decrease surface tension as a protective barrier against possible infections and provide efficient ventilation (6). Alveolar type II cell apoptosis is induced by the alveolar lining layer; destructive exudative changes occur, and lung edema occurs in the early phase of hypoxia; conversely, reactive hyperplasia of ATII cells is predominant in the later phase of alveolar hypoxia (7). Increasing evidence suggests that alveolar epithelial cells undergo apoptosis during pathological hypoxic lung injury and fibrosis, which correlates with their reduced ability to proliferate and restore the alveolar architecture (8–10). The regulation of hypoxia-related genes enables cellular adaptation to a hypoxic environment via increased glucose metabolism (11), cell proliferation and migration (12), increased survival (13), as well as by the promotion of angiogenesis (14). Advances in sequencing technology have revealed mRNA expression levels are regulated by miRNAs, lncRNAs, or circRNAs, and analysis of these factors could provide comprehensive insight into the complicated cellular biological process of hypoxia as a rapid approach to reveal the regulatory network (15, 16). Existing literature shows that lncRNAs and miRNAs participate in diverse biological processes in mammals, such as development, growth, immunity, and reproduction (17–19). lncRNAs often bind to the proximal promoters of protein-coding genes to regulate protein expression, and lncRNA FAF overexpression was shown to significantly inhibit cardiomyocyte apoptosis induced by hypoxia (20). Moreover, lncRNA BCRT1 is upregulated in response to hypoxia and has potential as a biomarker and therapeutic target for breast cancer (21). RMRP regulates hypoxia-induced injury by modulating the p53 signaling pathway, which is a direct target of miR-214-5p (22). In Tibetan pigs, mounting research evidence has also shown that RNAs are abundantly expressed in the lung tissue transcriptome in response to hypoxia (23), but no integrated analyses of how their networks regulate ATII cells have been performed. Thus, we compared the expression levels of mRNAs, miRNAs, and lncRNAs in the ATII cells of Tibetan and Landrace pigs under normoxic and hypoxic conditions. Then, we constructed and preliminarily verified lncRNA/miRNA/mRNA regulatory networks to identify the key factors involved in the ATII cell response to hypoxia.

Abbreviations: DElncRNAs, differentially expressed lncRNAs; DEmiRNAs, differentially expressed miRNAs; DEmRNAs, differentially expressed mRNAs; TN, ATII cells of Tibetan pigs were cultured under 21% O_2; TL, ATII cells of Tibetan pigs were cultured under 2% O_2; LN, ATII cells of Landrace pigs were cultured under 21% O_2; LL, ATII cells of Landrace pigs were cultured under 2% O_2.

MATERIALS AND METHODS

Sample Collection

Alveolar type II cell samples from Tibetan pigs (TN, $n = 4$) and Landrace pigs (LN, $n = 4$) were collected at 48 h after culture in 21% O_2, 5% CO_2, 79% N_2, and 37°C under normoxic conditions, and control group samples from Tibetan pigs (TL, $n = 4$) and Landrace pigs (LL, $n = 4$) were acquired at parallel time points at 2% O_2, 5% CO_2, 98% N_2, and 37°C under hypoxic conditions. Three of each group were flash-frozen in liquid nitrogen for RNA extraction, and the rest were used for the flow cytometric assay.

Total RNA Isolation and Illumina Sequencing

Total RNA was isolated from ATII cells using TRIzol reagent kit (Invitrogen, Carlsbad, CA, USA), and RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). All samples had an RNA integrity number (RIN) > 8. The enriched mRNAs and lncRNAs were reverse transcribed into cDNA, purified with the QiaQuick PCR extraction kit (Qiagen, Venlo, The Netherlands), end repaired, subjected to poly(A) addition, and ligated to Illumina sequencing adaptors. cDNA libraries were filtered and selected from the digested products and sequenced on the Illumina HiSeq^{TM} 4000 platform (or other platforms) by Gene Denovo Biotechnology Co. (Guangzhou, China).

RNA molecules were enriched for miRNA sequencing (miRNA-seq) by polyacrylamide gel electrophoresis (PAGE) at a size range of 18–30 nt. A 3′ adaptor was added to enrich the 36–44 nt RNAs, and the 5′ adaptor was then connected to the RNA. The ligation products were reverse transcribed by PCR amplification, and the 140–160 bp PCR products were enriched to generate a cDNA library and sequenced using Illumina HiSeq Xgten by Gene Denovo Biotechnology Co. (Guangzhou, China).

Identification of DERNAs

The protocols for mRNA and lncRNA identification were as follows: reads were filtered by fastp (24) (version 0.18.0) to obtain clean reads and further mapped to the ribosomal RNA (rRNA) database by the short read alignment tool Bowtie 2 (25). The reads were removed from the rRNA and used to assemble and analyze the transcriptome. Paired-end clean reads were mapped to the *Sus scrofa* RefSeq using HISAT2 (26). Transcript reconstruction was carried out with StringTie software (version 1.3.4) (https://ccb.jhu.edu/software/stringtie/index.shtml) (27) and HISAT2 (http://ccb.jhu.edu/software/hisat/index.shtml) (28) to identify new genes and new splice variants of known genes. The protein-coding potential of novel transcripts was assessed using CNCI (29) and CPC (30), and their intersection was used to identify long non-coding RNAs. The expression abundances of mRNA and lncRNA were calculated according to fragment per kilobase of transcript per million mapped reads (FPKM) values using StringTie software. We used a false discovery rate (FDR) < 0.05 and absolute fold change ≥2 as the thresholds for identifying

differentially expressed mRNAs (DEmRNAs) and differentially expressed lncRNAs (DElncRNAs) using DESeq2 (31).

Potential miRNAs were identified from raw reads, filtered to generate clean reads and aligned with small RNAs in the GenBank database. Clean reads were aligned with small RNAs in the Rfam database (32) to identify and remove others, which were also aligned with the reference genome (*Sus scrofa*) and searched against the miRbase database (33) to identify known miRNAs. The miRNA expression levels were calculated and normalized to transcripts per million (TPM) values. We identified miRNAs with fold changes $\geq$2 and *P*-values <0.05 as significant DEmiRNAs.

GO and KEGG Pathway Enrichment Analyses

The database for annotation, visualization, and integrated discovery (DAVID) was utilized to conduct Gene Ontology (GO) functional annotation (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (http://www.genome.jp/kegg/pathway.html) enrichment analyses to determine the roles, functions, and enrichment of different RNA biological pathways. GO terms and pathways with a q-value < 0.05 were considered significantly enriched.

Integrated Analysis of lncRNAs, miRNAs, and mRNAs

The DElncRNAs, DEmiRNAs, and mRNAs were analyzed using DESeq2 and miRanda. Target genes of lncRNAs were identified by analyzing the correlation between the expression levels of lncRNAs and protein-coding genes by RNAplex. Target genes of miRNAs were predicted by RNAhybrid 89 (version 2.1.2) + svm_light (version 6.01), miRanda (version 3.3a), and TargetScan (version 7.0). Furthermore, a coexpression network diagram of the lncRNAs, miRNAs, and mRNAs was generated by assembling all coexpression competing triplets and visualized using Cytoscape software (34).

Flow Cytometric Assay of Cell Apoptosis

Alveolar type II cells were collected and washed with PBS for the quantitative analysis of cell apoptosis. Furthermore, annexin V-FITC and 10 μl of PI staining solution were added to the cells, which were resuspended in Annexin V-FITC binding buffer at room temperature in the dark for 20 min. A BD FACSCanto II flow cytometer (BD Biosciences, San Jose, CA, USA) was used to analyze cell apoptosis, and the fractions of the cell populations in different quadrants were analyzed using quadrant statistics.

Quantitative Real-Time-qPCR Analysis

To verify the RNA-seq results, first-strand cDNA was synthesized from RNA samples returned by transcriptome sequencing using a FastQuant cDNA first-strand synthesis kit (TianGen, China), and cDNA was used as the template for gene expression analysis. qPCR was performed on a LightCycler 96 Real-Time System (Roche, Switzerland) using SYBR® Premix Ex Taq™ II (TaKaRa, China). Eight DElncRNAs, four DEmiRNAs, and six DEmRNAs were selected to determine the reliability of the data, and the amplification primers are listed (**Supplementary Table 1** of **Supplementary Material 1**).

RESULTS

Overview of RNA Sequencing

Twelve cDNA libraries were sequenced for lncRNA-seq and mRNA-seq analysis of ATII cells (**Supplementary Figure 1** of **Supplementary Material 2**). Averages of 100,489,128, 97,339,785, 64,417,965, and 44,846,728 raw reads were acquired from the TN, TL, LN, and LL groups, respectively. The raw reads, clean reads, clean bases, error rates, Q20 values, Q30 values, and GC contents of each library are shown (**Supplementary Table 2** of **Supplementary Material 1**). After quality filtering, an average of 44,821,226 clean reads were mapped to the porcine reference genome. lncRNAs were classified based on their genomic location, and the 10,964 lncRNAs (9,280 known lncRNAs and 1,684 novel lncRNAs) consisted of 62.30% long intergenic non-coding RNAs and 14.74% antisense lncRNAs but a minimal of intronic lncRNAs (**Figure 1A**; **Supplementary Table 3** of **Supplementary Material 1**). Another 12 libraries were constructed for miRNA-seq analysis of ATII cells. A total of 9,947,404–13,467,206 raw reads were generated, and 9,847,843–13,350,909 high-quality RNA sequences were obtained after removing reads at a suitable level, which accounted for more than 98.86% of the clean reads (**Supplementary Table 4** of **Supplementary Material 1**).

A total of 325 DElncRNAs (200 up- and 125 downregulated) and 124 DEmiRNAs (78 up- and 46 downregulated) were identified in the LN group compared to the LL group (**Figures 1B**, **2A**). Cluster analysis of DElncRNAs was conducted, and the results are shown as a heatmap (**Figures 1C,D**). Then, the most interesting candidates, including 80 DElncRNAs and 37 DEmiRNAs were identified and screened based on the intersection between the normoxic (TN vs. TL) and hypoxic (TL vs. LL) groups to assess the regulation of RNA responses to hypoxia (**Figures 1E**, **2B**). Eight lncRNAs, four miRNAs, and six mRNAs were randomly selected and detected using quantitative real-time (qRT)-PCR to validate the results (**Supplementary Figure 2** of **Supplementary Material 2**).

Prediction of DElncRNA- and DEmiRNA-Targeted mRNAs

In total, 1,716 (1,616 lncRNA and 1,054 mRNA) and 72 (43 lncRNA and 29 mRNA) lncRNA-mRNA pairs from total lncRNAs and DElncRNAs were obtained among the four groups (**Supplementary Materials 3**). The 2,679 target mRNAs of 157 DEmiRNAs were analyzed between the TN and TL groups (**Supplementary Materials 4**). Specifically, the target mRNAs of DElncRNAs and DEmiRNAs between the normoxic (TN vs. TL) and hypoxic (TL vs. LL) groups were examined to further investigate the potential lncRNA–mRNA and miRNA–mRNA interactions in response to hypoxia induction and to establish the potential roles of the lncRNAs and miRNAs in hypoxia adaptation (**Figure 2C**).

Functional Annotation of Target mRNAs

To further investigate the potential functions of DEmiRNAs, we performed GO enrichment and KEGG pathway analyses of their target mRNAs among the four groups (**Supplementary Figure 3**

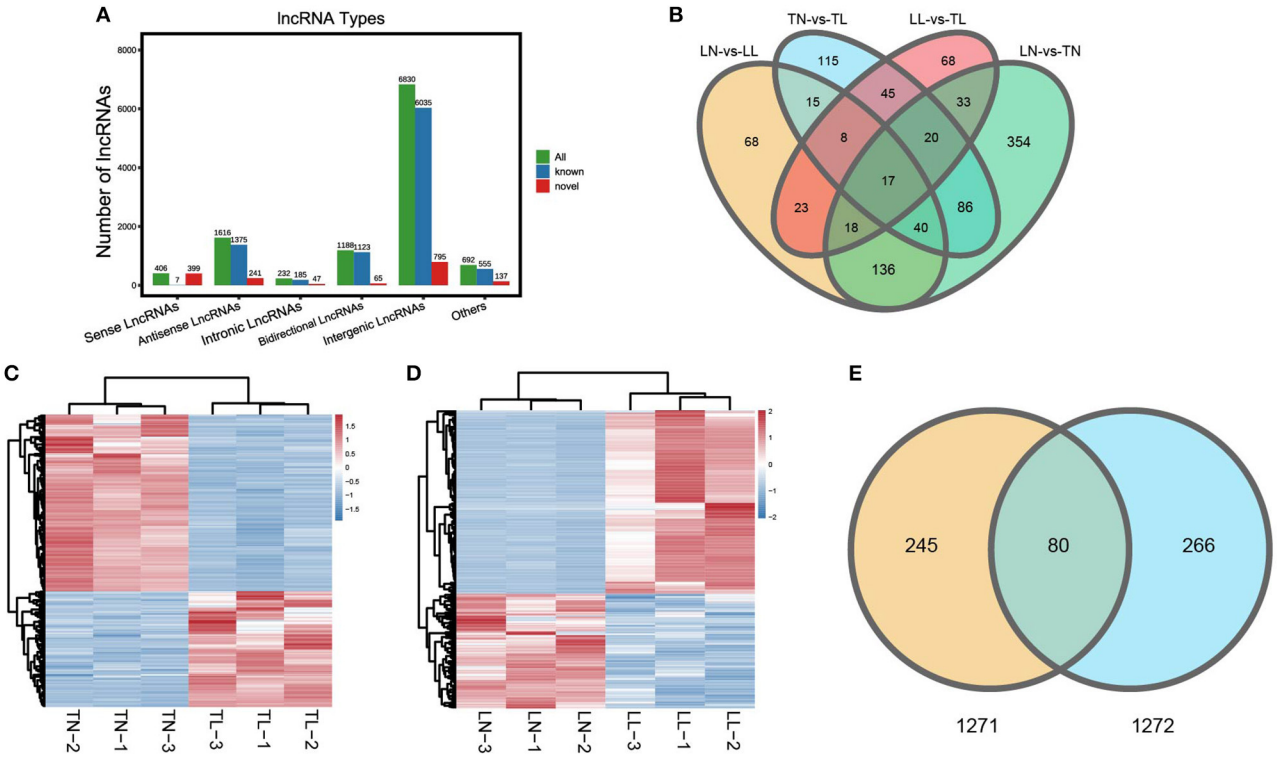


FIGURE 1 | Categories and features of the lncRNAs in ATII cells of Tibetan pigs and Landrace pigs. **(A)** Categories and types of lncRNAs originating from different genomic sites. **(B)** Venn diagram of lncRNAs interactions based on the overlapping DElncRNAs. **(C)** Heatmap of lncRNAs in ATII cells between the TN and TL groups. **(D)** Heatmap of lncRNAs in ATII cells between the LN and LL groups. **(E)** Common DElncRNAs shared between the normoxic (TN and LN) and hypoxic (TL and LL) groups.

of **Supplementary Material 2**). In this study, numerous target mRNAs of DElncRNAs (DEmiRNAs) were significantly enriched for the intracellular part (intracellular), binding (binding), and macromolecule metabolic process (localization) of cellular component, molecular function, and biological process between the normoxic (TN, LN) and hypoxic (TL, LL) groups, respectively (**Figures 3A,B**). The target mRNAs of DElncRNAs were enriched in 62 and 61 GO categories; moreover, the target mRNAs of DEmiRNAs were enriched in 135 and 130 categories between the Tibetan pigs (excluding DEmiRNAs shared between the normoxia and hypoxia groups) and Landrace pigs (excluding DEmiRNAs shared between the normoxia and hypoxia groups) at different oxygen concentrations (**Figures 3C–F**).

The comparison of normoxic (TN, LN) and hypoxic (TL, LL) groups revealed that the target genes of DElncRNAs were significantly enriched in the proteoglycans in cancer, renal cell carcinoma, and erbB signaling pathways, while the target genes of DEmiRNAs were significantly enriched in the axon guidance, focal adhesion, and MAPK signaling pathways (**Figures 4A,B**). Interestingly, the cell cycle, FOXO signaling, and proteoglycans in cancer pathways were significantly enriched among the target mRNAs of the DElncRNAs; however, the axon guidance, ras signaling, and mTOR signaling pathways were significantly enriched among the target mRNAs of the DEmiRNAs between the TN and TL groups (excluding DEmiRNAs shared between the normoxia and hypoxia groups) (**Figures 4C,D**). Numerous target mRNAs of the DElncRNAs were significantly enriched in the oxidative phosphorylation, Alzheimer's disease, and thermogenesis pathways; the target mRNAs of the DEmiRNAs were significantly enriched in the axon guidance, hepatocellular carcinoma, and focal adhesion pathways between the LN and LL groups (excluding DEmiRNAs shared between the normoxia and hypoxia groups) (**Figures 4E,F**).

lncRNA-miRNA-mRNA Networks Profiles

Three-component (lncRNA/miRNA/mRNA) ceRNA regulatory networks were constructed to identify the most effected RNAs associated with hypoxic adaptation in the ATII cells of Tibetan pigs using intersecting mRNAs of lncRNA–mRNA pairs and miRNA–mRNA pairs (**Figures 2C–E**). Three networks were constructed between (1) the normoxic (TN, LN) and hypoxic (TL, LL) groups, (2) the TN and TL groups (excluding DEmRNAs shared between the normoxia and hypoxia groups), and (3) the LN and LL groups (excluding DEmRNAs shared between the normoxia and hypoxia groups), and the top five gene pairs in each network were revealed. The network between normoxic (TN, LN) and hypoxic (TL, LL) groups contained fifteen lncRNA–miRNA pairs and seven miRNA–mRNA pairs, including nine lncRNAs, three miRNAs, and five mRNAs (**Supplementary Materials 5**). The network between the TN and TL groups (excluding

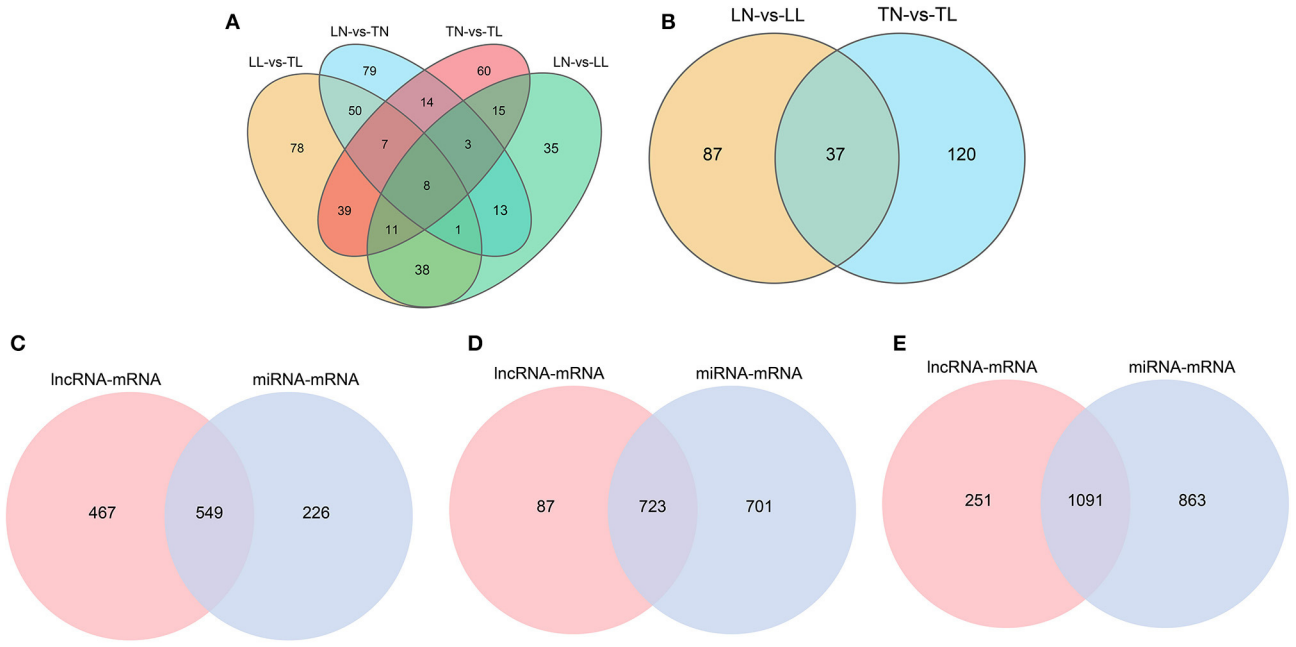


FIGURE 2 | Features of the miRNAs in ATII cells among the four groups. **(A)** Venn diagram of miRNAs interactions based on the overlapping DEmiRNAs. **(B)** Common DEmiRNAs shared between the normoxic (TN and LN) and hypoxic (TL and LL) groups. **(C)** Venn diagram of mRNAs interactions from lncRNA-mRNA and miRNA-mRNA pairs between the normoxic (TN and LN) and hypoxic (TL and LL) groups. **(D)** Venn diagram of mRNAs interactions from lncRNA-mRNA and miRNA-mRNA pairs between the TN and TL groups. **(E)** Venn diagram of mRNAs interactions from lncRNA-mRNA and miRNA-mRNA pairs between the LN and LL groups.

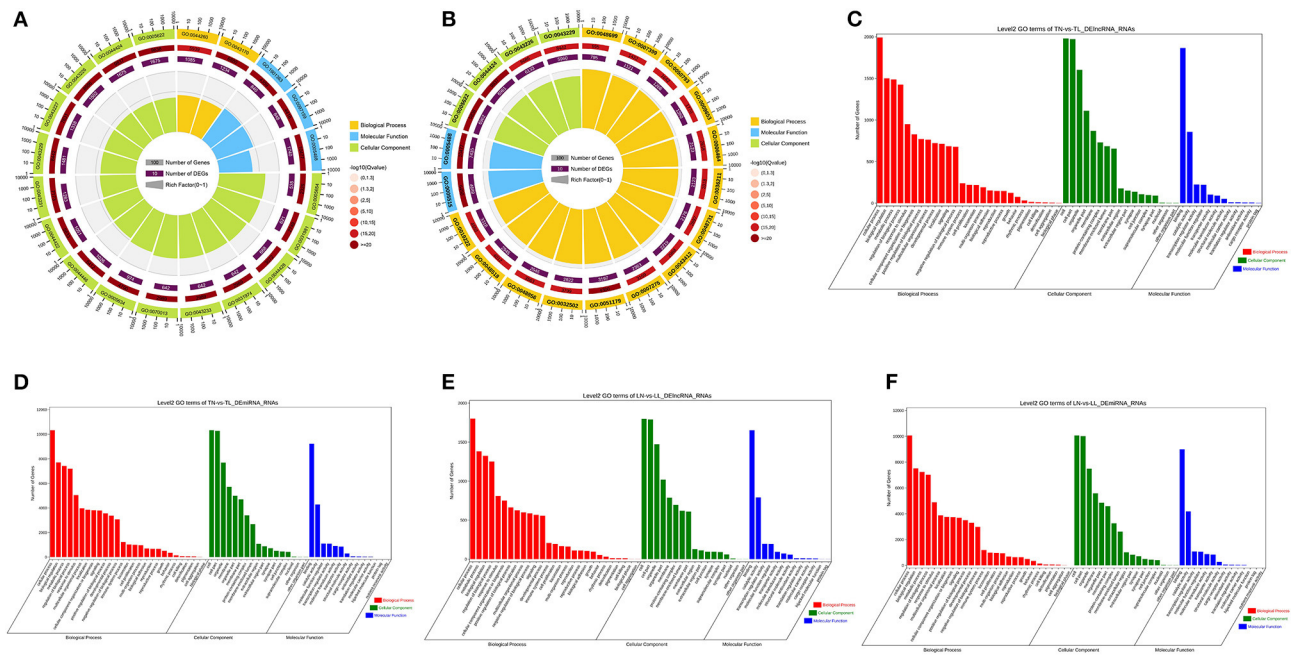


FIGURE 3 | GO functional annotation. Common target mRNAs of DElncRNAs **(A)** and DEmiRNAs **(B)** between the normoxic (TN and LN) and hypoxic (TL and LL) groups. Target mRNAs of DElncRNAs **(C)** and DEmiRNAs **(D)** between the TN and TL groups. Target mRNAs of DElncRNAs **(E)** and DEmiRNA **(F)** of Landrace pigs between the LN and LL groups.

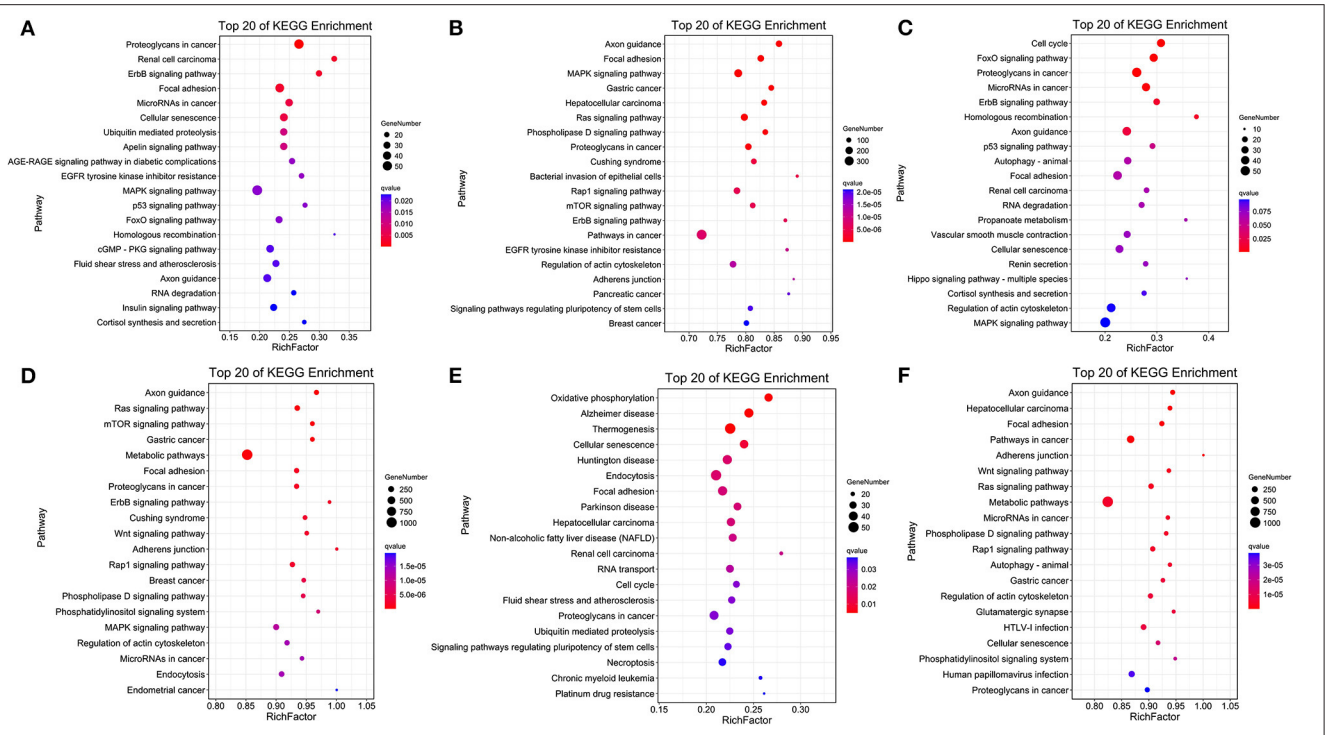


FIGURE 4 | KEGG pathway enrichment analysis. Common target mRNAs of DElncRNAs **(A)** and DEmiRNAs **(B)** between the normoxic (TN and LN) and hypoxic (TL and LL) groups. Target mRNAs of DElncRNAs **(C)** and DEmiRNAs **(D)** between the TN and TL groups. Target mRNAs of DElncRNAs **(E)** and DEmiRNA **(F)** between the LN and LL groups.

DEmRNAs shared between the normoxia and hypoxia groups) indicate an association among 11 lncRNA nodes. Twenty-three miRNA nodes, five mRNA nodes, and 58 edges, including ssc-miR-129b, CCDC12, and ssc-miR-129a-5p, were selected as the most affected RNAs. Here, the network of Landrace pigs (LN vs. LL) was composed of 38 lncRNA-miRNA pairs and 11 mRNA-miRNA pairs, 8 lncRNA nodes, 11 miRNAs node, and 5 mRNAs node, including ssc-miR-30c-3p, C1QC, and MSTRG.23871.1, which were selected as the most affected RNAs (**Figure 5**).

Apoptosis of ATII Cells

Cell apoptosis was investigated by flow cytometric assays, and our results showed that cell apoptosis was higher in the hypoxic groups (TL, LL) than in the normoxic groups (TN, LN). Notably, the rate of cell apoptosis in Tibetan pigs was lower than that in Landrace pigs under the same conditions. The order of total apoptosis rates among the four groups was as follows: TN<LN<TL<LL (**Figure 6**).

DISCUSSION

Under hypoxic conditions, the development of domestic mammals, including Tibetan pigs (35), yaks (36), and Tibetan sheep (3), is severely hampered. The hearts (37), lungs (38), and deciduous teeth (39) of Tibetan pigs adapted better than those of any other pig; these pigs were originally found exclusively on the Tibetan Plateau, which has an average altitude of 4,268 m above sea level (40). A total of 1,716 lncRNA–mRNA pairs were obtained to analyze their associated key regulatory pathways among the four groups (TL, LL, TN, LN), and the result could help to elucidate the regulatory mechanisms of ATII cells in Tibetan pigs in response to hypoxia and provide a theoretical basis for formulating strategies to relieve disease and injury during hypoxia (41). Simultaneously, it is important to understand the responses of ATII cells to hypoxia to reveal a series of adaptation activities in the lung. Hypoxia induces pulmonary injury via cell apoptosis, which leads to transdifferentiation of ATII cells; hence, the regeneration of ATII cells is an important indicator of hypoxia damage (42). Here, we generated comprehensive ceRNA network profiles, including the lncRNAs, miRNAs, and mRNAs of ATII cells. The comparisons between Tibetan pigs and Landrace pigs under normoxic and hypoxic conditions revealed genes and pathways that possibly undergo adaptive changes in response to hypoxia. Notably, the ATII cells of Tibetan pigs and Landrace pigs showed some genes with the same expression patterns and some with different expression patterns, and the majority of DERNAs were involved in network regulation.

Hypoxia Induction May Promote Apoptosis by Activating of the Focal Adhesion/PI3K-Akt/Glycolysis Pathway

Notably, the erbB signaling pathway, focal adhesion, and cellular senescence were clearly associated with different oxygen

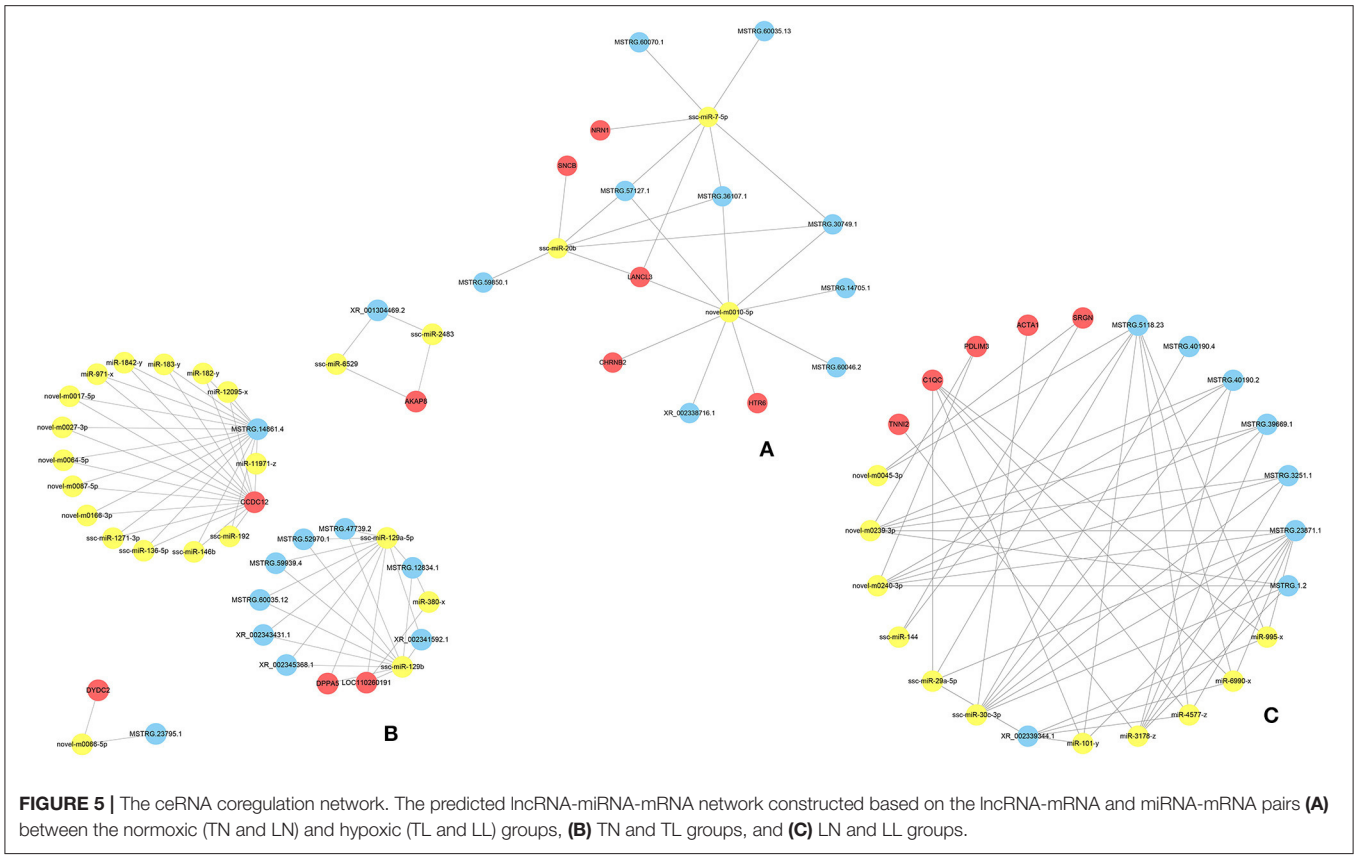


FIGURE 5 | The ceRNA coregulation network. The predicted lncRNA-miRNA-mRNA network constructed based on the lncRNA-mRNA and miRNA-mRNA pairs **(A)** between the normoxic (TN and LN) and hypoxic (TL and LL) groups, **(B)** TN and TL groups, and **(C)** LN and LL groups.

concentrations in ATII cells and enriched for a number of target mRNAs of DElncRNAs between the normoxic and hypoxic groups (**Figure 7**). Collagens that interact with integrins are widely represented in focal adhesion pathways, and focal adhesion kinases (FAKs) are focal adhesion complexes that play a key role in cell-substrate adhesions (43). Various inflammatory mediators and their receptors significant upregulate and affects normal life activities via damage the cell function caused by low oxygen levels (44, 45). In our analysis, we found *COL2A1* to be the target gene of ssc-miR-7-5p, miR-9277-z, and miR-148-z; the upregulation of this gene, along with *COL6A3*, *COL1A2* (target genes of miR-409-y, ssc-miR-1285, and miR-11980-z), and *COL5A1* (target gene of ssc-miR-139-3p, miR-1307-x, and miR-1197-y), indicated the enrichment of the focal adhesion pathway; these genes may link the extracellular matrix (ECM) and cytoskeleton in cells under hypoxia (46). Moreover, α and β chains form integrins that mediate differential cell interactions with specific ECM and cellular surface components, which are linked to arginine-glycine-aspartate (RGD) amino acid motifs as heterodimeric transmembrane adhesion receptors (47). Focal adhesion kinases are activated by integrins and integrate integrin signals by directly binding to signaling molecules, and FAK phosphorylation is important for the maintenance of normal cell adhesion (43, 48). We also identified *ITGA5* (target gene of miR-28-x, miR-411-y, ssc-miR-1285), *ITGA6*, and *ITGAL* (target gene of ssc-miR-326, ssc-miR-1, ssc-miR-9860-5p) as being significantly differentially expressed (DE) between the normoxic and hypoxic groups, and these are downstream genes related to the ECM, proteoglycans in cancer and focal adhesion pathways. *ITGA5* expression was significantly higher in the TL group than in the TN group, but was not significantly different in the LN and LL groups, indicating that it may influence and regulate the apoptosis, differentiation, migration, and proliferation of ATII cells in Tibetan pigs (49). In our study, the differential expression of protein kinase B among the four groups may have been mediated by ITGA5, which is also enriched in the PI3K-Akt signaling pathway and regulates a wide range of cellular activities, including cell proliferation (50), apoptosis (51), and metabolic progression (52). The antisense lncRNA MSTRG.36013.14 was complementary to the target gene AKT3, and its level was significantly higher in the TN and LL groups than that in the TL and LN groups; this lncRNA may regulate the phosphorylation of some vital downstream targets as the central mediator of the PI3K/Akt signaling pathway. Furthermore, the PI3K/Akt/mTOR (53) and VEGF/PI3K/Akt (54) signaling pathways regulate the expression of HIF-1α, which is closely associated with the concentration of oxygen in the environment and regulated by the PI3K/Akt signaling pathways as a downstream protein (55). The antioxidant function of enzymes could be activated following adaptive regulation to remove accumulated ROS in animals exposed to hypoxia as the main defense system (56). LDHA (the target genes of

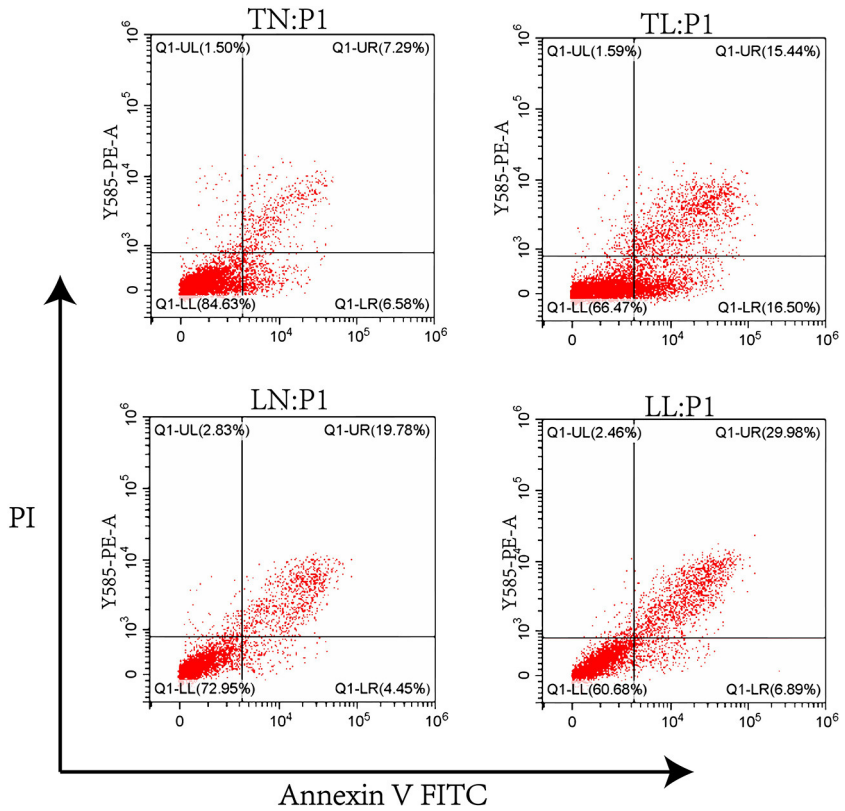


FIGURE 6 | The apoptosis of ATII cells from Tibetan pigs and Landrace pigs under normoxic (21% O_2) and hypoxic (2% O_2) condition was determined.

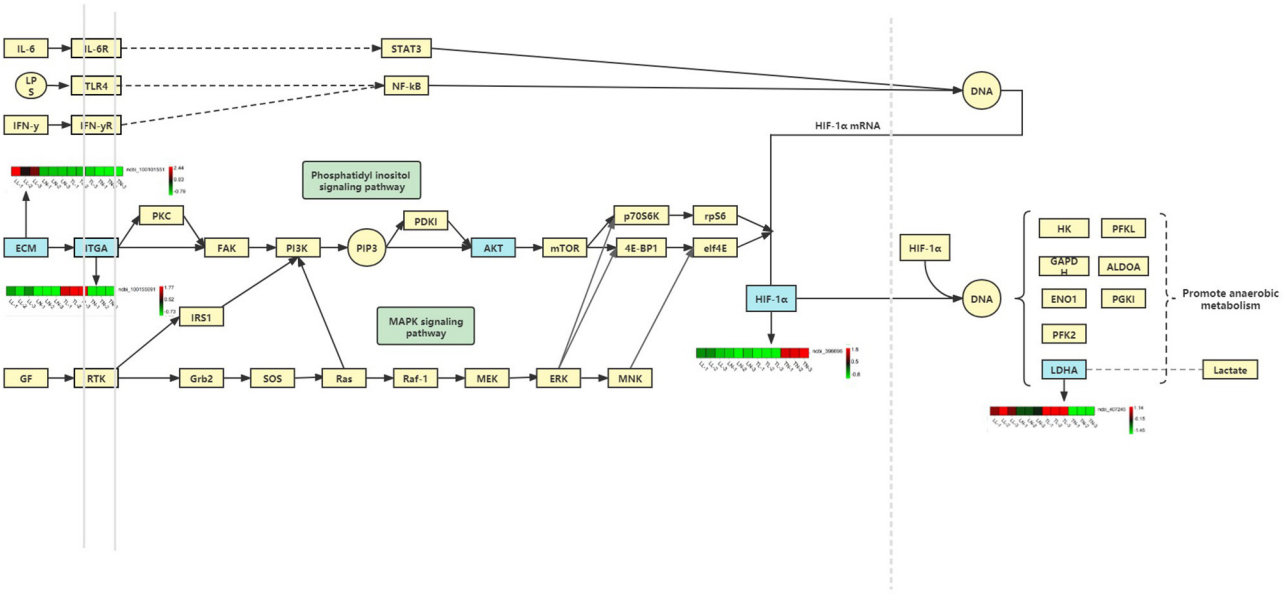


FIGURE 7 | Hypoxia induction may promote apoptosis by activating of the focal adhesion/PI3K-Akt/glycolysis pathway.

ssc-miR-429, ssc-miR-141, and miR-101-y) was expressed at a higher level in the hypoxia (TL and LL) groups than in the normoxia (TN and LN) groups and was significantly different in the TN and TL groups; this gene could be regulated by HIF-1α, and increased LDH expression may have promoted the accumulation of lactic acid in anaerobic glycolysis under the condition of hypoxia and further increased apoptosis, leading to numerous lung diseases in the hypoxic groups (TL and LL) (43, 57, 58). In our enrichment analysis, we found that exposure to hypoxia was mainly associated with the intracellular component GO term, and P4HA1, which could be induced by *HIF*-1*α* or hypoxia-independent factors, is a part of a gene expression signature associated with hypoxia and glycolysis in ATII cells (59, 60). *P4HA*1 and *P4HA*2 were the prolyl hydroxylase subtypes with significantly increased expression in cells under hypoxic conditions (61–63), which is similar to the results of our study, indicating that P4HA1 mediates hypoxia-induced invasion and migration (64).

lncRNA-miRNA-mRNA Networks in ATII Cells Under Hypoxic Conditions

In general, lncRNAs and miRNAs can maintain mRNA stability and regulate gene silencing and transcription. Cellular hypoxia responses involve the activation of complex signaling pathways, many of which lead to transcriptional cascades that are designed to alter cellular transcriptomes and proteomes to optimally combat hypoxia-induced damage. Many genes specifically increase related transcriptional activities and regulate a series of metabolic activities in the body during long-term adaptation to hypoxic environments. MiR-7-5p is a fascinating miRNA that plays diverse roles in tumor suppression; it inhibits the invasion abilities of tumor cells by directly targeting *PI3K*/*Akt*, *FAK,* and *KLF*4 expression and may be a useful therapeutic target for the diagnosis and treatment of patients with glioblastoma (65–67). miR-20b inhibits the expressions of autophagy-related proteins that are induced by hypoxia/reoxygenation injury via ULK1 (68). The central genes ssc-miR-7-5p, ssc-miR-20b, and novel-m0010-5p intersected at MSTRG.57127.1, MSTRG.36107.1, and LANCL3 between the hypoxia (TL and LL) and normoxia (TN and LN) groups, which may indicate that the ssc-miR-20b/MSTRG.57127.1/ssc-miR-7-5p axis plays a vital role in alleviating hypoxia-related injury by inhibiting the proliferation of tumor cells or regulating ATII cell autophagy (66). MSTRG.14861.4-miR-11971-z-CCDC12, the most affected axis, regulated numerous RNAs between the TN and TL groups, and thus may regulate ATII cell growth under hypoxic conditions (69). MSTRG.23871.1, miR-3178-z, and ssc-miR-30c-3p were identified as the most influential genes and regulated multiple RNAs. The ACTA1/ssc-miR-30c-3p/MSTRG.23871.1 axis is key for limiting ATII cell injury and improving dysfunction and fibrosis mediated by oxidative stress in Landrace pigs (70). The regulation of the expression levels of these factors by ceRNAs may improve the adaptation of Tibetan pigs vs. Landrace pigs to hypoxic environments.

CONCLUSION

In the present study, the ceRNA network of lncRNAs, miRNAs, and mRNAs from the ATII cells of Tibetan pigs and Landrace pigs under hypoxic environments was analyzed using whole-transcriptome techniques. Different regulatory ceRNA networks indicated that the MSTRG.14861.4/miR-11971-z/CCDC12 axis efficiently promotes ATII cell growth in Tibetan pigs under hypoxic conditions. In addition, oxidative stress may induce cell injury, increased dysfunction, and fibrosis in the ATII cells of Landrace pigs, but this could be relieved by the ACTA1/ssc-miR-30c-3p/MSTRG.23871.1 axis. These results may explain why Tibetan pigs adapted better to the plateau environment than Landrace pigs and will help to prevent hypoxia-induced injury in other mammals.

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 in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by 2006-398.

AUTHOR CONTRIBUTIONS

SZ was the overall project leader who provided financial support and experimental conception. YY was involved in data analyses, statistical analyses, language revisions, journal selection, and manuscript submissions and revisions. YL and HY contributed to the experimental design and implementation. CG contributed to the supervision and assistance of students in managing animals and collecting and analyzing samples. XL and YR were responsible for the trial implementation, supervision of students collecting and analyzing samples, and manuscript preparation. YC and TJ contributed to supervision of sample collection and analysis and manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (32060730, 31760644).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.834566/full#supplementary-material

REFERENCES

1. Ma YF, Han XM, Huang CP, Zhong L, Adeola AC, Irwin DM, et al. Population genomics analysis revealed origin and high-altitude adaptation of Tibetan pigs. *Sci Rep.* (2019) 9:11463. doi: 10.1038/s41598-019-47711-6
2. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet.* (2015) 47:217–25. doi: 10.1038/ng.3199
3. Wu DD, Yang CP, Wang MS, Dong KZ, Yan DW, Hao ZQ, et al. Convergent genomic signatures of high-altitude adaptation among domestic mammals. *Natl Sci Rev.* (2020) 7:952–63. doi: 10.1093/nsr/nwz213
4. Li J, Chen D, Yu B, He J, Huang Z, Mao X, et al. The fungal community and its interaction with the concentration of short-chain fatty acids in the faeces of Chenghua, Yorkshire and Tibetan pigs. *Microb Biotechnol.* (2020) 13:509–21. doi: 10.1111/1751-7915.13507
5. Cramer CL, Patterson A, Alchakaki A, Soubani AO. Immunomodulatory indications of azithromycin in respiratory disease: a concise review for the clinician. *Postgrad Med.* (2017) 129:493–9. doi: 10.1080/00325481.2017.1285677
6. Guillamat-Prats R, Puig F, Camprubí-Rimblas M, Herrero R, Serrano-Mollar A, GóMEZ MN, et al. Intratracheal instillation of alveolar type II cells enhances recovery from acute lung injury in rats. *J Heart Lung Transplant.* (2018) 37:782–91. doi: 10.1016/j.healun.2017.10.025
7. Sulkowska M. Morphological studies of the lungs in chronic hypobaric hypoxia. *Pol J Pathol.* (1997) 48:225–34.
8. Bernard O, Jeny F, Uzunhan Y, Dondi E, Terfous R, Label R, et al. Mesenchymal stem cells reduce hypoxia-induced apoptosis in alveolar epithelial cells by modulating HIF and ROS hypoxic signaling. *Am J Physiol Lung Cell Mol Physiol.* (2018) 314:L360–71. doi: 10.1152/ajplung.00153.2017
9. Guan R, Wang J, Li D, Li Z, Liu H, Ding M, et al. Hydrogen sulfide inhibits cigarette smoke-induced inflammation and injury in alveolar epithelial cells by suppressing PHD2/HIF-1α/MAPK signaling pathway. *Int Immunopharmacol.* (2020) 81:105979. doi: 10.1016/j.intimp.2019.105979
10. Sherman MA, Suresh MV, Dolgachev VA, McCandless LK, Xue X, Ziru L, et al. Molecular characterization of hypoxic alveolar epithelial cells after lung contusion indicates an important role for HIF-1α. *Ann Surg.* (2018) 267:382–91. doi: 10.1097/SLA.0000000000002070
11. Huang M, Yang L, Peng X, Wei S, Fan Q, Yang S, et al. Autonomous glucose metabolic reprogramming of tumour cells under hypoxia: opportunities for targeted therapy. *J Exp Clin Cancer Res.* (2020) 39:185. doi: 10.1186/s13046-020-01698-5
12. You B, Liu Y, Chen J, Huang X, Peng H, Liu Z, et al. Vascular peroxidase 1 mediates hypoxia-induced pulmonary artery smooth muscle cell proliferation, apoptosis resistance and migration. *Cardiovasc Res.* (2018) 114:188–99. doi: 10.1093/cvr/cvx234
13. Criscuoli M, Ulivieri C, Filippi I, Monaci S, Guerrini G, Crifò B, et al. The Shc protein Rai enhances T-cell survival under hypoxia. *J Cell Physiol.* (2020) 235:8058–70. doi: 10.1002/jcp.29461
14. Dulloo I, Hooi PB, Sabapathy K. Hypoxia-induced DNp73 stabilization regulates Vegf-a expression and tumor angiogenesis similar to TAp73. *Cell Cycle.* (2015) 14:3533–9. doi: 10.1080/15384101.2015.1078038
15. Zhao X, Tang DY, Zuo X, Zhang TD, Wang C. Identification of lncRNA-miRNA-mRNA regulatory network associated with epithelial ovarian cancer cisplatin-resistant. *J Cell Physiol.* (2019) 234:19886–94. doi: 10.1002/jcp.28587
16. Li W, Ma S, Bai X, Pan W, Ai L, Tan W. Long noncoding RNA WDFY3-AS2 suppresses tumor progression by acting as a competing endogenous RNA of microRNA-18a in ovarian cancer. *J Cell Physiol.* (2020) 235:1141–54. doi: 10.1002/jcp.29028
17. Lei GL, Niu Y, Cheng SJ, Li YY, Bai ZF, Yu LX, et al. Upregulation of long noncoding RNA W42 promotes tumor development by binding with DBN1 in hepatocellular carcinoma. *World J Gastroenterol.* (2021) 27:2586–602. doi: 10.3748/wjg.v27.i20.2586
18. Wang, Y., Wang, P., Zhang, Y., Xu, J., Li, Z., Li, Z., et al. (2020). Decreased expression of the host long-noncoding RNA-GM facilitates viral escape by inhibiting the kinase activity TBK1 via S-glutathionylation. *Immunity*. 53, 1168-1181.e7. doi: 10.1016/j.immuni.2020.11.010
19. Reza A, Choi YJ, Han SG, Song H, Park C, Hong K, et al. Roles of microRNAs in mammalian reproduction: from the commitment of germ cells to peri-implantation embryos. *Biol Rev Camb Philos Soc.* (2019) 94:415–38. doi: 10.1111/brv.12459
20. Shi HJ, Wang MW, Sun JT, Wang H, Li YF, Chen BR, et al. A novel long noncoding RNA FAF inhibits apoptosis via upregulating FGF9 through PI3K/AKT signaling pathway in ischemia-hypoxia cardiomyocytes. *J Cell Physiol.* (2019) 234:21973–87. doi: 10.1002/jcp.28760
21. Liang Y, Song X, Li Y, Chen B, Zhao W, Wang L, et al. LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. *Mol Cancer.* (2020) 19:85. doi: 10.1186/s12943-020-01206-5
22. Teng Y, Ding M, Wang X, Li H, Guo Q, Yan J, et al. LncRNA RMRP accelerates hypoxia-induced injury by targeting miR-214-5p in H9c2 cells. *J Pharmacol Sci.* (2020) 142:69–78. doi: 10.1016/j.jphs.2019.07.014
23. Yang Y, Yuan H, Yang T, Li Y, Gao C, Jiao T, et al. The expression regulatory network in the lung tissue of Tibetan pigs provides insight into hypoxia-sensitive pathways in high-altitude hypoxia. *Front Genet.* (2021) 12:691592. doi: 10.3389/fgene.2021.691592
24. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England).* (2018) 34:i884–90. doi: 10.1093/bioinformatics/bty560
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* (2012) 9:357–9. doi: 10.1038/nmeth.1923
26. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* (2015) 12:357–12, 360. doi: 10.1038/nmeth.3317
27. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290–5. doi: 10.1038/nbt.3122
28. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* (2016) 11:1650–67. doi: 10.1038/nprot.2016.095
29. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* (2013) 41:e166. doi: 10.1093/nar/gkt646
30. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* (2007) 35:W345–9. doi: 10.1093/nar/gkm391
31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* (2011) 12:323. doi: 10.1186/1471-2105-12-323
32. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res.* (2003) 31:439–41. doi: 10.1093/nar/gkg006
33. Griffiths-Jones S, Grocock RJ, Van DS, Bateman A, Enright AJ. MiRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* (2006) 34:D140–4. doi: 10.1093/nar/gkj112
34. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015) 43:D447–52. doi: 10.1093/nar/gku1003
35. Xin J, Zhang H, He Y, Duren Z, Bai C, Chen L, et al. Chromatin accessibility landscape and regulatory network of high-altitude hypoxia adaptation. *Nat Commun.* (2020) 11:4928. doi: 10.1038/s41467-020-18638-8
36. Ge Q, Guo Y, Zheng W, Cai Y, Qi X, Zhao S. A comparative analysis of differentially expressed mRNAs, miRNAs and circRNAs provides insights into the key genes involved in the high-altitude adaptation of yaks. *BMC Genomics.* (2021) 22:744. doi: 10.1186/s12864-021-08044-9
37. Yang Y, Gao C, Yang T, Sha Y, Cai Y, Wang X, et al. Vascular characteristics and expression of hypoxia genes in Tibetan pigs' hearts. *Vet Med Sci.* (2022) 8:177–86. doi: 10.1002/vms3.639
38. Yang Y, Gao C, Yang T, Sha Y, Cai Y, Wang X, et al. Characteristics of Tibetan pig lung tissue in response to a hypoxic environment on the Qinghai-Tibet Plateau. *Arch Anim Breed.* (2021) 64:283–92. doi: 10.5194/aab-64-283-2021
39. Lu T, Zheng Y, Yang H, Wu B, Xiong J, Huang C, et al. Structural characteristics of the deciduous teeth of Tibetan miniature pigs. *Nan Fang Yi Ke Da Xue Xue Bao.* (2019) 39:1113–7. doi: 10.12122/j.issn.1673-4254.2019.09.18
40. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* (2013) 45:1431–8. doi: 10.1038/ng.2811

41. Fan JL, Zhu TT, Xue ZY, Ren WQ, Guo JQ, Zhao HY, et al. LncRNA-XIST protects the hypoxia-induced cardiomyocyte injury through regulating the miR-125b-hexokianse 2 axis. *In Vitro Cell Dev Biol Anim.* (2020) 56:349–57. doi: 10.1007/s11626-020-00459-0
42. McClendon J, Jansing NL, Redente EF, Gandjeva A, Ito Y, Colgan SP, et al. Hypoxia-inducible factor 1α signaling promotes repair of the alveolar epithelium after acute lung injury. *Am J Pathol.* (2017) 187:1772–86. doi: 10.1016/j.ajpath.2017.04.012
43. Lu Q, Rounds S. Focal adhesion kinase and endothelial cell apoptosis. *Microvasc Res.* (2012) 83:56–63. doi: 10.1016/j.mvr.2011.05.003
44. Lin N, Shay JES, Xie H, Lee DSM, Skuli N, Tang Q, et al. Myeloid cell hypoxia-inducible factors promote resolution of inflammation in experimental colitis. *Front Immunol.* (2018) 9:2565. doi: 10.3389/fimmu.2018.02565
45. Saad A, Zhu XY, Herrmann S, Hickson L, Tang H, Dietz AB, et al. Adipose-derived mesenchymal stem cells from patients with atherosclerotic renovascular disease have increased DNA damage and reduced angiogenesis that can be modified by hypoxia. *Stem Cell Res Ther.* (2016) 7:128. doi: 10.1186/s13287-016-0389-x
46. Choi HJ, Sanders TA, Tormos KV, Ameri K, Tsai JD, Park AM, et al. ECM-dependent HIF induction directs trophoblast stem cell fate via LIMK1-mediated cytoskeletal rearrangement. *PLoS ONE.* (2013) 8:e56949. doi: 10.1371/journal.pone.0056949
47. Campbell ID, Humperies MJ. Integrin structure, activation, and interactions. *Cold Spring Harb Perspect Biol.* (2011) 3:a004994. doi: 10.1101/cshperspect.a004994
48. Zhao X, Guan JL. Focal adhesion kinase and its signaling pathways in cell migration and angiogenesis. *Adv Drug Deliv Rev.* (2011) 63:610–5. doi: 10.1016/j.addr.2010.11.001
49. Cui S, Yang CL, Chen DY. LncRNA EWSAT1 regulates the tumorigenesis of NSCLC as a ceRNA by modulating miR-330-5p/ITGA5 axis. *Biochem Genet.* (2021) 59:1441–56. doi: 10.1007/s10528-021-10069-4
50. Deng Y, Wan Q, Yan W. Integrin α5/ITGA5 promotes the proliferation, migration, invasion and progression of oral squamous carcinoma by epithelial-mesenchymal transition. *Cancer Manag Res.* (2019) 11:9609–20. doi: 10.2147/CMAR.S223201
51. Yu M, Chu S, Fei B, Fang X, Liu Z. O-GlcNAcylation of ITGA5 facilitates the occurrence and development of colorectal cancer. *Exp Cell Res.* (2019) 382:111464. doi: 10.1016/j.yexcr.2019.06.009
52. Li L, Qu Y, Mao M, Xiong Y, Mu D. The involvement of phosphoinositid 3-kinase/Akt pathway in the activation of hypoxia-inducible factor-1alpha in the developing rat brain after hypoxia-ischemia. *Brain Res.* (2008) 1197:152–8. doi: 10.1016/j.brainres.2007.12.059
53. Yang X, Gao M, Miao M, Jiang C, Zhang D, Yin Z, et al. Combining combretastatin A4 phosphate with ginsenoside Rd synergistically inhibited hepatocellular carcinoma by reducing HIF-1α via PI3K/AKT/mTOR signalling pathway. *J Pharm Pharmacol.* (2021) 73:263–71. doi: 10.1093/jpp/rgaa006
54. Wang D, Zhao W, Liu J, Wang Y, Yuan C, Zhang F, et al. Effects of HIF-1α on spermatogenesis of varicocele rats by regulating VEGF/PI3K/Akt signaling pathway. *Reprod Sci.* (2021) 28:1161–74. doi: 10.1007/s43032-020-00395-0
55. Gao T, Zhang X, Zhao J, Zhou F, Wang Y, Zhao Z, et al. SIK2 promotes reprogramming of glucose metabolism through PI3K/AKT/HIF-1α pathway and Drp1-mediated mitochondrial fission in ovarian cancer. *Cancer Lett.* (2020) 469:89–101. doi: 10.1016/j.canlet.2019.10.029
56. Ghosh S, Kamble NU, Verma P, Salvi P, Petla BP, Roy S, et al. Arabidopsis protein l-ISOASPARTYL METHYLTRANSFERASE repairs isoaspartyl damage to antioxidant enzymes and increases heat and oxidative stress tolerance. *J Biol Chem.* (2020) 295:783–99. doi: 10.1074/jbc.RA119.010779
57. Zhao LL, Wu H, Sun JL, Liao L, Cui C, Liu Q, et al. MicroRNA-124 regulates lactate transportation in the muscle of largemouth bass (micropterus salmoides) under hypoxia by targeting MCT1. *Aquat Toxicol.* (2020) 218:105359. doi: 10.1016/j.aquatox.2019.105359
58. Hu D, Linders A, Yamak A, Correia C, Kijlstra JD, Garakani A, et al. Metabolic maturation of human pluripotent stem cell-derived cardiomyocytes by inhibition of HIF1α and LDHA. *Circ Res.* (2018) 123:1066–79. doi: 10.1161/CIRCRESAHA.118.313249
59. Gilkes DM, Bajpai S, Chaturvedi P, Wirtz D, Semenza GL. Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *J Biol Chem.* (2013) 288:10819–29. doi: 10.1074/jbc.M112.442939
60. Chen L, Shen YH, Wang X, Wang J, Gan Y, Chen N, et al. Human prolyl-4-hydroxylase alpha (I) transcription is mediated by upstream stimulatory factors. *J Biol Chem.* (2006) 281:10849–55. doi: 10.1074/jbc.M511237200 Epub 2006 Feb 17
61. Morimoto C, Takedachi M, Kawasaki K, Shimomura J, Murata M, Hirai A, et al. Hypoxia stimulates collagen hydroxylation in gingival fibroblasts and periodontal ligament cells. *J Periodontol.* (2021) 92:1635–45. doi: 10.1002/JPER.20-0670
62. Trackman PC. Diverse biological functions of extracellular collagen processing enzymes. *J Cell Biochem.* (2005) 96:927–37. doi: 10.1002/jcb.20605
63. Hofbauer KH, Gess B, Lohaus C, Meyer HE, Katschinski D, Kurt Z, et al. Oxygen tension regulates the expression of a group of procollagen hydroxylases. *Eur J Biochem.* (2003) 270:4515–22. doi: 10.1046/j.1432-1033.2003.03846.x
64. Zhu X, Liu S, Yang X, Wang W, Shao W, Ji T. P4HA1 as an unfavorable prognostic marker promotes cell migration and invasion of glioblastoma via inducing EMT process under hypoxia microenvironment. *Am J Cancer Res.* (2021) 11:590–617.
65. Yin CY, Kong W, Jiang J, Xu H, Zhao W. miR-7-5p inhibits cell migration and invasion in glioblastoma through targeting SATB1. *Oncol Lett.* (2019) 17:1819–25. doi: 10.3892/ol.2018.9777
66. Xiao H. MiR-7-5p suppresses tumor metastasis of non-small cell lung cancer by targeting NOVA2. *Cell Mol Biol Lett.* (2019) 24:60. doi: 10.1186/s11658-019-0188-3
67. Tian S, Chen M, Wang B, Han Y, Shang H, Chen J. MiR-7-5p promotes hepatic stellate cell activation by targeting fibroblast growth factor receptor 4. *Gastroenterol Res Pract.* (2020) 2020:5346573. doi: 10.1155/2020/5346573
68. Lu Y, Wang S, Cai S, Gu X, Wang J, Yang Y, et al. Propofol-induced miR-20b expression initiates endogenous cellular signal changes mitigating hypoxia/re-oxygenation-induced endothelial autophagy *in vitro*. *Cell Death Dis.* (2020) 11:681. doi: 10.1038/s41419-020-02828-9
69. Fan C, Dong L, Zhu N, Xiong Y, Zhang J, Wang L, et al. Isolation of siRNA target by biotinylated siRNA reveals that human CCDC12 promotes early erythroid differentiation. *Leuk Res.* (2012) 36:779–83. doi: 10.1016/j.leukres.2011.12.017
70. Ning BB, Zhang Y, Wu DD, Cui JG, Liu L, Wang PW, et al. Luteolin-7-diglucuronide attenuates isoproterenol-induced myocardial injury and fibrosis in mice. *Acta Pharmacol Sin.* (2017) 38:331–41. doi: 10.1038/aps.2016.142

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: 11 March 2022
doi: 10.3389/fgene.2022.798076

Resequencing and Signatures of Selective Scans Point to Candidate Genetic Variants for Hair Length Traits in Long-Haired and Normal-Haired Tianzhu White Yak

Qi Bao[1,2†], Xiaoming Ma[1,2†], Congjun Jia[3], Xiaoyun Wu[1,2], Yi Wu[1], Guangyao Meng[1,2], Pengjia Bao[1,2], Min Chu[1,2], Xian Guo[1,2], Chunnian Liang[1,2] and Ping Yan[1,2*]*

[1]Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, China, [2]Key Laboratory of Yak Breeding Engineering, Lanzhou, China, [3]Guangdong Meizhou Vocational and Technical College, Meizhou, China

OPEN ACCESS

Edited by:
Klaus Wimmers,
Leibniz Institute for Farm Animal Biology (FBN), Germany

Reviewed by:
Fenghua Lyu,
China Agricultural University, China
Xiangdong Ding,
China Agricultural University, China

***Correspondence:**
Chunnian Liang
Chunnian2006@163.com
Ping Yan
pingyanlz@163.com

[†]These authors have contributed equally to this work.

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

Received: 30 October 2021
Accepted: 17 February 2022
Published: 11 March 2022

Citation:
Bao Q, Ma X, Jia C, Wu X, Wu Y, Meng G, Bao P, Chu M, Guo X, Liang C and Yan P (2022) Resequencing and Signatures of Selective Scans Point to Candidate Genetic Variants for Hair Length Traits in Long-Haired and Normal-Haired Tianzhu White Yak. Front. Genet. 13:798076. doi: 10.3389/fgene.2022.798076

Tianzhu white yak is a rare local yak breed with a pure white coat in China. In recent years, breeders have discovered long-haired individuals characterized by long hair on the forehead in the Tianzhu white yak, and the length and density of the hair on these two parts of the body are higher than that of the normal Tianzhu white yak. To elucidate the genetic mechanism of hair length in Tianzhu white yak, we re-sequence the whole genome of long-haired Tianzhu White yak (LTWY) (n = 10) and normal Tianzhu White yak (NTWY) (n = 10). Then, fixation index (F_{ST}), θπ ratio, cross-population composite likelihood ratio (XP-CLR), integrated haplotype score (iHS), cross-population extended haplotype homozygosity (XP-EHH), and one composite method, the de-correlated composite of multiple signals (DCMS) were performed to discover the loci and genes related to long-haired traits. Based on five single methods, we found two hotspots of 0.2 and 1.1 MB in length on chromosome 6, annotating two (*FGF5*, *CFAP299*) and four genes (*ATP8A1*, *SLC30A9*, *SHISA3*, *TMEM33*), respectively. Function enrichment analysis of genes in two hotspots revealed Ras signaling pathway, MAPK signaling pathway, PI3K-Akt signaling pathway, and Rap1 signaling pathway were involved in the process of hair length differences. Besides, the DCMS method further found that four genes (*ACOXL*, *PDPK1*, *MAGEL2*, *CDH1*) were associated with hair follicle development. Henceforth, our work provides novel genetic insights into the mechanisms of hair growth in the LTWY.

Keywords: yak, long-haired trait, DCMS, selection signal, resequencing

INTRODUCTION

In taxonomy, yak (*Bos grunniens*) is a member of the *Artiodactyla*, family Bovidae, genus *Bos*, which is endemic to the alpine region of the Qinghai-Tibet Plateau (Qiu et al., 2015). Yak adapts to the cold climate and is distributed in China's Qinghai-Tibet Plateau with an altitude of more than 3,000 m. The yak has been well known for its reputation as "boat on the plateau", providing an indispensable transportation source for the production and life of local herdsmen (Qiu et al., 2015). Compared with cattle at a lower altitude, yak has long, thick skirt hair on the chest, legs, and flanks, forming a natural

thermal insulation layer. Yak has more villi on their side, shoulder, and back, and the content of abdominal coarse hair (group hair) is the highest (Danzan et al., 2014).

Among all the 16 million yaks in the world, white individuals are rare, and the coat color is genetically unstable. Generally, in the domestic yak, most of the individuals' coat colors are black, brown, black-brown, or with a small number of white patches, and only about 2–3% of the individuals are white (Wiener et al., 2003). The Tianzhu white yak population, however, is relatively large, with stable genetic properties, and is a unique local group breed in Tianzhu, Gansu province of China. As the special case of directional breeding for pure breeding in yak and the iconic white coat, Tianzhu white yak is a precious local yak group. In recent years, breeders have discovered a subgroup of Tianzhu white yak, which is characterized by the long hair on the forehead, and the length and density of the hair on the side of the body are higher than that of the normal Tianzhu white yak. We conducted statistical analysis on the production performance record and fur quality record data of Tianzhu White Yak in Tianzhu White Yak Breeding Base in Tianzhu County and concluded that Tianzhu White Yak can be divided into normal-haired type (≤13 cm) and long-haired type (>13 cm). Due to the economic benefit and landscape use of hair of Tianzhu white yak, breeders hope to breed stable offspring of this subgroup of Tianzhu white yak.

The selection signature of the genome includes the free-riding effect and selective clearance. The free-riding effect refers to that when a favorable mutation site with high fitness is fixed quickly, the polymorphism of the gene sequences linked to this site change accordingly (Smith and Haigh, 1974; Fay and Wu, 2000). Selective clearance is the phenomenon that the polymorphism of the chromosomal region linked closely around the site is reduced due to the free-riding effect (Smith and Haigh, 1974). And the selection in genetics often leads to corresponding changes in biological traits. These selected genes determine the traits of the organism. Therefore, it is equivalent to finding candidate genes that perform corresponding functions when the selection signals are identified. This is of great significance for understanding the evolutionary process of species and finding genes controlling traits with economic importance. Methods of the selection signal detection mainly include three categories, including population differentiation-based methods: fixation index (F_{ST}) test (Pearse and Crandall, 2004), locus-specific branch lengths (LSBL) test (Shriver et al., 2004), and di test (Akey et al., 2010); allele frequency spectrum-based methods: Tajima's D test (Tajima, 1989) and Hp test (Fay and Wu, 2000), etc.; haplotype-based methods: cross-population extended haplotype homozygosity (XP-EHH) test (Sabeti et al., 2007), extended haplotype homozygosity (EHH) test (Sabeti et al., 2002) and integrated haplotype score (iHS) test (Voight et al., 2006). In addition, the HapFLK method based on the hierarchical structure of the sample population and the cross-population composite likelihood ratio (XP-CLR) method based on the difference in multilocus allele frequency between two populations are always used in the selection signal detection (Hua et al., 2010; Fariello et al., 2013).

Multiple methods can be used to detect selection signals, and each method has its limitations. Results obtained from algorithms based on the genetic differentiation are interfered by the population history, and methods based on the unit points are affected by the linkage factors (Lewontin and Krakauer, 1973; Tajima, 1989; Shriver et al., 2004). In addition, methods based on the linkage-disequilibrium can only judge recent selection signals (Sabeti et al., 2002). Compared to the single-statistic tests or other meta-analyses, more recent studies manifested that composite measures of multiple signals selection have higher efficiency and present a reliable positional resolution (Grossman et al., 2010; Lotterhos et al., 2017). In our study, the de-correlated composite of multiple signals (DCMS), one of the composite analysis strategies was performed here (Ma et al., 2015). The DCMS method can combine *p*-values from different selection signal statistics into a single DCMS framework and correct for the overall correlation between the statistics based on the covariance matrix (Ma et al., 2015).

The purpose of this study is to identify the imprints left on the genome of LTWY and NTWY populations during the process of natural and artificial selection and to identify genes involved in the determination of hair length. To solve these problems, the genome-wide haplotype data of the long-haired Tianzhu white yak (LTWY) and the normal Tianzhu white yak (NTWY) were used, and five single methods (F_{ST}, θπ ratio, XP-CLR, iHs, XP-EHH) and one composite method (DCMS) were conducted to detect the population selection signal and dig out the sites or candidate genes related to hair length where selection occurs. Our work provides an important reference for the selection and improvement of long-haired yak breeding.

MATERIALS AND METHODS

Sample Collection and Sequencing

All blood samples of the LTWY and NTWY were collected from the Tianzhu white yak farmed in Gansu province, China. For each population, genomic DNA was extracted from blood samples using the EasyPure Blood Genomic DNA Kit (TransGen Biotech, Beijing, China) according to the manufacturer's instructions. The quality and integrity of the extracted DNA were checked by measuring the A260/A280 ratio and screening by agarose gel electrophoresis. Qualified genomic DNA samples were randomly broken into fragments with a Covaris instrument. The interrupted samples were selected and concentrated around 200–300 bp using the Agencourt AMPure XP-Medium kit. The end-repair was performed on the fragmented DNA, the base A was added to the 3′ end to connect the sequencing adapter, and subsequent PCR amplification was performed on the ligated product. Then the PCR product was denatured to single-stranded, and the circularization reaction system was prepared, samples were mixed thoroughly and reacted at a suitable temperature for a certain time to obtain a single-stranded circular product. After digesting the linear DNA

molecules that have not been circularized, the final libraries were obtained. The Agilent 2,100 Bioanalyzer (Agilent DNA 1000 Reagents) was used to detect the fragment size and concentration of the libraries, and then the qualified libraries were sequenced on the BGISEQ-500 platform. The raw image data obtained by sequencing were converted into raw reads by the BGISEQ-500 Base Calling software. The data were stored in the FASTQ file format.

Reads Mapping and Single-Nucleotide Polymorphisms (SNPs) Calling

After removing adaptor sequences, contamination, and low-quality reads, high-quality reads were aligned to the latest *Bos grunniens* reference genome (accession number: GCA_005887515.1) using BWA-MEM (0.7.10-r789) with default parameters (Li and Durbin, 2009). The SAM files were sorted and converted to binary format (BAM, Binary sequence Alignment Map) to relieve computer memory and storage pressure via using SAMtools (version 1.9) (Li et al., 2009). The Genome Analysis ToolKit (GATK) (v4.1.8.0) was used to call variants. And alignments were marked for PCR duplicates using MarkDuplicates module of GATK following the BAM construction. For all the BAM files obtained, variants were called with HaplotypeCaller module. The g.vcf files were combined with the GenotypeGVCFs module of GATK. Finally, the original SNP files were obtained by using SelectVariants module (McKenna et al., 2010). With the VariantFiltration parameter, the filter conditions were set as "QUAL <30.0, QualByDepth (QD) < 2.0, Fisher`s exact test (FS) > 60.0, RMS Mapping Quality (MQ) < 40.0, HaplotypeScore >13.0". After filtering, a VCF file containing high-quality SNPs was obtained. The command *samtools flagstat* was used to discover the statistic information of each sample, including average coverage, count of raw reads, mapped reads, and properly paired reads. The sequencing depth of each sample was analyzed using the VCFtools software. Considering that the low-quality genotype data may affect the subsequent analysis, samples were removed when individual call rate was <0.95 of all SNPs and SNPs with low call rate (geno<0.99), SNPs with low minor allele frequencies (MAFs) (MAF<0.05), SNPs without chromosomal assignments, and SNPs on sex chromosomes were excluded (Yurchenko et al., 2019). The parameter of PLINK was set as follow: -maf 0.05, --mind 0.05, --geno 0.01, and --chr 1–29.

Principal Component Analysis (PCA)

Based on SNP information, PLINK was performed to the PCA to determine the genetic structure between populations. The visualization of PCA was based on the R package ggplot2.

Selective Scans of Five Single Methods

Considering that method based on unit point SNP scanning is susceptible to factors such as genetic drift, therefore, a sliding window calculation strategy was selected here to raise the sensitivity of the selected signal and reduce false positives (Ma et al., 2015). The F_{ST} values were calculated using VCFtools software (v1.1.0) (Danecek et al., 2011) with parameter: -fst-window-size 50,000. Negative values of F_{ST} were converted into zeros. Nucleotide diversity (π) is the ratio of polymorphic sites in two randomly selected nucleotide sequences, which is evaluated on the difference between the sequences and the relative frequency (Tajima, 1983). In this study, the π values were also calculated by VCFtools software, and the parameter was set as -window-pi 50,000. The $\theta\pi$ ratio was calculated as π_L/π_N, where π_L and π_N were the nucleotide diversity values for the LTWY and NTWY, respectively. The values of the XP-CLR were calculated using the python script XPCLR, which was downloaded from GitHub (https://github.com/hardingnj/xpclr). The corresponding parameters were set as: maximum of SNPs 600, ld value 0.95, window size 50,000. The integrated haplotype score (iHS) was used to calculate values of a window of SNPs (Voight et al., 2006) through the R package rehh (Gautier and Vitalis, 2012). The software BEDtools was used to obtain the 50 kb window coordinate file (Quinlan and Hall, 2010). And the in-house python script was used to average absolute values of iHS into non-overlapping sliding windows of 50 kb. Lastly, the rehh package was also used for XP-EHH calculation (Sabeti et al., 2007). The NTWY population was selected as the reference population. And then, the same script was used to average the XP-EHH scores into non-overlapping sliding windows of 50 kb. To shorten computing time, the rehh package was performed in a parallel mode and the R package Parallel (Vera et al., 2008).

De-Correlated Composite of Multiple Signals (DCMS)

After performing the statistics of five selection signal methods (F_{ST}, $\theta\pi$ ratio, XP-CLR, iHS, XP-EHH), all the values were combined into a matrix based on the window name. The DCMS values were calculated using the MINOTAUR package (Verity et al., 2017). Firstly, the results of five statistics were converted to *p*-values using the function *stat_to_pvalue*. Each column in the input data frame was converted to fractional ranks between 0 and 1. These values were then transformed to rank-based *p*-values based on the one-tailed test (iHS- left-tailed; $\theta\pi$ ratio, XP-EHH, XP-CLR, and F_{ST}–right-tailed). Final values were then transformed again to occupy the range 0–1 exclusive. Then, to obtain the correlations among these statistics, the covariance matrix was calculated using the *Cov-NAMcd* function with the parameters: alpha = 0.75, nsamp = 300,000. Combined with the matrix obtained in the last step, the *DCMS* function was used to calculate the DCMS scores. Robust estimations of the mean and variance of the DCMS scores were obtained using the R MASS package *rlm* function to eliminate the influence of isolated values (Boitard et al., 2016). And then the fitted DCMS scores were converted into *p*-values using the *pnorm* function. Finally, to control for false discovery rate, the R package *q*-value was used to adjust *p*-values using multiple hypothesis testing (Storey and Tibshirani, 2003). The adjusted *p*-values (*q*-values) were

TABLE 1 | Summary statistics of NTWY and LTWY re-sequenced reads.

Sample name	Number	Raw reads	Mapped reads	Properly paired reads	Average coverage	Average fold
NTWY	10	1,544,406,470	1,518,525,222	1,428,802,730	98.33%	7.58
LTWY	10	1,520,934,131	1,495,623,348	1,409,965,496	98.33%	7.38
Total	20	3,065,340,601	3,014,148,570	2,838,768,226	98.33%	7.48

TABLE 2 | Functional annotation of the identified single-nucleotide polymorphisms (SNPs) in NTWY and LTWY.

Fields		NTWY	LTWY	Total
Sample counts		10	10	20
SNP count		15,331,905	15,124,083	16,708,655
Ts/Tv ratio		2.497	2.496	***
Hom/Het		0.61	0.63	***
SNP types				
Exon	Synonymous variant	128,679	132,764	152,356
	Initiator codon variant	28	23	16
	Start lost	243	271	208
	start_retained_variant	2	2	39
	Stop gained	1,840	1,780	1,857
	Stop lost	227	235	197
	Stop retained variant	130	123	190
Splice site	Splice region variant	25,046	25,556	26,836
	Splice acceptor variant	546	546	429
	Splice donor variant	724	725	629
Intron	Intron variant	13,948,749	13,869,588	15,421,077
	Intragenic variant	568	632	607
UTR	5 prime UTR variant	24,657	25,413	23,110
	5 prime UTR premature start codon gain variant	3,687	3,836	3,914
	3 prime UTR variant	55,993	57,992	60,922
Intergenic	Upstream gene variant	1,174,257	1,194,925	1,277,565
	Downstream gene variant	1,190,326	1,209,439	1,296,616
	Intergenic region	9,907,119	9,727,032	10,736,862
Functional classes	Missense	106,774	107,373	209,965
	Nonsense	1,840	1,780	3,746
	Silent	128,813	132,891	252,509

visualized by *manhattan* function of R package qqman (Turner, 2018).

Variant Functional Annotations

Genes annotated in the BosGru3.0 genome version included in a selected interval were extracted using SnpEff (v4.5) software (Cingolani, 2012). The values of five selective signal methods in the top 1% of the empirical distribution (F_{ST}>0.119, θπ ratio>2.558, $|iHS|_{LTWY}$>1.532, XP-CLR>18.531, XP-EHH>2.203) were designated as candidate selection scans and genes that in those window region were defined as potential candidate genes. Next, to identify selection regions under the DCMS method of both populations, all the intervals with SNPs expressing decorrelated composite of multiple signal *q*-values less than 0.05 were obtained. BEDtools was used to extract the annotation file of these strong selection signal intervals. To get the meaningful mutations, the intron region and synonymous mutation sites were removed. The genes corresponding to the remaining sites were defined as potential candidate genes. The overlapping genes identified by DCMS were visualized using the circos package (Krzywinski et al., 2009). GO and KEGG enrichment was employed by KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/index.php).

RESULTS

Sequencing and Variation Calling

In this study, a total of 20 samples from LTWY and NTWY were re-sequenced, and an average of 7.48× coverage was generated. High-quality reads were aligned using the LU_Bosgru_v3.0 reference genome through BWA MEM algorithm. Statistical results showed that a total of 3,014,148,570 reads were obtained, covering 98.33% of the reference sequences across the region (**Table 1**). The SnpEff software was used to evaluate the genomic polymorphism of the LTWY and NTWY populations (**Table 2**). Our results showed that a total of 15,124,083 SNPs were identified in the LTWY population, with an average of one mutation site per 169 bases on the chromosome. And a total of 15,331,905 SNPs were identified, with an average of one mutation site per 166 bases in the NTWY population. For downstream selective signal analysis, the g.vcf

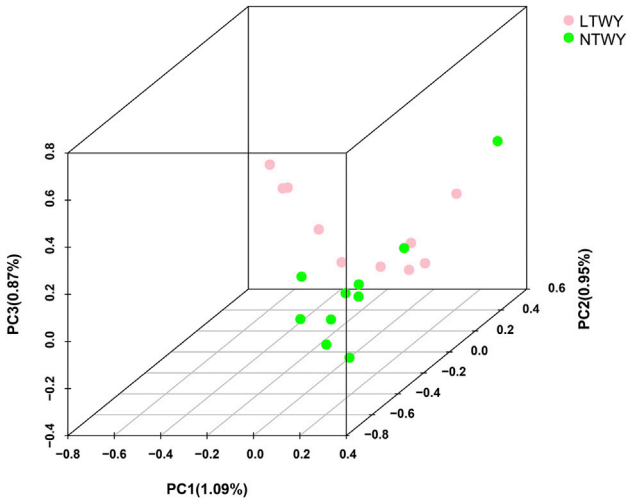


FIGURE 1 | Principal component analysis of NTWY and LTWY populations.

files of both populations were combined to identify SNPs, and finally, 16,708,655 SNP sites were obtained. In addition, the distribution region of SNPs in the LTWY and the NWTY was also analyzed (**Table 2**). Our results showed that the SNP variations in the LTWY and NTWY populations mainly occurred in the genetic interval (Intergenic), followed by the downstream interval (Downstream), upstream interval (Upstream), intron interval (Intron), exon interval (Exon) and so on (**Table 2**). And the Ts/Tv ratio which can be evaluated for the quality of the SNP call were 2.496 and 2.497 in the LTWY and NTWY, respectively. Principal component analysis results showed that two populations could be distinguished according to the three principal components. Three components captured 1.09, 0.95, and 0.87% of the total eigenvalue, respectively (**Figure 1**).

Analysis Results of Single Selection Signal Method

Based on different principles, five selection signal methods were used to screen selected regions and candidate genes. By two haplotype-based selection methods (iHS, XP-EHH), a hotspot (chr6: 25,200,001-25,400,000) with a length of 0.2-MB was detected (**Figure 2A**). Part of the segment of two genes (*FGF5*, *CFAP299*) was located in this region (**Figure 2B**). One missense mutation (c.302G > C, p.Ser101Thr) sites with large allele frequency differences were identified in *FGF5* (**Table 3**). Enrichment analysis results showed that two genes in the 0.2-MB hotspot were involved in eight GO items and six KEGG pathways, including signal transduction involved in the regulation of gene expression, fibroblast growth factor receptor binding, growth factor activity, positive regulation of cell population proliferation. The KEGG pathway includes MAPK signaling pathway, PI3K-Akt signaling pathway, Rap1 signaling pathway, Ras signaling pathway, melanoma, regulation of actin cytoskeleton (**Figure 2C**). Among the selected areas identified by the five methods, we found another hotspot (chr6: 61,650,001-62,750,000) with a length of 1.1-MB (**Figure 2A**). Four genes (*ATP8A1*, *SHISA3*, *SLC30A9*, *TMEM33*) were annotated in this hotspot (**Figure 2B**). In *SHISA3*, we identified one missense mutation (c.199G > A, p.Ala67Thr) sites (**Table 3**). Enrichment analysis results showed that four genes in 1.1-MB hotspot were enriched in 18 GO items, including magnesium ion binding, trans-Golgi network, cation transmembrane transport, membrane organization, phospholipid transport, etc (**Figure 2D**). There was no significant KEGG pathway enrichment in this segment.

Analysis of the Multi-Signal De-correlation Composite

The five selection signal statistics were combined into a single DCMS framework using the MINOTAUR. 400 and 420 genomic intervals under putative selection in LTWY and NTWY genomes were obtained after fitting for normal distribution, calculation of *p*-values, and correction for multiple testing (q-value < 0.05) (**Figure 3A**). According to the DCMS method, we screened the loci in the overlapping region for annotation. A total of 254 intervals were obtained, and 71 genes were annotated (**Figure 3B**). GO enrichment analysis resulted in 34 significantly enriched pathways (q-value < 0.05). Hair follicle development, including positive regulation of hair follicle development, positive regulation of cytokine-mediated signaling pathway, mitotic cell cycle, positive regulation of apoptotic process, negative regulation of transforming growth factor beta receptor signaling pathway, and negative regulation of cell-substrate adhesion were involved as major enrichment pathway, which may play an important role in the hair growth of LTWY. In addition, three significant pathways (platinum drug resistance, aldosterone-regulated sodium reabsorption, ECM-receptor interaction) were enriched in these genes (**Figure 3C**).

The DCMS method also identified strong signals detected near the 1.1-MB hotspot, further verifying the feasibility of this method. Further screening identified some highly significant genes related to hair follicle development. Among all the overlapping genes, we noticed that both LTWY and NTWY populations had a common signature of selection near the *ACOXL*, *PDPK1*, *MAGEL2*, and *CDH1* (**Figure 3B**). *ACOXL* encoded the rate-limiting enzyme of the fatty acid β-oxidation pathway (Schrader and Fahimi, 2006). One missense mutation (c.958G > A, p.Asp320Asn) site was found in this gene. In the 50 kb window (20:10300001-10350000), one missense mutation site (c.1360G > T, p.Val454Leu) of *CDH1* was identified (**Table 3**). *CDH1* played an important role in maintaining the adhesive properties and proper skin differentiation in keratinocytes (Hodivala and Watt, 1994). Two missense mutations (c.2274C > A, p.His758Gin; c.325A > G, p.Met109Val) were identified in *MAGEL2* in the 50 kb window (17: 76450001-76500000) (**Table 3**). *MAGEL2* was primarily expressed in the paraventricular nucleus, supraoptic nucleus, and in the suprachiasmatic nucleus (SCN) of the hypothalamus, which played role in circadian rhythm (Panda

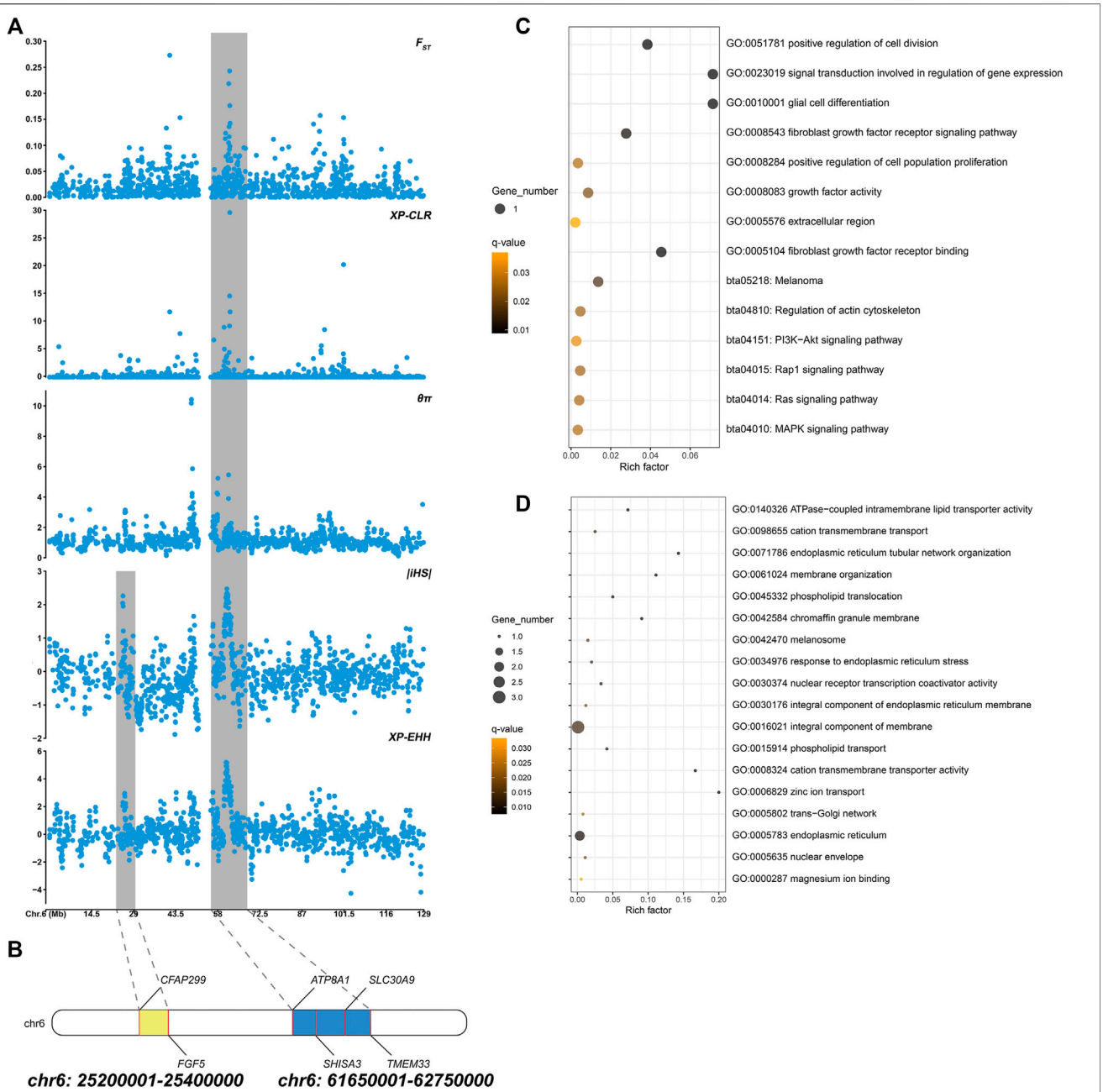


FIGURE 2 | (A) Manhattan plot of five selection signals on chromosome 6. The two hotspots are marked with shading. **(B)** Genes identified in two hotspots. **(C)** Functional enrichment analysis of 0.2-MB hotspot. **(D)** Functional enrichment analysis of 1.1-MB hotspot.

et al., 2002). *PDPK1* was associated with glucose metabolism (Beg et al., 2017), and one missense mutation (c.1579T > C, p.er527Pro) was found (**Table 3**).

DISCUSSION

The resequencing data from LTWK and NTWK was used to reveal the potential selective sweeps. And detailed genomic information along with candidate genes associated with the phenotypic change in the long-haired population was identified here. Our results showed that the total number and the distribution density of SNPs, Ts/Tn, and heterozygosity of the two yak populations were close (**Table 2**), indicating that the genetic diversity of the two yak populations was similarity. The PCA results showed that the degree of differentiation between the two populations was low, suggesting that two population had a closer relatedness (**Figure 1**). Tianzhu white yak is a local breed that has been bred artificially for a long time. Previous studies showed that the Tianzhu white yak has a large variation within

TABLE 3 | Allele frequencies for missense mutations in the candidate genes identified in NTWY and LTWY.

Sites	Gene	Amino acid variation	Allele frequency (NTWY)		Allele frequency (LTWY)		Genotype	Genotype frequency (NTWY)	Genotype frequency (LTWY)
			Before mutation	After mutation	Before mutation	After mutation			
c.302G > C	*FGF5*	Ser101Thr	1.00	0.00	0.78	0.22	CC	1.00	0.67
							CG	0.00	0.22
							GG	0.00	0.11
c.199G > A	*SHISA3*	Ala67Thr	0.90	0.10	1.00	0.00	CC	0.90	1.00
							CT	0.00	0.00
							TT	0.10	0.00
c.958G > A	*ACOXL*	Asp320Asn	0.95	0.05	1.00	0.00	CC	0.90	1.00
							CT	0.10	0.00
							TT	0.00	0.00
c.1360G > T	*CDH1*	Val454Leu	1.00	0.00	0.85	0.15	CC	1.00	0.80
							CA	0.00	0.10
							AA	0.00	0.10
c.2274C > A	*MAGEL2*	His758Gln	0.00	1.00	0.06	0.94	GG	0.00	0.00
							GT	0.00	0.11
							TT	1.00	0.89
c.325A > G	*MAGEL2*	Met109Val	0.86	0.14	1.00	0.00	TT	0.90	1.00
							TC	0.00	0.00
							CC	0.10	0.00
c.1579T > C	*PDPK1*	Ser527Pro	0.95	0.05	1.00	0.00	AA	0.90	1.00
							AG	0.10	0.00
							GG	0.00	0.00

the population. The long-haired type should be a subgroup that appeared in a short period, and the degree of differentiation from the normal type of white yak is lower (Qiu et al., 2015). The PCA result was consistent with the current population situation of long-haired white yak.

To reveal the genetic mechanisms of the long-haired phenotype, five selection (F_{ST}, XP-EHH, iHS, XP-CLR, θπ ratio) methods were performed to find candidate genes and pathways. Window scan results showed a shared strong selection region of 0.2 MB in length (CHR6:25,200,001-25,400,000) was detected on two haplotype-based selection methods, and two genes (*FGF5*, *CFAP299*) were located (**Figure 2B**). Currently, *FGF5* is the famous mutant gene found in mammalian species that causes the hairy phenotype variation. Long hair is inherited as a simple recessive trait in animals (Drögemüller et al., 2010). Studies on long-haired mice, dogs, rabbits, and donkeys have shown that the inherited hair length is caused by mutations within *FGF5* gene (Drögemüller et al., 2010; Dierks et al., 2013; Legrand et al., 2014; Zhao et al., 2018). In addition, these genes were enriched into four hair follicle-related pathways (Ras signaling pathway, MAPK signaling pathway, PI3K-Akt signaling pathway, and Rap1 signaling pathway) (**Figure 2C**). Ras signaling is essential for skin development (Drosten et al., 2014). He et al. found Ras and Rap1 signaling pathways were involved in the growth of hair follicle stem cells cultured *in vitro* (He et al., 2020). The MAPK signaling pathway can induce the proliferation and differentiation of hair follicle cells, promote the periodic development of hair follicles, and then affect the growth of villi and the distribution of hair follicles and the number of hair shafts (Zhang et al., 2008; Öztürk et al., 2015; Lu et al., 2021). PI3K/Akt signaling pathway is essential for *de novo* hair follicle regeneration (Chen et al., 2020). Previous studies have found that the PI3K/AKT and ERK1/2 signaling pathways in hair follicle cells can work together to accelerate the transformation of hair follicles from resting phase to growth phase, extend the growth phase of hair follicles, and promote hair follicle development and hair growth (Liu et al., 2020). In addition, a 1.1-MB hotspot (CHR6:61,650,001-62,750,000) was found in five selection signals, and four genes (*ATP8A1*, *SHISA3*, *SLC30A9*, *TMEM33*) were annotated (**Figure 2B**). *SLC30A9* and *ATP8A1* were involved in several pathways related to ion transport, including cation transmembrane transport, magnesium ion binding, zinc ion transport, and cation transmembrane transporter activity pathway (**Figure 2D**). Hair development is closely related to the content of various ions, such as zinc, which plays an important role in animal hair growth (Vallee and Falchuk, 1993). Suliman et al. (1988) found that their wool was sparse and their growth rate slowed down, and the wool fell off on both sides of the back and neck when sheep were zinc deficient (Suliman et al., 1988). Zinc deficiency was also leading to rough fur and shedding in cattle (Ott et al., 1965; Tomlinson et al., 2004). *TMEM33*, *SHISA3*, and *SLC30A9* were also enriched in phospholipid translocation, phospholipid transport, trans-Golgi network, response to endoplasmic reticulum stress (**Figure 2D**). These pathways were involved in the synthesis of extracellular proteins, which related to may be related to the synthesis of hair growth-related proteins (Shore and Tata, 1977; Vitale et al., 1993). In summary, these genes may affect hair growth through ion transport or the synthesis of extracellular proteins.

Due to the low degree of differentiation between the two populations, the method based on genetic differentiation may not be able to identify different genes. A combination of several

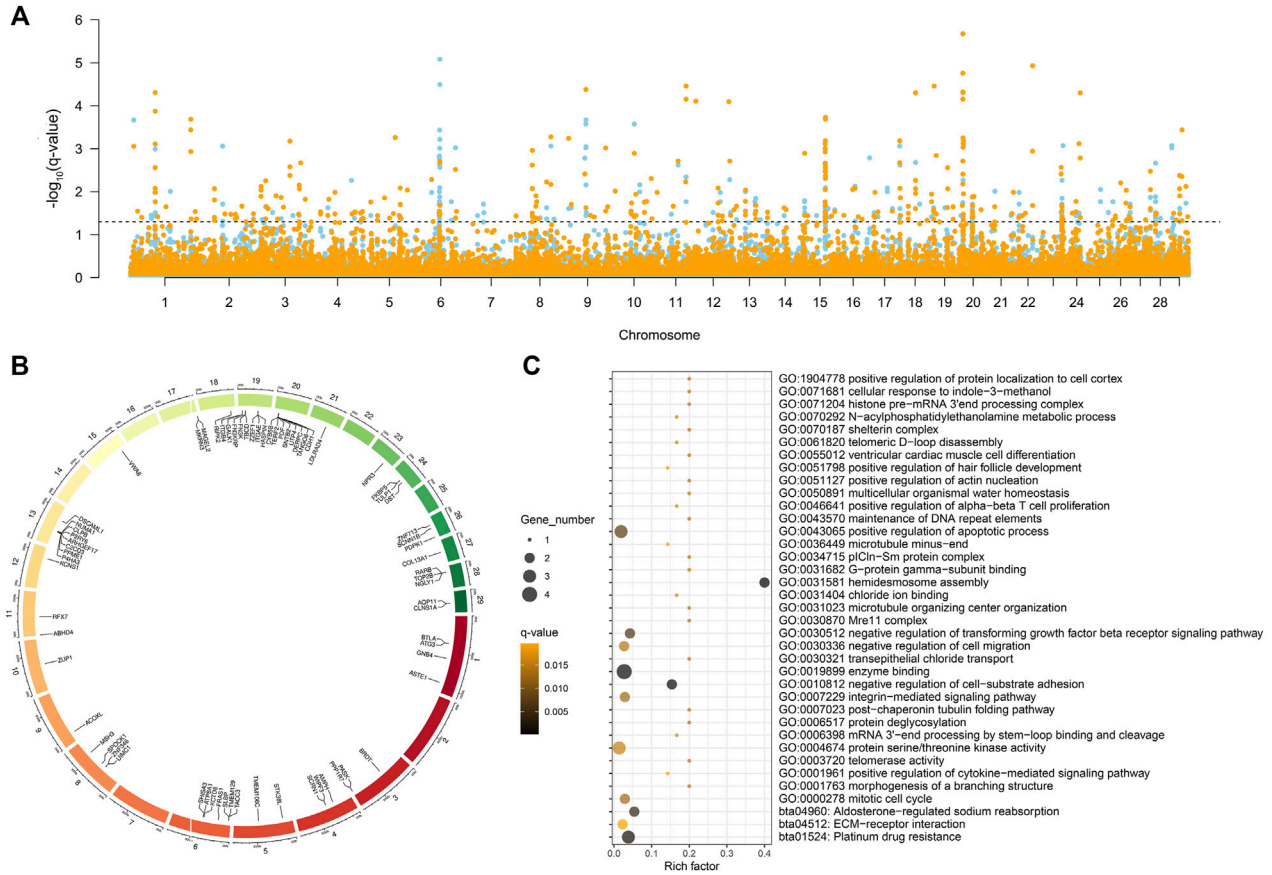


FIGURE 3 | (A) Manhattan plot of decorrelated composite of multiple signals *q*-values of NTWY (skyblue) and LTWY (yellow) populations. **(B)** Overlapping genes identified by DCMS method between NTWY and LTWY. **(C)** Functional enrichment analysis of overlapping genes.

selection methods may be more conducive to this research, and the DCMS method allows more precisely and filters out spurious results specific to other methods (Ma et al., 2015). We calculated DCMS statistics for each population and the overlapping genes were selected as candidate genes associated with phenotypes (**Figure 3A**). In our study, a total of 71 overlapping genes were obtained using the DCMS method (**Figure 3B**). These overlapping genes were enriched into pathways involved in hair follicle development, including positive regulation of hair follicle development, positive regulation of cytokine-mediated signaling pathway, mitotic cell cycle, positive regulation of apoptotic process, negative regulation of transforming growth factor beta receptor signaling pathway, negative regulation of cell-substrate adhesion (**Figure 3C**). Classic studies showed that during embryogenesis, the embryonic *epidermis* and mesenchyme communicated with each other to form a hair follicle (McElwee and Hoffmann, 2000). The strong selection signal of DCMS found on Chromosome 20 (20:10300001-10350000) contained the *CDH1* gene, which mediated the intercellular adhesion in the mammalian *epidermis* and hair follicles as the adhesive component of adherens junctions (Hodivala and Watt, 1994). *CDH1* was weakly expressed in the dermis, while was highly expressed in the *epidermis* and hair follicles (She et al., 2016). Reports showed that *CDH1* played an important role in the formation of melanin in hair follicles and the adhesion of hair follicles and *epidermis* (Larue et al., 1994; Perl et al., 1998; Young et al., 2003; Kuphal and Bosserhoff, 2012). Previous studies also found that continuous hair follicle cycling was dependent on *CDH1* (Young et al., 2003). *ACOXL*, a typical lipid metabolism-related gene, was strongly selected in our study (**Figure 3B**). This enzyme could catalyze the desaturation of acyl-CoAs to 2-trans-enoyl-CoAs in the reductive half-reaction (Brown and Baker, 2003). Festa et al. (2011) found that dermal white adipose tissue (WAT) not only provided animals with thermo-insulation but also modulated regeneration dynamics of pelage hair follicles via the production of paracrine growth factor. Regeneration of the dermal WAT periodically cycles was in synchrony with the hair cycle, undergoing the cycles of expansion and collapse (Chase et al., 1953; Donati et al., 2014). These pieces of evidence suggested that the lipogenesis and lipolysis of WAT could be influenced by the β-oxidation process, so we inferred that *ACOXL* may affect the metabolism of WAT to synchrony affect the hair follicle cycle in yak. One circadian rhythm-related gene (*MAGEL2*) was identified among the overlapping genes (**Figure 3B**). *MAGEL2* has been found to modulate the circadian rhythm: it was

primarily expressed in the suprachiasmatic nucleus where the transcription of *MAGEL2* oscillated in phase with clock-controlled genes. In addition to local paracrine modulators, hair follicles are also regulated by physiological changes that take place throughout the body. For example, several results suggested the involvement of the circadian clock regulate the hair cycle and hair follicle pigmentation (Al-Nuaimi et al., 2014; Hardman et al., 2015). In the mature anagen, clock genes were prominently expressed in the hair matrix, dermal papilla, and other follicular compartments (Plikus et al., 2013). Previous research showed that mice deficient in *Magel2* expression will disrupt circadian rhythm, metabolic and endocrine deficits (Kozlov et al., 2007; Mercer et al., 2013). Therefore, *MAGEL2* may affect the hair growth cycle by influencing the robust rhythmicity of *MAGEL2* expression, which may be one of the reasons for the different hair lengths of Tianzhu white yak. *PDPK1* was involved in the negative regulation of the transforming growth factor beta receptor (TGF-β) signaling pathway. During the development process of the hair follicle, TGF-β1, TGF-β2, and their receptors were locationally and cyclically specifically expressed in hair follicles and were proved to be involved in regulating the growth and development of hair follicle through multiple signaling pathways. Studies of transgene or gene knockout of TGF-β also confirmed that TGF-β related signaling was necessary for hair follicle development. It is indicated that PDPK1 may play an important role in hair development and cycle through TGF-β (Paus et al., 1997; Foitzik et al., 1999).

Through five selection signal methods (F_{ST}, XP-EHH, iHS, XP-CLR, θπ ratio), 0.2-MB and 1.1-MB hotspot were identified, both located on chromosome 6. *FGF5* was identified as the key gene affecting hair length in 0.2-MB hotspot. The enriched pathways (Ras signaling pathway, MAPK signaling pathway, PI3K-Akt signaling pathway, and Rap1 signaling pathway) were involved in the process of hair length differences. The genes (*ATP8A1*, *SHISA3*, *SLC30A9*, *TMEM33*) annotated in 1.1-MB hotspot mainly enriched into two types of pathways, one was ion transport-related pathways, another was endoplasmic reticulum related pathways, which may affect hair follicle development through protein synthesis. The DMCS method further obtained four genes related to hair follicle development (*ACOXL*, *PDPK1*, *MAGEL2*, *CDH1*), which influenced the hair follicle cycle through fat metabolism, growth factors, circadian rhythm, and cell adhesion pathways. The candidate genes and pathways screened in this study were involved in the formation mechanism of hair length in yak. In the next step, further experiments will be performed to verify the function of candidate genes. Our study provided an important reference for breeding, breed improvement, and functional genome research of landscape Tianzhu White yak in China.

DATA AVAILABILITY STATEMENT

The bioproject number of the sequencing data information about long-haired Tianzhu white yak and normal-haired Tianzhu white yak is PRJNA766811 in the NCBI Sequence Read Archive.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Ethics Committee of Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

QB and XM contributed equally towards the construction and execution of this manuscript. CJ and XW revised the manuscript, YW, GM, and PB helped in sample collection, MC and XG provided with the valuable suggestion. CL and PY contributed in the funding for the research.

FUNDING

This research was supported by the National Key Research Program (2021YFD1600200), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2014-LIHPS-01), the Science and Technology Aid Qinghai Cooperation Special Project (2020-QY-212), the National Beef Cattle Industry Technology & System (CARS-37), the Science and Technology program of Gansu Province (20JR5RA580). The National Natural Science Foundation of China (32102524), the Science and Technology program of Gansu Province (21JR7RA032).

REFERENCES

Akey, J. M., Ruhe, A. L., Akey, D. T., Wong, A. K., Connelly, C. F., Madeoy, J., et al. (2010). Tracking Footprints of Artificial Selection in the Dog Genome. *Proc. Natl. Acad. Sci.* 107 (3), 1160–1165. doi:10.1073/pnas.0909918107

Akilli Öztürk, Ö., Pakula, H., Chmielowiec, J., Qi, J., Stein, S., Lan, L., et al. (2015). Gab1 and Mapk Signaling Are Essential in the Hair Cycle and Hair Follicle Stem Cell Quiescence. *Cel Rep.* 13 (3), 561–572. doi:10.1016/j.celrep.2015.09.015

Al-Nuaimi, Y., Hardman, J. A., Bíró, T., Haslam, I. S., Philpott, M. P., Tóth, B. I., et al. (2014). A Meeting of Two Chronobiological Systems: Circadian Proteins Period1 and BMAL1 Modulate the Human Hair Cycle Clock. *J. Invest. Dermatol.* 134 (3), 610–619. doi:10.1038/jid.2013.366

Beg, M., Abdullah, N., Thowfeik, F. S., Altorki, N. K., and McGraw, T. E. (2017). Distinct Akt Phosphorylation States Are Required for Insulin Regulated Glut4 and Glut1-Mediated Glucose Uptake. *Elife* 6, e26896. doi:10.7554/eLife.26896

Boitard, S., Boussaha, M., Capitan, A., Rocha, D., and Servin, B. (2016). Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics* 203 (1), 433–450. doi:10.1534/genetics.115.181594

Brown, L.-A., and Baker, A. (2003). Peroxisome Biogenesis and the Role of Protein Import. *J. Cell. Mol. Med.* 7 (4), 388–400. doi:10.1111/j.1582-4934.2003.tb00241.x

Chase, H. B., Montagna, W., and Malone, J. D. (1953). Changes in the Skin in Relation to the Hair Growth Cycle. *Anat. Rec.* 116 (1), 75–81. doi:10.1002/ar.1091160107

Chen, Y., Fan, Z., Wang, X., Mo, M., Zeng, S. B., Xu, R. H., et al. (2020). PI3K/Akt Signaling Pathway Is Essential for De Novo Hair Follicle Regeneration. *Stem Cel Res Ther* 11 (1), 144–210. doi:10.1186/s13287-020-01650-6

Cingolani, P., Platts, A., Wang, le. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila melanogaster* Strain W1118; Iso-2; Iso-3. *Fly (Austin)* 6 (2), 80–92. doi:10.4161/fly.19695

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 (15), 2156–2158. doi:10.1093/bioinformatics/btr330

Danzan, B., Tsedev, K., and Luvsandorj, N. (2014). The Shedding and Growth Dynamics of Yak Down Wool and Links to Habitat Ecological Conditions. *Asian J. Agric. Rural Dev.* 4 (2), 156–161. doi:10.22004/ag.econ.198394

Dierks, C., Mã¶Mke, S., Philipp, U., and Distl, O. (2013). Allelic Heterogeneity of FGF5 Mutations Causes the Long-Hair Phenotype in Dogs. *Anim. Genet.* 44, 425–431. doi:10.1111/age.12010

Donati, G., Proserpio, V., Lichtenberger, B. M., Natsuga, K., Sinclair, R., Fujiwara, H., et al. (2014). Epidermal Wnt/-catenin Signaling Regulates Adipocyte Differentiation via Secretion of Adipogenic Factors. *Proc. Natl. Acad. Sci.* 111 (15), E1501–E1509. doi:10.1073/pnas.1312880111

Drögemüller, C., Rüfenacht, S., Wichert, B., and Leeb, T. (2010). Mutations within the FGF5 Gene Are Associated with Hair Length in Cats. *Anim. Genet.* 38 (3), 218–221. doi:10.1111/j.1365-2052.2007.01590.x

Drosten, M., Lechuga, C. G., and Barbacid, M. (2014). Ras Signaling Is Essential for Skin Development. *Oncogene* 33 (22), 2857–2865. doi:10.4161/sgtp.2690510.1038/onc.2013.254

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

Fay, J. C., and Wu, C. (2000). Hitchhiking under Positive Darwinian Selection. *Genetics* 155 (3), 1405–1413. doi:10.1093/genetics/155.3.1405

Festa, E., Fretz, J., Berry, R., Schmidt, B., Rodeheffer, M., Horowitz, M., et al. (2011). Adipocyte Lineage Cells Contribute to the Skin Stem Cell Niche to Drive Hair Cycling. *Cell* 146 (5), 761–771. doi:10.1016/j.cell.2011.07.019

Foitzik, K., Paus, R., Doetschman, T., and Paolo Dotto, G. (1999). The TGF-B2 Isoform Is Both a Required and Sufficient Inducer of Murine Hair Follicle Morphogenesis. *Dev. Biol.* 212 (2), 278–289. doi:10.1006/dbio.1999.9325

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 (8), 1176–1177. doi:10.1093/bioinformatics/bts115

Grossman, S. R., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., et al. (2010). A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science* 327 (5967), 883–886. doi:10.1126/science.1183863

Hardman, J. A., Tobin, D. J., Haslam, I. S., Farjo, N., Farjo, B., Al-Nuaimi, Y., et al. (2015). The Peripheral Clock Regulates Human Pigmentation. *J. Invest. Dermatol.* 135 (4), 1053–1064. doi:10.1038/jid.2014.442

He, N., Su, R., Wang, Z., Zhang, Y., and Li, J. (2020). Exploring Differentially Expressed Genes between Anagen and Telogen Secondary Hair Follicle Stem Cells from the Cashmere Goat (*Capra hircus*) by RNA-Seq. *PLoS ONE* 15 (4), e0231376. doi:10.1371/journal.pone.0231376

Hodivala, K., and Watt, F. (1994). Evidence that Cadherins Play a Role in the Downregulation of Integrin Expression that Occurs during Keratinocyte Terminal Differentiation. *J. Cel Biol.* 124 (4), 589–600. doi:10.1083/jcb.124.4.589

Hua, C., Patterson, N., and Reich, D. (2010). Population Differentiation as a Test for Selective Sweeps. *Genome Res.* 20 (3), 393–402. doi:10.1101/gr.100545.109

Kozlov, S. V., Bogenpohl, J. W., Howell, M. P., Wevrick, R., Panda, S., Hogenesch, J. B., et al. (2007). The Imprinted Gene Magel2 Regulates normal Circadian Output. *Nat. Genet.* 39 (10), 1266–1272. doi:10.1038/ng2114

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an Information Aesthetic for Comparative Genomics. *Genome Res.* 19 (9), 1639–1645. doi:10.1101/gr.092759.109

Kuphal, S., and Bosserhoff, A. K. (2012). E-cadherin Cell-Cell Communication in Melanogenesis and during Development of Malignant Melanoma. *Arch. Biochem. Biophys.* 524 (1), 43–47. doi:10.1016/j.abb.2011.10.020

Larue, L., Ohsugi, M., Hirchenhain, J., and Kemler, R. (1994). E-cadherin Null Mutant Embryos Fail to Form a Trophectoderm Epithelium. *Proc. Natl. Acad. Sci.* 91 (17), 8263–8267. doi:10.1073/pnas.91.17.8263

Legrand, R., Tiret, L., and Abitbol, M. (2014). Two Recessive Mutations in FGF5 Are Associated with the Long-Hair Phenotype in Donkeys. *Genet. Sel Evol.* 46 (1), 65–67. doi:10.1186/s12711-014-0065-5

Lewontin, R. C., and Krakauer, J. (1973). Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. *Genetics* 74 (1), 175–195. doi:10.1093/genetics/74.1.175

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Liu, X., Zhang, P., Zhang, X., Li, X., Bai, Y., Ao, Y., et al. (2020). Fgf21 Knockout Mice Generated Using CRISPR/Cas9 Reveal Genetic Alterations that May Affect Hair Growth. *Gene* 733, 144242. doi:10.1016/j.gene.2019.144242

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., Verity, B., et al. (2017). Composite Measures of Selection Can Improve the Signal-to-noise Ratio in Genome Scans. *Methods Ecol. Evol.* 8 (6), 717–727. doi:10.1111/2041-210X.12774

Lu, Q., Gao, Y., Fan, Z., Xiao, X., Chen, Y., Si, Y., et al. (2021). Amphiregulin Promotes Hair Regeneration of Skin-Derived Precursors via the PI3K and MAPK Pathways. *Cell Prolif* 54 (9), e13106. doi:10.1111/cpr.13106

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., and Simianer, H. (2015). Properties of Different Selection Signature Statistics and a New Strategy for Combining Them. *Heredity* 115 (5), 426–436. doi:10.1038/hdy.2015.42

McElwee, K., and Hoffmann, R. (2000). Growth Factors in Early Hair Follicle Morphogenesis. *Eur. J. Dermatol.* 10 (5), 341–350. doi:10.1046/j.1524-4725.2000.026007714.x

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 (9), 1297–1303. doi:10.1101/gr.107524.110

Mercer, R. E., Michaelson, S. D., Chee, M. J. S., Atallah, T. A., Wevrick, R., Colmers, W. F., et al. (2013). Magel2 Is Required for Leptin-Mediated Depolarization of POMC Neurons in the Hypothalamic Arcuate Nucleus in Mice. *Plos Genet.* 9 (1), e1003207. doi:10.4161/rdis.2442110.1371/journal.pgen.1003207

Ott, E., Smith, W. H., Martin, S., Parker, H. E., and Beeson, W. M. (1965). ZINC DEFICIENCY SYNDROME IN THE YOUNG CALF. *J. Anim. Sci.* 3, 735–741. doi:10.2527/jas1965.243735x

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., et al. (2002). Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. *Cell* 109 (3), 307–320. doi:10.1016/s0092-8674(02)00722-5

Paus, R., Foitzik, K., Welker, P., Bulfone-Paus, S., and Eichmüller, S. (1997). Transforming Growth Factor-β Receptor Type I and Type II Expression during Murine Hair Follicle Development and Cycling. *J. Invest. Dermatol.* 109 (4), 518–526. doi:10.1111/1523-1747.ep12336635

Pearse, D. E., and Crandall, K. A. (2004). Beyond FST: Analysis of Population Genetic Data for Conservation. *Conservation Genet.* 5 (5), 585–602. doi:10.1007/s10592-003-1863-4

Perl, A.-K., Wilgenbus, P., Dahl, U., Semb, H., and Christofori, G. (1998). A Causal Role for E-Cadherin in the Transition from Adenoma to Carcinoma. *Nature* 392 (6672), 190–193. doi:10.1038/32433

Plikus, M. V., Vollmers, C., de la Cruz, D., Chaix, A., Ramos, R., Panda, S., et al. (2013). Local Circadian Clock gates Cell Cycle Progression of Transient Amplifying Cells during Regenerative Hair Cycling. *Proc. Natl. Acad. Sci. USA* 110 (23), E2106–E2115. doi:10.1073/pnas.1215935110

Qiu, Q., Wang, L., Wang, K., Yang, Y., Ma, T., Wang, Z., et al. (2015). Yak Whole-Genome Resequencing Reveals Domestication Signatures and Prehistoric Population Expansions. *Nat. Commun.* 6, 10283. doi:10.1038/ncomms10283

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting Recent Positive Selection in the Human Genome from Haplotype Structure. *Nature* 419 (6909), 832–837. doi:10.1111/jpi.1261010.1038/nature01140

Sabeti, P. C., Varilly, P., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., et al. (2007). Genome-wide Detection and Characterization of Positive Selection in Human Populations. *Nature* 449, 913–918. doi:10.1038/nature06250

Schrader, M., and Fahimi, H. D. (2006). Peroxisomes and Oxidative Stress. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1763 (12), 1755–1766. doi:10.1016/j.bbamcr.2006.09.006

She, P. C., Liang, C. N., Pei, J., Chu, M., Guo, X., and Yan, P. (2016). Fetal Skin Hair Follicle Morphogenesis and E-Cadherin Expression of the Yak. *Chin. J. Anim. Vet. Sci.* 47 (02), 397–403. doi:10.11843/j.issn.0366-6964.2016.02.025

Shore, G. C., and Tata, J. R. (1977). Functions for Polyribosome-Membrane Interactions in Protein Synthesis. *Biochim. Biophys. Acta (Bba) - Rev. Biomembranes* 472 (2), 197–236. doi:10.1016/0304-4157(77)90017-x

Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J., et al. (2004). The Genomic Distribution of Population Substructure in Four Populations Using 8,525 Autosomal SNPs. *Hum. genomics* 1 (4), 274–286. doi:10.1186/1479-7364-1-4-274

Smith, J. M., and Haigh, J. (1974). The Hitch-Hiking Effect of a Favourable Gene. *Genet. Res.* 23 391–403. doi:10.1017/s0016672300014634

Storey, J. D., and Tibshirani, R. (2003). Statistical Significance for Genomewide Studies. *Proc. Natl. Acad. Sci.* 100 (16), 9440–9445. doi:10.1073/pnas.1530509100

Suliman, H. B., Abdelrahim, A. I., Zakia, A. M., and Shommein, A. M. (1988). Zinc Deficiency in Sheep: Field Cases. *Trop. Anim. Health Prod.* 20 (1), 47–51. doi:10.1007/BF02239646

Tajima, F. (1983). Evolutionary Relationship of DNA Sequences in Finite Populations. *Genetics* 105 (2), 437–460. doi:10.1093/genetics/105.2.437

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123 (3), 585–595. doi:10.1093/genetics/123.3.585

Tomlinson, D. J., Mülling, C. H., and Fakler, T. M. (2004). Invited Review: Formation of Keratins in the Bovine Claw: Roles of Hormones, Minerals, and Vitamins in Functional Claw Integrity. *J. Dairy Sci.* 87 (4), 797–809. doi:10.3168/jds.S0022-0302(04)73223-3

Turner, S. D. (2018). Qqman: An R Package for Visualizing GWAS Results Using QQ and Manhattan Plots. *J. Open Source Softw.* 3 (25), 731. doi:10.21105/joss.00731

Vallee, B. L., and Falchuk, K. H. (1993). The Biochemical Basis of Zinc Physiology. *Physiol. Rev.* 73 (1), 79–118. doi:10.1152/physrev.1993.73.1.79

Vera, G., Jansen, R. C., and Suppi, R. L. (2008). R/parallel--speeding up Bioinformatics Analysis with R. *BMC bioinformatics* 9 (1), 390–396. doi:10.1186/1471-2105-9-390

Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L., and Lotterhos, K. E. (2017). Minotaur: A Platform for the Analysis and Visualization of Multivariate Results from Genome Scans with R Shiny. *Mol. Ecol. Resour.* 17 (1), 33–43. doi:10.1111/1755-0998.12579

Vitale, A., Ceriotti, A., and Denecke, J. (1993). The Role of the Endoplasmic Reticulum in Protein Synthesis, Modification and Intracellular Transport. *J. Exp. Bot.* 9, 1417–1444. doi:10.1093/jxb/44.9.1417

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *Plos Biol.* 4, e72. doi:10.1371/journal.pbio.0040072

Wiener, G., Han, J. L., and Long, R.J. (2003). "Origins, Domestication and Distribution of Yak," in *The yak*. Editors G. Wiener, J. L. Han, and R. J. Long (Bangkok: FAO Regional Office for Asia and the Pacific), 1–16.

Young, P., Boussadia, O., Halfter, H., Grose, R., Berger, P., Leone, D. P., et al. (2003). E-cadherin Controls Adherens Junctions in the Epidermis and the Renewal of Hair Follicles. *EMBO J.* 22 (21), 5723–5733. doi:10.1093/emboj/cdg560

Yurchenko, A. A., Deniskova, T. E., Yudin, N. S., Dotsev, A. V., Khamiruev, T. N., Selionova, M. I, et al. (2019). High-Density Genotyping Reveals Signatures of Selection Related to Acclimation and Economically Important Traits in 15 Local Sheep Breeds From Russia. *BMC Genom.* 20 (3), 1–19. doi:10.1186/s12864-019-5537-0

Zhang, J., Teng, X., Si, L., Zhou, P., Kong, X., and Hu, L. (2008). New Evidence for the Involvement of the EGF Receptor Pathway in Hair Follicle Morphogenesis in Uncv Mice. *Genes and Genomics* 30 (4), 347–353. doi:10.1007/s00705-002-0974-4

Zhao, B., Chen, Y., Hao, Y., Yang, N., Wang, M., Mei, M., et al. (2018). Transcriptomic Analysis Reveals Differentially Expressed Genes Associated with Wool Length in Rabbit. *Anim. Genet.* 49, 428–437. doi:10.1111/age.12701

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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 01 April 2022
doi: 10.3389/fgene.2022.858232

Multi-Omics Reveals Different Strategies in the Immune and Metabolic Systems of High-Yielding Strains of Laying Hens

*Muhammad Arsalan Iqbal[1], Henry Reyer[1], Michael Oster[1], Frieder Hadlich[1], Nares Trakooljul[1], Alvaro Perdomo-Sabogal[1], Sonja Schmucker[2], Volker Stefanski[2], Christoph Roth[2], Amélia Camarinha Silva[2], Korinna Huber[2], Vera Sommerfeld[2], Markus Rodehutscord[2], Klaus Wimmers[1,3] and Siriluck Ponsuksili[1]**

[1]Research Institute for Farm Animal Biology, Institute of Genome Biology, Dummerstorf, Germany, [2]University of Hohenheim, Institute of Animal Science, Stuttgart, Germany, [3]University Rostock, Faculty of Agricultural and Environmental Sciences, Rostock, Germany

OPEN ACCESS

Edited by:
Gerson Barreto Mourão,
University of São Paulo, Brazil

Reviewed by:
Monica Correa Ledur,
Embrapa Suínos e Aves, Brazil
Ottmar Distl,
University of Veterinary Medicine Hannover, Germany

***Correspondence:**
Siriluck Ponsuksili
ponsuksili@fbn-dummerstorf.de

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

Received: 19 January 2022
Accepted: 10 March 2022
Published: 01 April 2022

Citation:
Iqbal MA, Reyer H, Oster M, Hadlich F, Trakooljul N, Perdomo-Sabogal A, Schmucker S, Stefanski V, Roth C, Camarinha Silva A, Huber K, Sommerfeld V, Rodehutscord M, Wimmers K and Ponsuksili S (2022) Multi-Omics Reveals Different Strategies in the Immune and Metabolic Systems of High-Yielding Strains of Laying Hens. Front. Genet. 13:858232. doi: 10.3389/fgene.2022.858232

Lohmann Brown (LB) and Lohmann Selected Leghorn (LSL) are two commercially important laying hen strains due to their high egg production and excellent commercial suitability. The present study integrated multiple data sets along the genotype-phenotype map to better understand how the genetic background of the two strains influences their molecular pathways. In total, 71 individuals were analyzed (LB, $n = 36$; LSL, $n = 35$). Data sets include gut miRNA and mRNA transcriptome data, microbiota composition, immune cells, inositol phosphate metabolites, minerals, and hormones from different organs of the two hen strains. All complex data sets were pre-processed, normalized, and compatible with the mixOmics platform. The most discriminant features between two laying strains included 20 miRNAs, 20 mRNAs, 16 immune cells, 10 microbes, 11 phenotypic traits, and 16 metabolites. The expression of specific miRNAs and the abundance of immune cell types were related to the enrichment of immune pathways in the LSL strain. In contrast, more microbial taxa specific to the LB strain were identified, and the abundance of certain microbes strongly correlated with host gut transcripts enriched in immunological and metabolic pathways. Our findings indicate that both strains employ distinct inherent strategies to acquire and maintain their immune and metabolic systems under high-performance conditions. In addition, the study provides a new perspective on a view of the functional biodiversity that emerges during strain selection and contributes to the understanding of the role of host–gut interaction, including immune phenotype, microbiota, gut transcriptome, and metabolome.

Keywords: data integration, mixOmics, microbiota, miRNA, mRNA, immune cells, metabolites, laying hen

INTRODUCTION

The two commercially important laying hen strains, Lohmann Brown (LB) and Lohmann Selected Leghorn (LSL), are selected for high egg production (Singh et al., 2009; Habig et al., 2012; Reyer et al., 2021). Although LB and LSL are nearly identical in egg production performance, these strains differ considerably in other phenotypic traits including body weight, gene expression, immunological

traits, bone metabolism, and gastrointestinal phytate degradation (Kaufmann et al., 2011; Silversides et al., 2012; Sommerfeld et al., 2020; Ponsuksili et al., 2021).

Previously, comparative transcriptomics from brain tissue revealed that while transcripts' upregulation contributed to immune system processes in LSL, the downregulation was involved in phosphorus (P) metabolism and signaling pathways (Habig et al., 2012). Following dietary interventions with reduced P and calcium (Ca) intake, the intestinal transcriptional profile of the two strains showed strain-specific alterations, for example, in the cell proliferation rate and extracellular matrix formation, which might also be due to different mineral and vitamin D requirements of LB and LSL (Reyer et al., 2021). Similarly, the miRNAs expression profiles from the jejunum of LB and LSL hens fed with different amounts of Ca and P indicate that miRNA targets contributed to metabolic pathways in LB; miRNA targets were involved in Ca signaling pathways and mitochondrial dysfunction in LSL (Iqbal et al., 2021). In addition, the deep-sequenced miRNAs of the jejunum of LB and LSL at different production periods demonstrated that miRNAs play a pivotal role in regulating gene expression and impacting intestinal homeostasis differently in both strains (Ponsuksili et al., 2021).

Several studies have suggested that the host's genetic background is a factor that might influence gut microbiota composition (Org et al., 2015; Schokker et al., 2015; Han et al., 2016; Kers et al., 2018). The host intestinal epithelia and gut microbiota ecosystem are complex and consist of diverse molecular activities, including immune and metabolic functions (Simon et al., 2016; Broom and Kogut, 2018; Borda-Molina et al., 2021). Recent studies demonstrated a shift in the microbiota during the laying hens' lifespan or change in metabolite profiles between the LB and LSL strains (Gonzalez-Uarquin et al., 2021; Joat et al., 2021). Concerns about the environment, nutrient supply, and farm profit contribute to the increasing attention to the Ca and mineral P supplements in animal feed (Delezie et al., 2015). Numerous studies indicated that adding additional P and Ca to the feed of broiler chickens significantly reduced endogenous Inositol hexakisphosphate ($InsP_6$) degradation (Tamim et al., 2004; Shastak et al., 2014; Zeller et al., 2015). Compared with broilers, laying hens need less P in their diet, but they require substantially more Ca because of eggshell formation (Ahmadi and Rodehutscord, 2012). Consequently, the processes of $InsP_6$ degradation, myo-inositol (MI) release, and P/Ca utilization might be more distinct in laying hens than in broilers (Sommerfeld et al., 2020; Sommerfeld et al., 2020). Sommerfeld et al. (2020) described the effects of dietary Ca and P on intestinal phytate degradation and mineral utilization during the laying phase in LB and LSL and also elaborated that measured traits depend on the hen strain. It was concluded that to meet their respective mineral demands, LB and LSL use different mechanisms, including transcellular transport in LB and more effective phytate degradation in LSL. In addition, results indicated that the variation in Ca and P concentration in feed affects the body weight of the LB strain, while no effect was observed in the LSL strain.

Since poultry production intensifies and antibiotics are under pressure to be used less frequently, maintaining and improving poultry health by promoting animal-intrinsic mechanisms become more relevant (Swaggerty et al., 2019). Here, factors such as nutrition and genetics play a critical role in modulating immunity in commercial poultry production (Koenen et al., 2002; Kidd, 2004; Kjærup et al., 2017; Nie et al., 2018; Hofmann et al., 2021). Interestingly, laying hens fed higher nonphytate-P levels were found to have higher interferon (IFN) levels in the blood, suggesting an improved immune system (Nie et al., 2018).

Using advancements in high-throughput approaches and the availability of multiple omics data generated from similar and different experiments, data integration provides the material for a more comprehensive biological interpretation at multiple levels, which can support the unfolding of the complex biological processes scientifically and holistically (Subramanian et al., 2020). An earlier study identified a highly correlated multi-omics signature including host mRNA, miRNA, and microbial data for identifying the molecular drivers for P utilization (PU) in Japanese quail (Ponsuksili et al., 2020) using the multi-block discriminant analysis with DIABLO (Data Integration Analysis for Biomarker discovery using a Latent cOmponents) embedded in the R package "mixOmics" (Rohart et al., 2017; Singh et al., 2019).

In our preliminary work, individual datasets covering physiological data, hormones, metabolites, immune traits, host transcriptome, or microbiome were retrieved from the same animals of the LB and LSL strains in separate work packages of the DFG (Deutsche Forschungsgemeinschaft) Research Unit "P-Fowl" in the context of divergent dietary Ca and P supply (Sommerfeld et al., 2020; Hofmann et al., 2021; Iqbal et al., 2021; Reyer et al., 2021). Individual analysis of each dataset for the two laying hen strains in the various studies revealed specific differences at each level and lacked an integrated multi-omics view of these specific physiological changes. The integration of multi-omics provides a new perspective on the functional biodiversity that arises during strain selection and contributes to our understanding of the role of host–gut interactions. In the present study, we integrate these datasets in a holistic biological analysis to characterize the functional biodiversity of the two strains that contribute to attaining comparable performance *via* different molecular mechanisms. Finally, strain-specific bio-signatures are being uncovered to deepen our understanding of the relationships between underlying immunology, genetics, metabolism, developmental processes, and gut microbial community composition.

MATERIALS AND METHODS

Experimental Design and Sample Selection

We used previously published datasets (Sommerfeld et al., 2020; Hofmann et al., 2021; Iqbal et al., 2021; Reyer et al., 2021). In brief, hatchlings of brown (LB) and white (LSL) layer strains, representing two distinct genetic backgrounds, were obtained

from a breeding company (Lohmann Tierzucht GmbH, Cuxhaven, Germany). LB and LSL were selected for their egg production performance while being monitored for bone quality, egg quality, and behavior (Preisinger, 2018). During rearing, the standard management conditions for the pullet phase of the experiment station of the University of Hohenheim were applied (Sommerfeld et al., 2020). All hens received identical corn-soybean meal-based diets meeting or exceeding the supply recommendations of the breeding company for the starter, grower, pre-laying, and layer phases. From week 27 to 31, forty individuals per strain received experimental diets with reduced Ca, P, or a combination thereof, as previously described by Sommerfeld et al. (2020). In week 31, the blood and intestinal contents were sampled. Trunk blood or vein blood was collected depending on the targets of analysis (immune features, metabolites, P, Ca, MI, or hormones). Tissue samples for transcriptome analysis were collected from jejunum mucosa, while digesta and mucosa from the duodenum were collected for the microbiota investigations. Since the study design was balanced between LB and LSL in terms of diet and the effects of strain clearly outweighed the dietary alterations [as reported earlier (Sommerfeld et al., 2020; Iqbal et al., 2021; Reyer et al., 2021)], the dietary groups were combined in the present study for the downstream analysis. Specifically, residuals were calculated for each parameter after adjusting for diet and father effects and used for further downstream analysis.

Data Pre-Processing

The miRNA and mRNA expression profiles were gathered from the jejunum. The read count matrices of miRNA and mRNA from the same birds were obtained from our previous studies under accession numbers (E-MTAB-9136) and (E-MTAB-9109), respectively (Iqbal et al., 2021; Reyer et al., 2021). The samples with lower counts, that is, outliers deviating from the mean by more than two SD, were excluded. Finally, the total number of samples analyzed in this study was $n = 71$, whereas $n = 36$ for LB and $n = 35$ for LSL. Microbiota data were represented as amplicon sequence variants (ASVs) that were deduced from 16S rRNA sequencing from the duodenal mucosa and the duodenal digesta of LB and LSL. Initially, ASVs were assigned to taxa at the genus level and were filtered so that only taxa with more than one observation in at least half of the samples were considered. Subsequently, miRNA and mRNA read count matrices and microbiota ASVs count matrices were transformed by variance-stabilizing transformation (VST) using the DESeq2 R package (Love et al., 2014).

Previously collected data from the same birds include immunological traits such as the counts of many types of T cells, B cells, heterophils, thrombocytes, and monocytes from blood, the spleen, and pharyngeal tonsils. Data on metabolites are from blood and digesta, including InsP3-6 isomers, P, Ca, MI, and hormones, and other phenotypic traits include Ca and P intake, Ca and P utilization, Ca and P excretion, feed intake, and body weight (Sommerfeld et al., 2020; Hofmann et al., 2021).

After pre-processing by transformation or normalization of each data type, data were adjusted for systemic effects using JMP Genomics (SAS Institute, Cary, NC, United States) mixed analyses of variance. Diet was used as a fixed effect and hen father as a random effect. The residuals after adjustment for diet and father were further used for the integration analysis.

Data Integration

To integrate the data, we used the R package mixOmics (version.6.10.9) (Rohart et al., 2017). All preprocessed input data matrices were subjected to mixOmics for further filtering, removing predictors with zero or near-zero variance. We integrated six data blocks: mRNA (13,455 features), miRNA (185 features), immune system (54 features), metabolome (29 features), microbiome (111 features), and phenotypic features (11 features).

Data integration, classification, feature selection, and visualization were carried out by Data Integration Analysis and Biomarker discovery using Latent variable approaches for Omics studies (DIABLO) (Singh et al., 2019). The DIABLO multi-omic approach instantly predicts significant biomarkers, including mRNAs, metabolites, proteins, and miRNA. DIABLO is the first multivariate integrative classification method to identify correlated or co-expressed features from heterogeneous data sets. We used the N-integration supervised Sparse Partial Least Square Discriminant Analysis (SPLS-DA) approach for feature selection (Lê Cao et al., 2011; Gley et al., 2021). The *block. splsda*() function was used to identify signatures composed of highly correlated variables across the multiple matrix sets, enabling us to detect a confident relationship between the data sets (Singh et al., 2019).

To evaluate the number of parameters and global performance, to select the appropriate metric distance, and to determine the number of components for our block. splsda analysis, we used DIABLO's pref () function. The parameters set for the pref () function were Mfold validation ($n = 5$) and cross-validation (nrepeat = 10). For all supervised N-integration models, the tuning function was crucial for selecting the optimal components and features. We ran the tuning function tune. block.splsda () to predict the optimal number of features that were finally used for our block. splsda analysis. The best performance was obtained with the optimal component selection based on the balanced error rate (BER). The parameters used for the tune. block.splsda () function were M-fold validation ($n = 5$) and cross-validation (nrepeat = 10). In addition, the parameter distance metric for splsda to estimate the classification error rate (dist = max dist) and two misclassification measures, the total error rate and balanced error rate (BER) were used.

Subsequently, visualization of the block. splsda results was accomplished by various plotting functions in a mixomics environment. The discriminant analysis results were visualized by the *PlotArrow*() and *plotindiv*() functions. The *PlotArrow*() function plotted an arrow plot, which indicates the components' scores from multiple datasets. It generates arrows between the scores associated with two or more groups, in our case, the two groups LB and LSL. In contrast, the

plotindiv() function generates the individual plot, which depicts different data blocks, where each block sample shows differentiation between two or more groups. The components of the block. splsda results were visualized by the *plotLoadings*() function. The *plotLoadings*() function plotted loading plots that represent the loading weights of each feature selected on each dataset block. At the same time, the size bar indicates the significance of the selected feature between two or more groups. In addition, the bar's color is associated with the group in which the selected features are most prevalent. A Circos correlation plot was generated by the *circosplot*() function; the plot demonstrates the significant signature from multiple datasets and their correlation coefficient. LB could be distinguished from LSL laying hens *via* two components of biomarkers from the heterogeneous data sets.

Correlation Analysis of Bio-Signature Features and Gut mRNA Transcripts

The miRNAs selected using the above approaches (DIABLO) were used for the prediction of target mRNAs and the study of negative correlation. Therefore, variance-stabilized counts were used. A heatmap of bio-signature miRNAs selected from mixomics was generated using gplots (version 3.1.1) R package function *heatmap.2*() (Warnes et al., 2016). The scatter plot of differentially expressed mRNAs was visualized by *plot*() within the R programming environment (https://www.R-project.org/).

The miRNAs and their potential mRNA targets obtained from the recent chicken genome assembly (GRC6a) were predicted using RNAhybrid version 2.1.2 by setting the parameter binding energy with a cut-off of 25 k, the helix constraint in a range from 2 to 7, and one hit per target (Rehmsmeier et al., 2004). MiRNAs and their downstream mRNA targets were selected based on their minimum free energy and *p*-value, as previously described (Iqbal et al., 2021). The Pearson correlation was calculated between bio-signature miRNAs and differentially expressed (DE) mRNAs at FDR ≤ 5%. In addition, mRNAs inversely correlated with miRNAs were included for further downstream analysis. Moreover, strain-specific correlation analyses were performed between jejunal mRNA profiles and bio-signature revealed for data sets of immune cells, microbes, blood/digesta parameters (metabolites, P, Ca, MI, and hormones), and phenotypic traits. Pearson correlation was calculated between the complete set of mRNAs with datasets mentioned above within LB and LSL. The correlation was considered significant at FDR ≤ 5%.

Gene Ontology and KEGG Pathway Enrichment Analysis

Functional annotation enrichment analysis was performed on the identified miRNAs and their inversely correlated DE mRNA targets. The mRNAs negatively correlated with miRNAs were subjected to ClueGO (version.2.5.1) and Cluepedia (version 1.5.7) plug-ins within Cytoscape (3.6.1.) for gene ontology (immune system processes) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (Bindea et al., 2009; Saito et al., 2012; Bindea et al., 2013). Likewise, pairs of inversely correlated miRNAs/mRNAs were subject to DAVID (version 6.8) for gene ontology (biological processes) enrichment analysis (Huang et al., 2007). GOplot (version 1.0.2) within the R programming environment generates GO circular and Circos plots for gene ontology enrichment analysis (Walter et al., 2015). ClueGO was employed to plot a functionally annotated KEGG pathways network.

Moreover, gene ontology and KEGG pathway enrichment analyses were performed considering the complete set of jejunal mRNAs correlated with bio-signature of immune cells, microbes, blood/digesta parameters, and phenotypic traits at the cutoff criteria of FDR ≤ 0.05 and ($r \geq \pm 0.5$) within LB and LSL. For gene ontology and KEGG pathway enrichment analyses, these were further subjected to DAVID (version 6.8) and ClueGO, respectively. ClueGO revealed KEGG pathway enrichment networks, while DAVID results were imported to the ggplot2 R package to create dot plots for gene ontology enrichment analysis (Wickham et al., 2016). Parameters used for ClueGO and DAVID were right-sided hypergeometric tests to calculate the *p*-value, Benjamin–Hochberg for multiple testing correction, and *Gallus gallus* as a reference genome. The KEGG pathways, biological processes, and immune system processes with $p \leq 0.05$ were considered significant.

RESULTS

Our study combines miRNA expression, mRNA expression, immune cell profiles, microbial composition, blood/digesta data (metabolites, P, Ca, MI, and hormones), and phenotypic traits from different tissues, blood, and plasma of LB and LSL chicken strains. The initial datasets included 80 birds. After pre-processing and filtration, data from 71 animals (36 from LB and 35 from LSL) were considered for further downstream analysis. We developed a framework for integrating these data sets into one analysis using the mixOmics platform and predicting significant bio-signatures from the heterogeneous dataset (**Figure 1A**). To examine the data variation between LB and LSL, we performed a discriminatory analysis using the Sparse Partial Least Square Discriminant Analysis (SPLS-DA) supervised approach available in the mixOmics R package. A mixOmics revealed the selection of the most discriminant features between the two laying hen strains, including 20 miRNAs, 20 mRNAs, 16 immune parameters, 10 microbes, 11 phenotypic traits, and 16 blood/digesta parameters (**Figure 1A**). In order to characterize the molecular and metabolic routes linked to these selected biomarkers, the individual features were further used for correlation with selected miRNAs and mRNAs pairs (**Figure 1B**) and with the whole set of intestinal transcripts (**Figure 1C**).

Considering the different datasets available for the laying hens, a considerable degree of separation between LB and LSL samples was achieved (**Figure 2A**). Discriminant analyses performed for

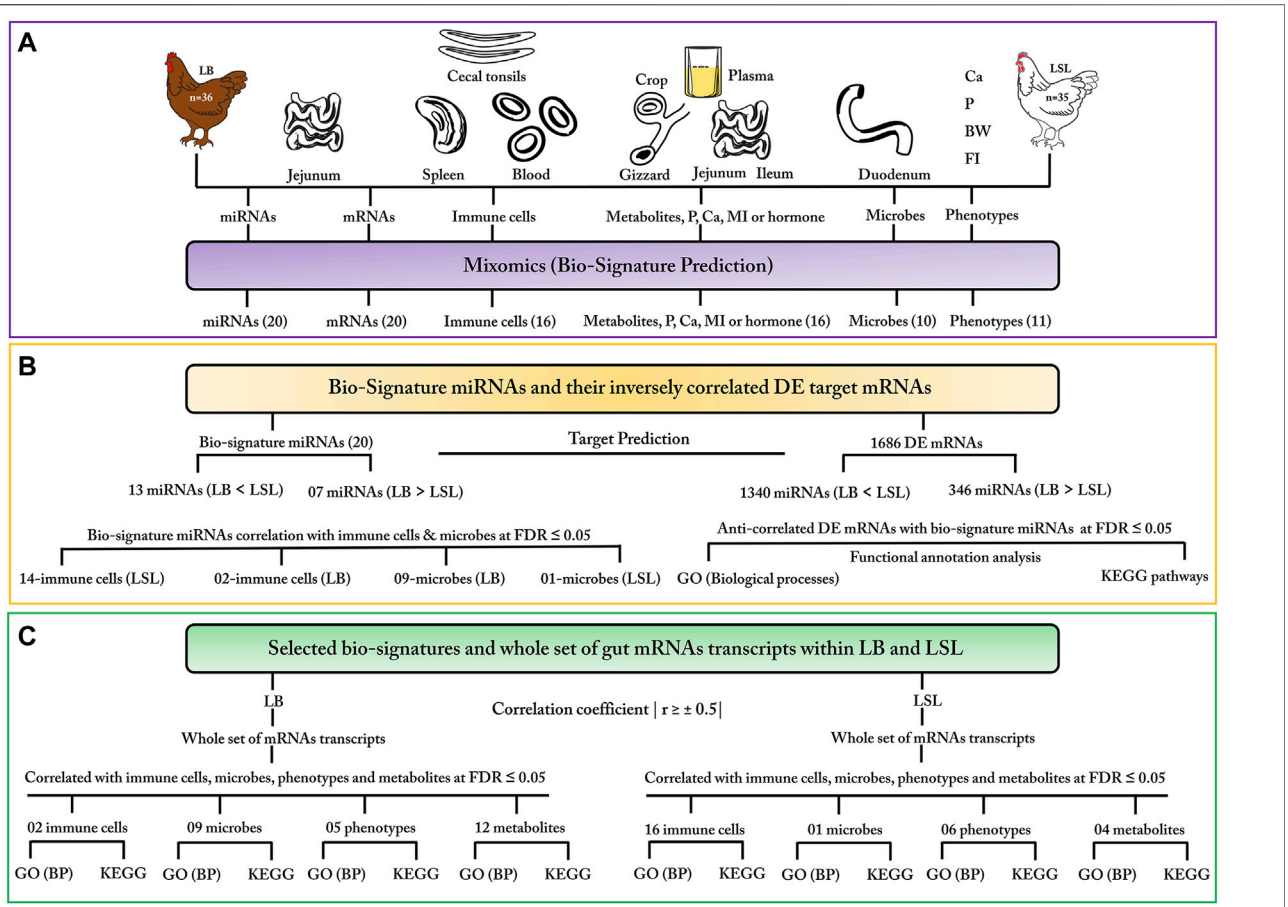


FIGURE 1 | Framework for multi-omics data integration and downstream analysis. **(A)** Data types: miRNAs, mRNAs, immune cells, microbes, blood/digesta parameters, and phenotypic traits were measured from different tissues, blood, and plasma of the LB and LSL laying hen strains. Discriminant analysis in mixOmics was performed using the Sparse Partial Least Square Discriminant Analysis (SPLS-DA) supervised method for data classification and feature selection. The number of selected features in each dataset is given in brackets. **(B)** Pearson correlations between selected miRNAs and bio-signature of immune cells and microbes were calculated for LB and LSL at FDR ≤5%. Moreover, miRNA–mRNA negatively correlated pairs, derived from the twenty selected miRNAs and differentially expressed (DE) mRNAs in the jejunum, were subjected to functional annotation analysis. **(C)** Strain-specific bio-signatures were applied to correlation analysis and functional enrichment. Within LB and LSL, Pearson correlation was calculated between the whole set of gut mRNAs and bio-signatures of immune cells, microbes, phenotypic traits, and blood/digesta parameters at an FDR ≤5% and ($r \geq \pm 0.5$). Finally, mRNAs correlated with the different datasets were subjected to functional enrichment analysis.

the individual data blocks comprising miRNA, mRNA, immune traits, microbiome, blood/digesta parameters, and phenotypic traits showed a clear separation of groups (**Figure 2B**). Compared to other blocks, miRNA, mRNA, and immune parameters revealed the highest degree of separation between LB and LSL.

Integration and Identification of Biomarkers Specifying the Lohmann Brown and Lohmann Selected Leghorn Strains

Significant features for distinguishing LB and LSL hens were selected based on their loading weights and correlation between the two components for each block. The loadings were represented as a bar plot that indicates each selected feature's contribution on each data block: mRNA, miRNA, phenotype, immune traits, metabolome, and microbiota (**Figures 3**, **4**). Out of these sets, for miRNAs 13/20 (65%), mRNAs 6/20 (30%), immune cells 14/16 (88%), microbes 1/10 (10%), metabolites 4/16 (25%), and phenotypic traits 6/11 (55%) were most prevalent in LSL based on the two components, as shown in **Figures 3**, **4**. The results suggest that strain-discriminating immune features and miRNAs were prominently abundant in LSL, whereas microbes, blood/digesta parameters, and mRNAs were visibly more abundant in LB. The Circos plot demonstrates the correlation among the selected biomarker set and their profiling in LB and LSL (**Figure 5**).

Bio-Signature miRNAs and Their Negatively Correlated DE Target mRNAs

The co-expression clustering analysis of 20 miRNAs selected as features revealed clusters composed of 13 miRNAs and seven miRNAs in clusters 1 and 2, respectively (**Figure 6A**). Results

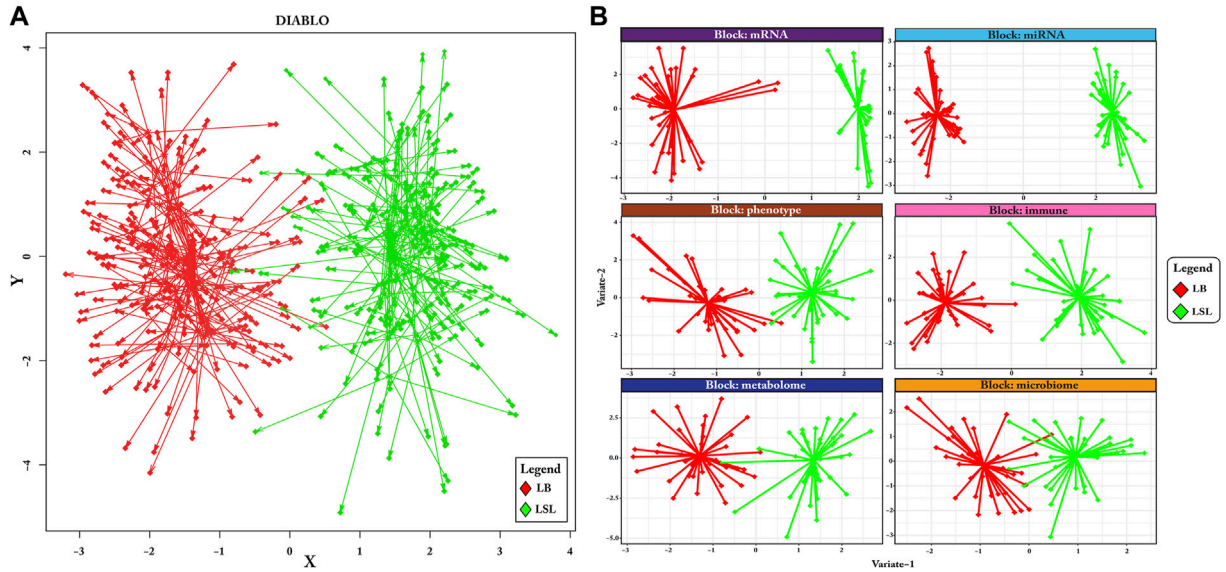


FIGURE 2 | Discriminant analysis between the two strains of laying hens, LB and LSL. **(A)** The arrow plot indicates the discrimination of LB and LSL hens as depicted as consensus components based on the integration of all datasets. The start of the arrow represents the components' score in the first dimension, while the diamond symbol on the arrow's tip gives the components' scores related to the second dimension for each individual. Short arrows indicate substantial agreement between the matching datasets, whereas long arrows depict disagreement between the matching datasets. The red and green colors represent the LB and LSL, respectively. **(B)** The individual plots depict six data blocks: miRNA, mRNA, phenotype, immune, metabolome, and microbiota, where each block sample showed differentiation between LB and LSL. The red is for LB and green is for LSL.

show that 65% (13/20) of the miRNAs were downregulated, and the remaining 35% (7/20) were upregulated in LB (**Figure 6A**). Furthermore, differential expression (DE) analysis on the complete set of mRNAs revealed 1,686 transcripts were differentially expressed between LB and LSL, with about 80% of them (1,340/1,686) downregulated and about 20% (346/1,686) of them upregulated in LB compared to LSL (**Figure 6B**).

In addition, we performed the Pearson correlation analyses between bio-signature miRNAs and their DE putative target genes, as shown in **Figure 1B**. The DE mRNAs negatively correlated with bio-signature miRNAs at FDR ≤ 5% and revealed that 13 downregulated miRNAs in LB were negatively correlated with 3,956 genes, while seven upregulated miRNAs in LB were negatively correlated with 9,153 genes as shown in **Figures 6C,D**.

Functional Enrichment Analysis of DE mRNAs Negatively Correlated With Bio-Signature miRNAs

Initially, bio-signature miRNAs and their negatively correlated target genes were subjected to ClueGO and Cluepedia for gene ontology enrichment analyses (immune system processes). Immune system processes were enriched with 30 DE genes, including T-cell activation or regulation, hemopoiesis, lymphocyte activation or proliferation, and leukocyte activation or differentiation. Interestingly, 29/30 genes were downregulated in LB but upregulated in LSL; these findings indicated that genes involved in the respective immune system processes were highly expressed in LSL compared to LB (**Figure 7A**).

Moreover, gene ontology analysis (biological processes) was performed using DAVID. We shortlisted ten biological processes related to immune response, regulation of cell proliferation, inflammatory response, and cellular processes based on 189 DE mRNAs. The majority of these genes (~93% 175/189) were upregulated in LSL and downregulated in LB. The remaining 14 upregulated genes in LB are also involved in the aforementioned biological processes, as shown in **Figure 7B**.

Moreover, KEGG pathways enrichment analysis of bio-signature miRNAs and their negatively correlated transcripts were performed (**Figure 8A**). In total, 13 KEGG pathways were found to be enriched (p-value ≤ 0.05). Most of the downregulated genes in LB and upregulated genes in LSL were enriched in immune-related pathways, including cytokine–cytokine receptor interaction, intestinal immune network for IgA production, the MAPK signaling pathway, the TGF-β signaling pathway, the FoxO signaling pathway, the Wnt signaling pathway, and others as represented in **Figure 8B**. In contrast, a lower proportion of transcripts, upregulated in LB and downregulated in LSL, were enriched in Gap junction, phagosome, the Ca signaling pathway, and cell adhesion molecule. Surprisingly, these findings also depict that transcripts enriched in immune-related KEGG pathways were more abundant in LSL than LB (**Figure 8A**).

We additionally performed correlation network analysis revealing that most immune cells were positive correlated with miRNAs in LSL including miR-7460-5p, miR146c-3p, miR-212-

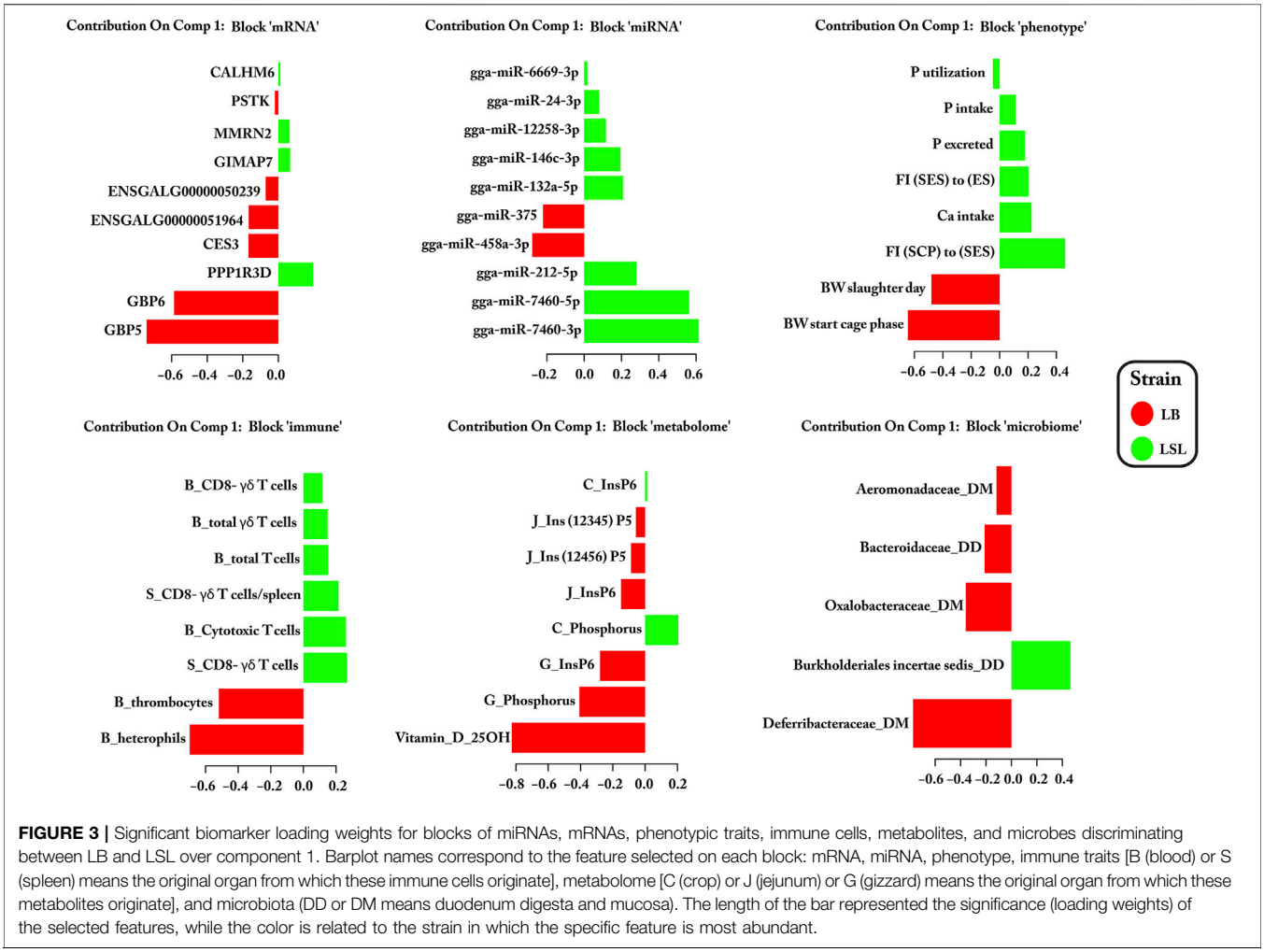


FIGURE 3 | Significant biomarker loading weights for blocks of miRNAs, mRNAs, phenotypic traits, immune cells, metabolites, and microbes discriminating between LB and LSL over component 1. Barplot names correspond to the feature selected on each block: mRNA, miRNA, phenotype, immune traits [B (blood) or S (spleen) means the original organ from which these immune cells originate], metabolome [C (crop) or J (jejunum) or G (gizzard) means the original organ from which these metabolites originate], and microbiota (DD or DM means duodenum digesta and mucosa). The length of the bar represented the significance (loading weights) of the selected features, while the color is related to the strain in which the specific feature is most abundant.

5p, and miR-24-3p, while microbes were more connected to LB (**Figure 8B**).

Strain-Specific Bio-Signature of Immune Cells and Functional Enrichment Analysis of Correlated Gut mRNAs Transcripts

Sixteen immune cell types were selected as features for LB and LSL using mixOmics. About 88% of these cell types, 14/16, were shown to be more abundant in LSL, including immune cell types in blood (B cells, $CD8^+$ T helper cells, cytotoxic T cells, total T cells, and monocytes), in the spleen ($CD8^-$ γδ T cells/g, $CD8^-$ T-helper cells/g, $CD8^+$ γδ T-cells/g, total γδ T cells/g, and $CD8^+$ γδ T cells), and both in blood and spleen (total γδ T cells and $CD8^-$ γδ T cells). The remaining 12 percent was shown to be abundant in LB: blood heterophils and thrombocytes. Subsequently, Pearson correlation was calculated between selected immune features for LB and LSL with the complete set of mRNA profiles from the jejunum (**Figure 9A**).

In LB, the transcripts correlated with blood thrombocytes and heterophils were used to predict significantly enriched biological processes (gene ontology; p-value ≤ 0.05). Immune-related biological processes were identified, including programmed cell death, regulation MAPK cascade, cell differentiation, immune system regulation, the TGF-β receptor signaling pathway, and others shown in **Figure 9B**. In LSL, the transcripts correlated with 7/14 (50%) immune cells from blood and 7/14 (50%) immune cells from the spleen were also used to examine the significantly enriched biological processes. Similarly, our results indicate that most of the transcripts correlated with immune cells in LSL were predominantly involved in immune-related biological processes, including apoptotic processes, cell cycle, immune response, regulation of immune system processes, cell proliferation, activation of the immune response, and others represented in **Figure 9C**.

In addition, KEGG pathways of genes correlated with immune cells within LB and LSL were analyzed and revealed the enrichment in immune-related and mitochondrial dysfunctional pathways, including autophagy, mitophagy, endocytosis, phagosome, apoptosis, TGF-β signaling pathway, cytokine–cytokine receptor interaction, cell cycle, and others mentioned in **Figure 9D** (p ≤ 0.05). Interestingly, the major

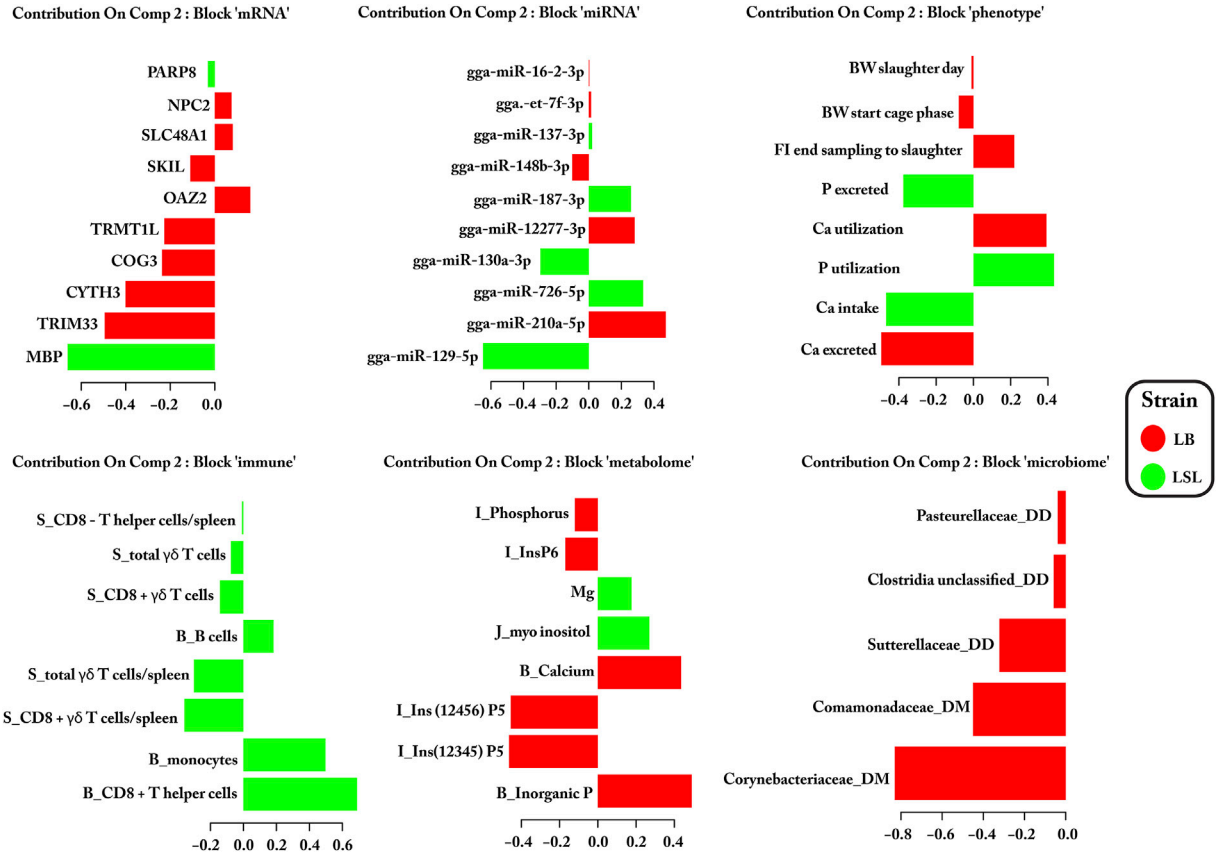


FIGURE 4 | Significant biomarker loading weights for blocks of miRNAs, mRNAs, phenotypic traits, immune cells, metabolites, and microbes discriminating between LB and LSL over component 2. Barplot names correspond to the feature selected on each block: mRNA, miRNA, phenotype, immune traits [B (blood) or S (spleen) means the original organ from which these immune cells originate], metabolome [I (ileum) or J (jejunum) or B (blood) means the original organ from which these metabolites originate], and microbiota (DD and DM means duodenum digesta and mucosa). The length of the bar represented the significance (loading weights) of the selected features, while the color is related to the strain in which the specific feature is most abundant.

portion of these enriched pathways was covered by the transcripts highlighted in the LSL strain. At the same time, the transcripts correlated with immune cells in LB were also involved in these pathways, but to a lesser extent than in LSL.

Strain-Specific Bio-Signature of Blood/Digesta Parameters and Functional Enrichment Analysis of Correlated Gut mRNAs Transcripts

Sixteen blood/digesta parameters (metabolites, P, Ca, MI, and hormones) were shortlisted as features for LB and LSL. About 76% (12/16) of these blood/digesta parameters were more abundant in LB: MI (jejunum), Ins (1,2,3,4,5) P_5 and Ins (1,2,4,5,6) P_5 (jejunum and ileum), $InsP_6$ (jejunum and gizzard), P (ileum and gizzard), and Ca, inorganic P, and vitamin-D-25OH (plasma). The remaining 14% (4/16) was shown to be more abundant in LSL: Mg (Plasma), MI (jejunum), P, and $InsP_6$ (crop). Subsequently, Pearson correlation was calculated between selected blood/digesta parameters for LB and LSL with the complete set of mRNA profiles from the jejunum (**Figure 10A**).

Based on the correlated transcripts, gene ontology (biological processes) enrichment analyses were performed. In LB, these genes were primarily enriched in metabolism, development, and immune systems, such as the P, lipid, and carbohydrate derivative metabolic processes, immune system development, inflammatory response, tissue development, growth factors, cell proliferation, and inorganic ion homeostasis (**Figure 10B**). The most enriched biological process was the P metabolic process, which included 508 correlated transcripts (p-value = 0.03).

In LSL, the results revealed that the correlated transcripts were mainly involved in metabolism and immune functions, such as cellular metabolic process, lipid metabolic process, glycoprotein metabolic process, fatty acid metabolic process, lymphocyte activation, proliferation processes, and regulation of immune response (**Figure 10C**). The predominantly enriched biological process was the cellular metabolic process containing 1,144 correlated mRNAs (p-value = 0.01).

Similarly, KEGG pathway analyses were performed with the transcripts correlated with blood/digesta parameters. The

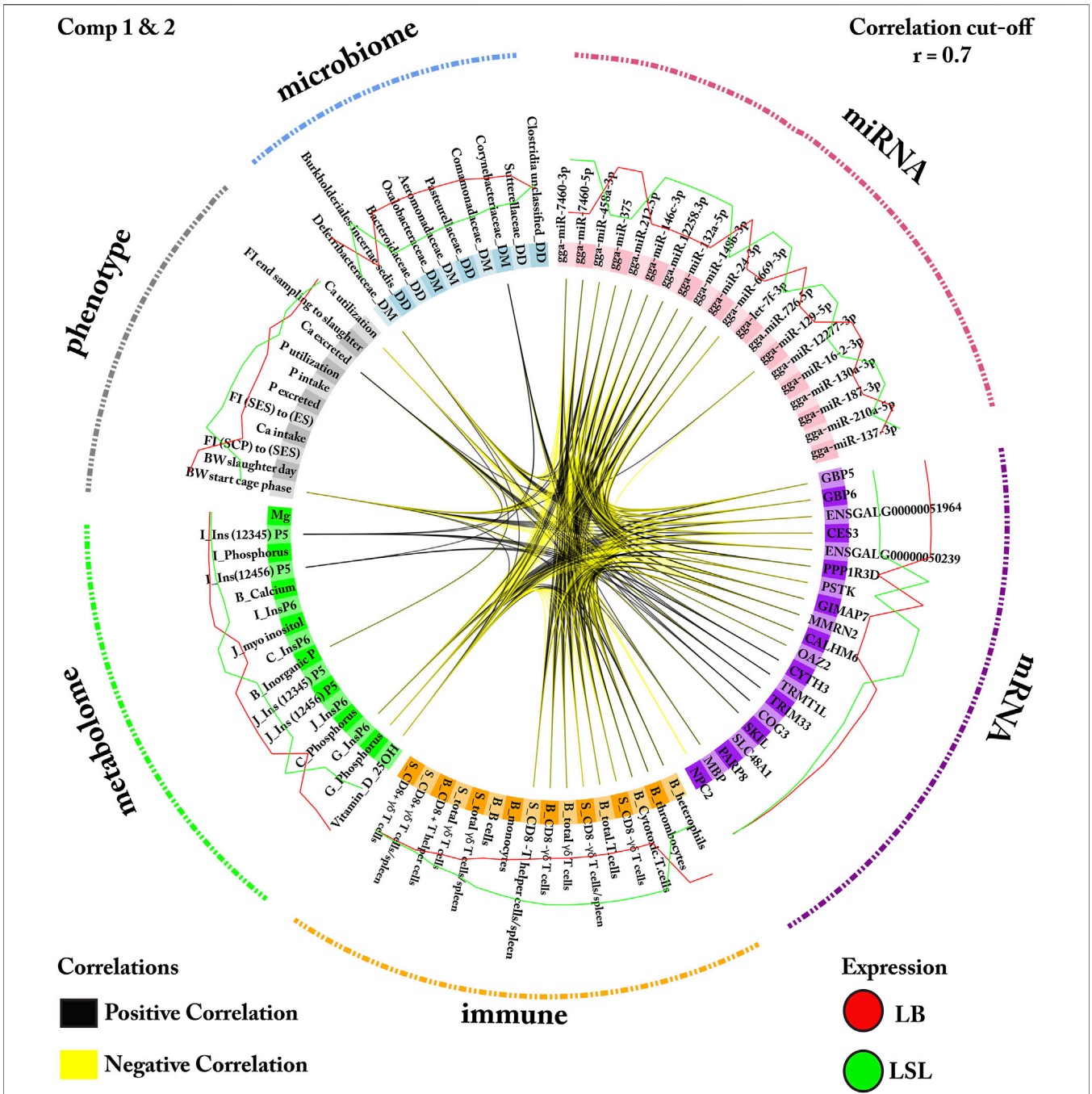


FIGURE 5 | Variable plot of miRNAs, mRNAs, phenotypic, immune traits, metabolites, and microbes related to LB and LSL strains. Circos plot demonstrates the bio-signature from multiple datasets over the two components. The selected biomarkers were represented in the inner circle. Similarly, the pink, purple, orange, green, gray, and blue dashed lines outside the Circos indicate each data type. The black link suggests a positive correlation, while the yellow link depicts a negative correlation. The red and green lines represent the feature expression level in LB and LSL, respectively.

transcripts were enriched in metabolic and immune pathways, including glycerophospholipid metabolism, inositol phosphate metabolism, glycerolipid metabolism, glutathione metabolism, fructose, and mannose metabolism, autophagy, phagosome, apoptosis, and cytokine–cytokine receptor interaction (**Figure 10D**). In addition, we found that transcripts derived from parameters more abundant in the LB strain dominated the enrichment of pathways compared to LSL, including sphingolipid metabolism, glycerophospholipid metabolism, cytokine–cytokine receptor interaction, glycolysis, and gluconeogenesis (**Figure 10D**).

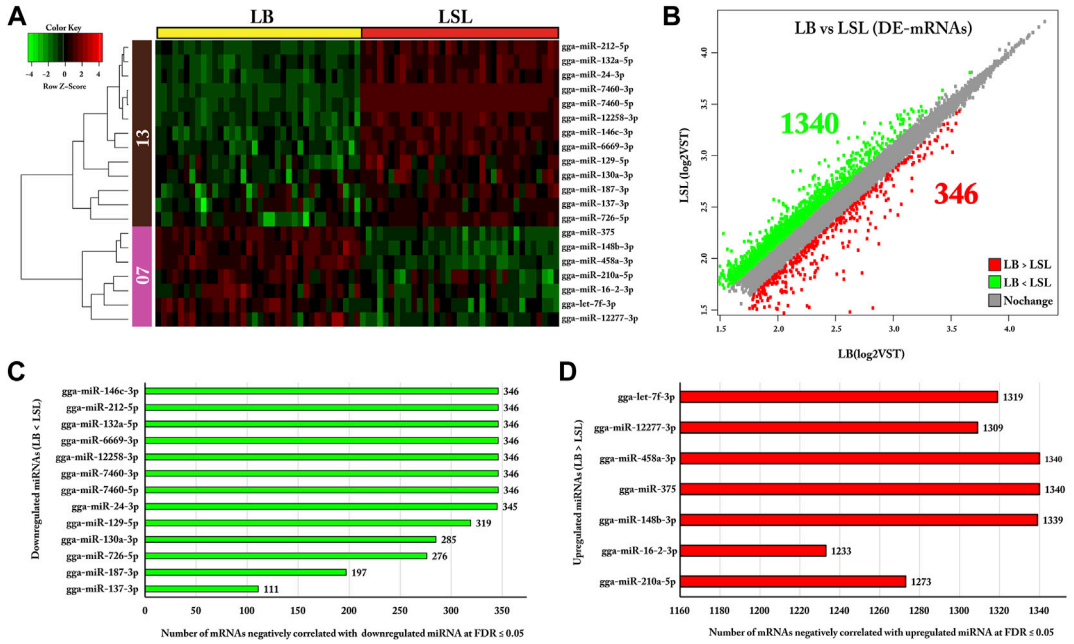


FIGURE 6 | Twenty bio-signature miRNAs, their negatively correlated DE target transcripts, and differentially expressed genes in contrast for LB vs. LSL. **(A)** Heatmap representing the LB and LSL expression profiles of 20 shortlisted miRNAs from the mixOmics. According to their co-expression, these 20 miRNAs clustered apart in two groups, 13 and seven, in clusters one and two, respectively. The green color represents downregulation, while the red indicates upregulation. **(B)** Scatter plot showing differentially expressed mRNAs between LB and LSL (cutoff |VST ≥2, logFC ≥1.2|). The green and red colors indicated the downregulated and upregulated transcripts in LB, respectively, and the gray color indicated no significant difference in expression between the strains. **(C)** Bar chart showing the downregulated miRNAs in LB and upregulated in LSL and the number of negatively correlated target transcripts at FDR ≤5%. **(D)** Bar chart representing the upregulated miRNAs in LB and downregulated in LSL and the number of negatively correlated target transcripts at FDR ≤ 5%.

Strain-Specific Bio-Signature of Microbes and Functional Enrichment Analysis of Correlated Gut mRNAs Transcripts

Ten microbes were shortlisted as features for LB and LSL from duodenal digesta/mucosa. About 90% (9/10) of these microbial taxa were more abundant in LB: Aeromonadaceae, Bacteroidaceae, Clostridia, Comamonadaceae, Corynebacteriaceae, Deferribacteraceae, Oxalobacteraceae, Pasteurellaceae, and Sutterellaceae. The remaining 10% (1/10) was abundant in LSL: Burkholderiales. Subsequently, Pearson correlation was calculated between selected microbes for LB and LSL with the complete set of mRNA profiles from the jejunum (**Figure 11A**).

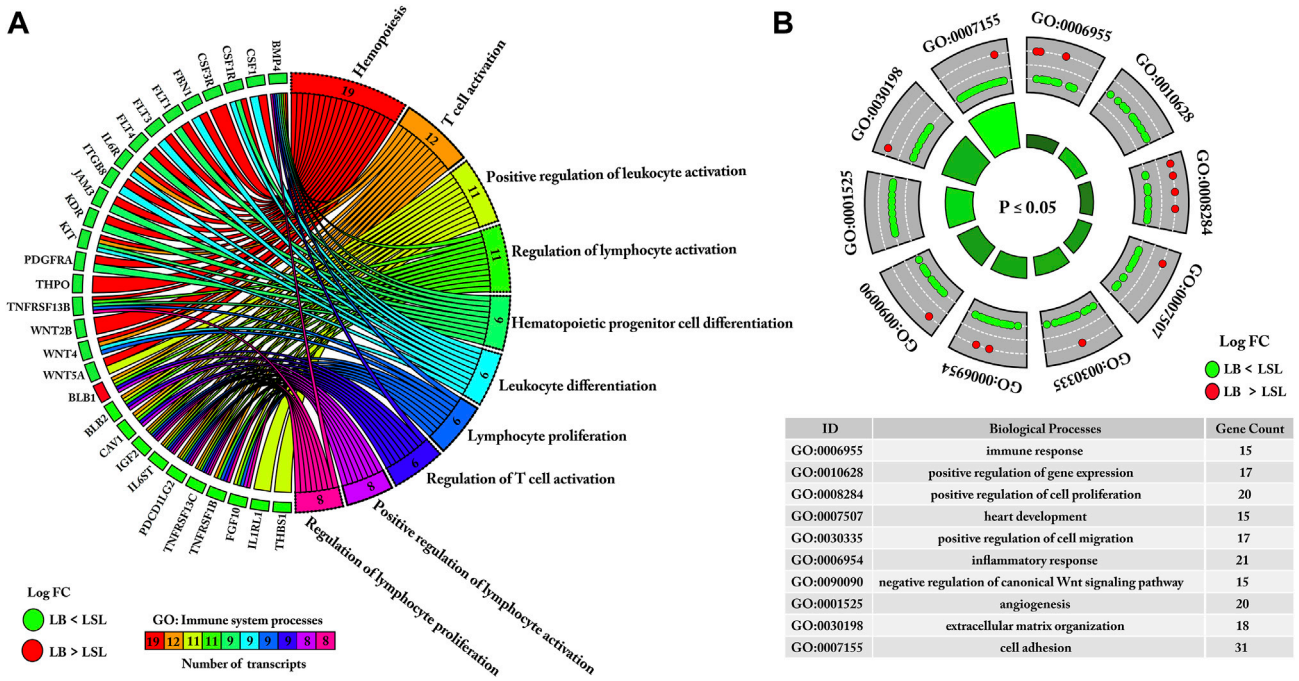
For gene ontology analysis (biological processes), the transcripts identified for LB revealed enrichment in metabolic and immune functions such as the phosphatidylinositol metabolic process, glycerophospholipid metabolic process, P metabolic process, hemostasis, T- cell differentiation, and cytokine response (**Figure 11B**). The most enriched biological process was the P metabolic process comprising 614 correlated transcripts (p-value = 0.04).

In LSL, the 554 transcripts correlated with Burkholderiales from duodenal digesta were used to investigate the function of these transcripts. The role of these 554 genes was observed in metabolic processes, including the cellular metabolic process, primary metabolic process, macromolecule metabolic process, nitrogen compound metabolic process, and nucleic acid metabolic process (**Figure 11C**).

KEGG pathway analysis was performed for the same correlated gene sets of LB and LSL. The transcripts correlated with microbes within LB were predominantly involved in immune or metabolic signaling pathways, covering Ca, MAPK, adipocytokine, Toll-like receptor, TGF-β, FoxO, and phosphatidylinositol as well as metabolic pathways, including glycerophospholipid, inositol phosphate, and glycerolipid metabolism (**Figure 11D**). In contrast, the proportion of transcripts correlated with microbes within LSL was too low for consideration.

Strain-Specific Bio-Signature of Phenotypic Traits and Functional Enrichment Analysis of Correlated Gut mRNAs Transcripts

Interestingly, all eleven phenotypic traits used as input were selected as features that differ between LB and LSL. The traits Ca intake, P utilization, P intake, P excretion, feed intake (start excreta sampling to end sampling), and feed intake (start cage phase to start excreta sampling) had higher values in LSL. The remaining 45% was higher in LB than in LSL: Feed intake (31 weeks), Ca utilization, Ca excretion, bodyweight (start cage phase), and bodyweight (31 weeks). Subsequently, Pearson correlation was calculated between selected phenotypic traits for LB and LSL with the complete set of mRNA profiles from the jejunum (**Figure 12A**).



ID	Biological Processes	Gene Count
GO:0006955	immune response	15
GO:0010628	positive regulation of gene expression	17
GO:0008284	positive regulation of cell proliferation	20
GO:0007507	heart development	15
GO:0030335	positive regulation of cell migration	17
GO:0006954	inflammatory response	21
GO:0090090	negative regulation of canonical Wnt signaling pathway	15
GO:0001525	angiogenesis	20
GO:0030198	extracellular matrix organization	18
GO:0007155	cell adhesion	31

FIGURE 7 | Gene ontology enrichment analysis of DE mRNAs negatively correlated with bio-signature miRNAs. **(A)** miRNAs and their negatively correlated DE target genes were subjected to enrichment analysis of gene ontology (GO) (immune system processes). The inner part of the Circos plot represents links of DE genes and identified GO terms. **(B)** Gene Ontology (biological processes) enrichment analysis was performed on bio-signature miRNAs and their negatively correlated target genes. The inner white dotted line represents the partition of logFC from lower to higher levels. The green is downregulated in LB and the red is upregulated in LB. Immune system processes and biological processes with $p \leq 0.05$ were considered significant.

In LB, the transcripts correlating with the bio-signature of phenotypic traits were used to predict the functions and pathways. The correlated genes are involved in metabolic and catabolic biological processes, including the glycerolipid metabolism, lipid metabolism, lipid catabolic process, and fatty acid catabolic process (**Figure 12B**). The most noticeable enriched biological process was a metabolic process comprising 1,497 correlated transcripts (p-value = 0.03).

In LSL, most of the correlated genes were involved in metabolism and the immune system, such as P metabolism, carbohydrate metabolism, glycoprotein metabolism, ATP metabolic process, ribose phosphate metabolism, and immune responses (**Figure 12C**). One of the most enriched biological processes was the P metabolic process, including 1,146 correlated transcripts (p-value = 0.02).

For KEGG pathway analysis, the proportion of correlated transcripts within LB was more enriched than LSL for mitophagy, autophagy sphingolipid metabolism, the PPAR signaling pathway, apoptosis, the Toll-like receptor signaling pathway, and the adipocytokine signaling pathway (**Figure 12D**).

DISCUSSION

Over the past decade, data integration methods have become increasingly popular due to the plethora of biological data generated from different biological experiments (Gligorijević and Pržulj, 2015). A multi-omics data integration approach can identify novel biomarkers and gain profound insight into biological mechanisms when integrating data from different experimental designs (Graw et al., 2021). In the present study, we applied a muti-omics data integration approach to analyze the miscellaneous dataset; our results established that the two strains were noticeably different regarding their immune system, transcriptional responses, metabolism, gut microbial activity, and physiological traits such as development and body weight, although both layer lines had comparable egg production performance. Moreover, mix-omics provided a shortlist of significant bio-signatures from pan-omics data comprising 20 miRNAs, 20 mRNAs, 16 immune cell types, 10 intestinal microbes, 11 phenotypic traits, and 16 blood/digesta parameters (metabolites, P, Ca, MI, and hormones). These biomarkers of the twolayer strains revealed distinct modes of adaptation in metabolic and immune pathways.

The mRNAs transcripts are the master regulator in almost every biological process and pathway (Mattick et al., 2010). In the current study, 20 mRNAs were categorized as bio-signature different between LB and LSL. 14/20 were upregulated in LB and downregulated in LSL, while 6/20 were downregulated in LB and upregulated in LSL (**Figure 3** and **Figure 4**). Guanylate-binding proteins (GBPs) are the major component of cellular immunity and are pivotal in controlling intercellular infections (Sohrabi et al., 2018). Previously, a study indicated that copy number variations of *GBP2* and *GBP4* are related to growth traits

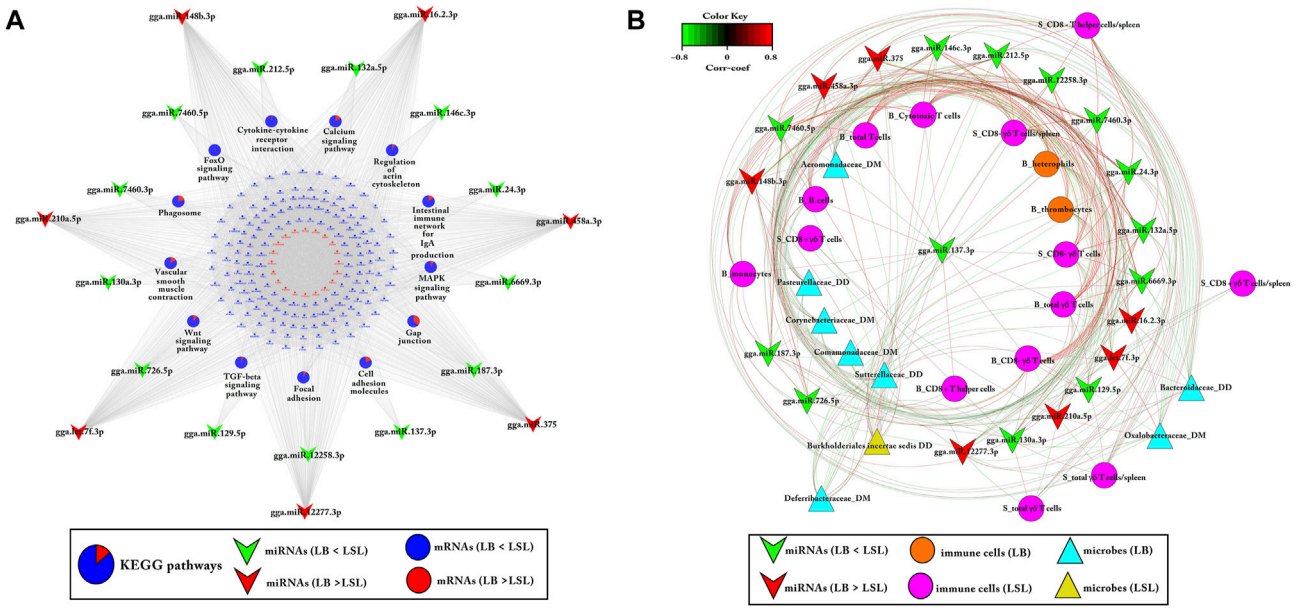


FIGURE 8 | Correlation network analysis of selected features for miRNAs, immune cells, and microbes, and KEGG pathway enrichment analysis of these miRNAs and their negatively correlated DE target transcripts. **(A)** KEGG pathway enrichment analysis of bio-signature miRNAs and their negatively correlated DE target genes ($p \leq 0.05$). The pie charts indicate the proportion of the DE gene's contribution to KEGG pathways. The blue ellipse showed that mRNAs were downregulated in LB and upregulated in LSL, while the red ellipse showed the opposite. The green symbol depicts miRNAs downregulated in LB and upregulated in LSL, whereas the red symbol indicates those upregulated in LB and downregulated in LSL. **(B)** Correlation network derived from Pearson correlations calculated between bio-signatures of miRNAs, immune cells, and microbes (FDR ≤ 5%). Circular shape showed immune cells linked to LB (orange) or LSL (pink). Likewise, the sky blue triangle indicated microbial taxa linked with LB, while the yellow triangle showed microbial taxa associated with LSL. The red and green connections indicated the positive and negative correlation, respectively.

in Chinese domestic cattle (Hao et al., 2020). An earlier study revealed that viremia levels and weight gain in response to PRRSV infection in pigs are affected by genotype-dependent alterations in *GBP5* and *GBP6* expression (Kommadath et al., 2017). Another study indicated that *GBP5* is involved in host defense, the assembly of inflammasomes, and inflammatory responses against pathogenic bacteria in *GBP5* knockout mice (Shenoy et al., 2012). Interestingly, our study also found both of these genes as highly expressed biomarkers in the LB strain. The myelin basic protein (*MBP*) and its related transcripts are widely expressed in cells of the immune system, such as T lymphocytes, B lymphocytes, and macrophages (Feng, 2007; Xu et al., 2016). Previously, a study indicated that the bone marrow and the immune system contain MBP-related transcripts and are predominantly expressed in T cells (Marty et al., 2002; Xu et al., 2016). In the present study, we identified the *MBP* gene as a highly expressed biomarker in the LSL strain.

There is considerable evidence that miRNAs are highly conserved among species and significantly regulate gene expression (Zhang et al., 2012; O'Brien et al., 2018). In the present study, we focus on 20 miRNA biomarkers and their mRNA targets that differ between LB and LSL strains. Regarding biological processes related to the immune system, transcripts inversely correlating with bio-signature miRNAs were predominantly higher in terms of expression in LSL than in LB, which was valid for 29 out of the 30 identified transcripts. The related processes include hemopoiesis, T-cell activation, lymphocyte activation, proliferation, leukocyte differentiation, and activation. The overrepresentation of transcripts assigned to immune system processes in LSL compared to LB provides additional evidence to previous studies (Habig et al., 2012, 2014; Iqbal et al., 2021; Ponsuksili et al., 2021). Stressors affected the heterophil-to-lymphocyte (H/L) ratio, which can be used to assess the level of stress imposed on laying hens (Gross and Siegel, 1983). An earlier study indicated that according to H/L ratio calculations, LB hens had ratios 2.6-fold higher than the LSL hens, and the H/Lratios of LB hens indicate prolonged stress exposure. In addition, previous studies argue that LSL hens seem to have a more adaptive immunological phenotype, while LB hens have a distinct innate immunological phenotype (Hofmann et al., 2021), and it was also shown in this study that T-cell activation, lymphocyte activation, and proliferation are higher in LSL hens. Stress influences the immune system to downregulate its responsiveness (Habig et al., 2014; Monson et al., 2018; Abbas et al., 2020; Goel et al., 2021). Accordingly, it is conceivable that higher susceptibility of LB hens to stress might be responsible for the lower abundance of transcripts related to immune function compared to LSL.

MiR-375 has been reported to be highly enriched in intestinal endocrine cells (EECs), and these cells play an essential role in systemic energy homeostasis (Hung and Sethupathy, 2018). A higher level of miR-375 was also found in the jejunal mucosa of the LB hens, which is consistent with our previous findings that it may be associated with the higher growth rate of LB compared to

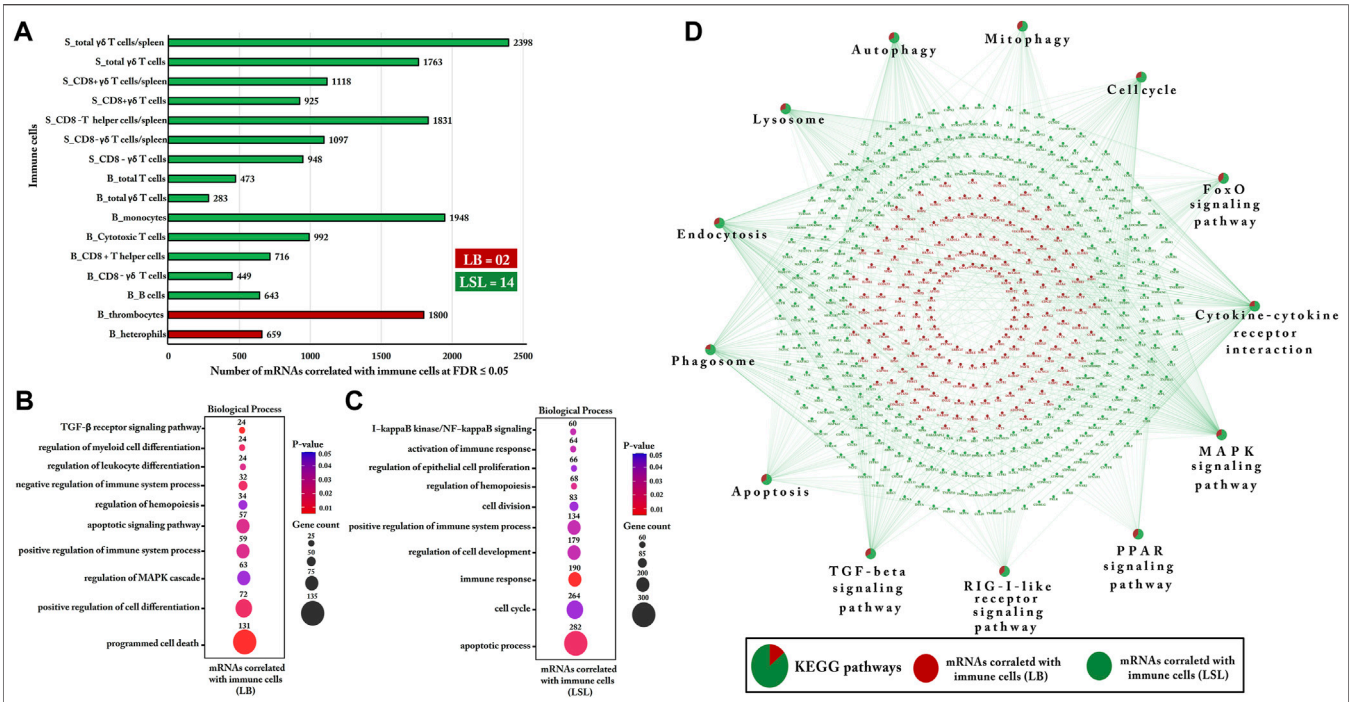


FIGURE 9 | Gene Ontology and KEGG pathways enrichment analysis of mRNAs correlated with immune cells within LB and LSL. **(A)** The bar chart indicates the number of mRNAs correlated with immune cells at FDR ≤ 5% within LB and LSL (|r ≥ 0.5|). The green bar shows the immune cells that are more abundant in LSL, while the red bar depicts immune cells that are more abundant in LB. **(B)** The transcripts correlated with immune cells in LB were subject to DAVID (version 6.8) for Gene Ontology (biological processes) enrichment analysis. **(C)** Gene Ontology (biological processes) enrichment analysis of transcripts correlated with immune cells in LSL. The dot size represents the number of transcripts involved in each biological process, while the dot's color indicates the *p*-value. **(D)** KEGG pathway enrichment analysis of mRNAs correlated with immune cells within LB and LSL. The pie charts indicate the strain-specific proportions of mRNAs correlated with immune cells to the KEGG pathways. The red ellipse shows mRNAs correlated with immune cells in LB, and the green ellipse depicts mRNAs correlated with immune cells in LSL. KEGG pathways and biological processes with $p \leq 0.05$ were considered significant.

LSL (Ponsuksili et al., 2021). Our findings show that miR-375 is less abundant in LSL than in LB and can upregulate 29 transcripts significantly associated with immunity, as shown in **Figures 7A**, **8A**. The abundance of miR-375 in the gut is also strongly negatively correlated with total T cells, total γδ T cells, and $CD8^-$ γδ T cells ($r < -0.7$; FDR < 10^{-11}) and positively correlated with thrombocytes in the blood. Recently, a study in mice reported that miR-375 might improve immune functions by regulating Kupffer cells (Ke et al., 2019). We speculate that miR-375 is likely to be a new therapeutic target for immune-mediated diseases in layer chickens. Another study revealed that miR-148b-3p was downregulated in LSL and upregulated in LB. This miRNA contributes to osteogenic differentiation and bone remodeling (Manochantr et al., 2017). For the let-7f miRNA, a recent study suggests that let-7f functions as a crucial component of the miRNA network regulating immunity (Kumar et al., 2015). Consistent with these findings, our results showed that let-7f-3p is downregulated in LSL, and its potential targets are concomitantly upregulated and govern the immune cell activation, proliferation, and differentiation processes in the gut. Furthermore, we identified bona fide gene noggin (*NOG*) that was upregulated in LB and downregulated in LSL; this gene plays an essential role in body tissue development such as muscle and bones. *NOG* regulates the TGF-β signaling pathway, which plays a significant role in bone development by stimulating osteoprogenitor enrichment (Wu et al., 2016).

Signal transduction in complex immune responses is triggered by *TNFRSF13B/TNFRSF13C* combined with *TNFSF13B* (Maeda et al., 2014). The expression of TNF Receptor Superfamily Member 13 B/C (*TNFRSF13B and TNFRSF13C*) was higher in LSL and may play an important role in humoral immunity by regulating the intestinal immune system network for the IgA production signaling pathway (**Figure 8B**).

B-cell survival and maturation are dependent on this *TNFSF13B/TNFRSF13C* system. Earlier studies showed that blocking *TNFSF13B/TNFRSF13C* signaling may effectively treat autoimmune diseases mediated by B-cells in humans (Ferrer et al., 2014). As a result, we speculate that *TNFSF13B/TNFRSF13C* are two essential components of gut immunity among laying hens. Macrophages play a crucial role in innate and acquired immune systems and are an integral component of the mononuclear phagocytic system (Elhelu, 1983). Macrophages need cytokines for their proper functioning (Arango Duque and Descoteaux, 2014). We identified two upregulated genes in LSL, colony-stimulating Factor 1 and receptor for colony-stimulating factor 1 (*CSF1* and *CSF1R*), which were downregulated in LB. The protein encoded by the *CSF1* gene is a cytokine. The *CSF1/CSF1R* system plays a pivotal role in controlling macrophages'

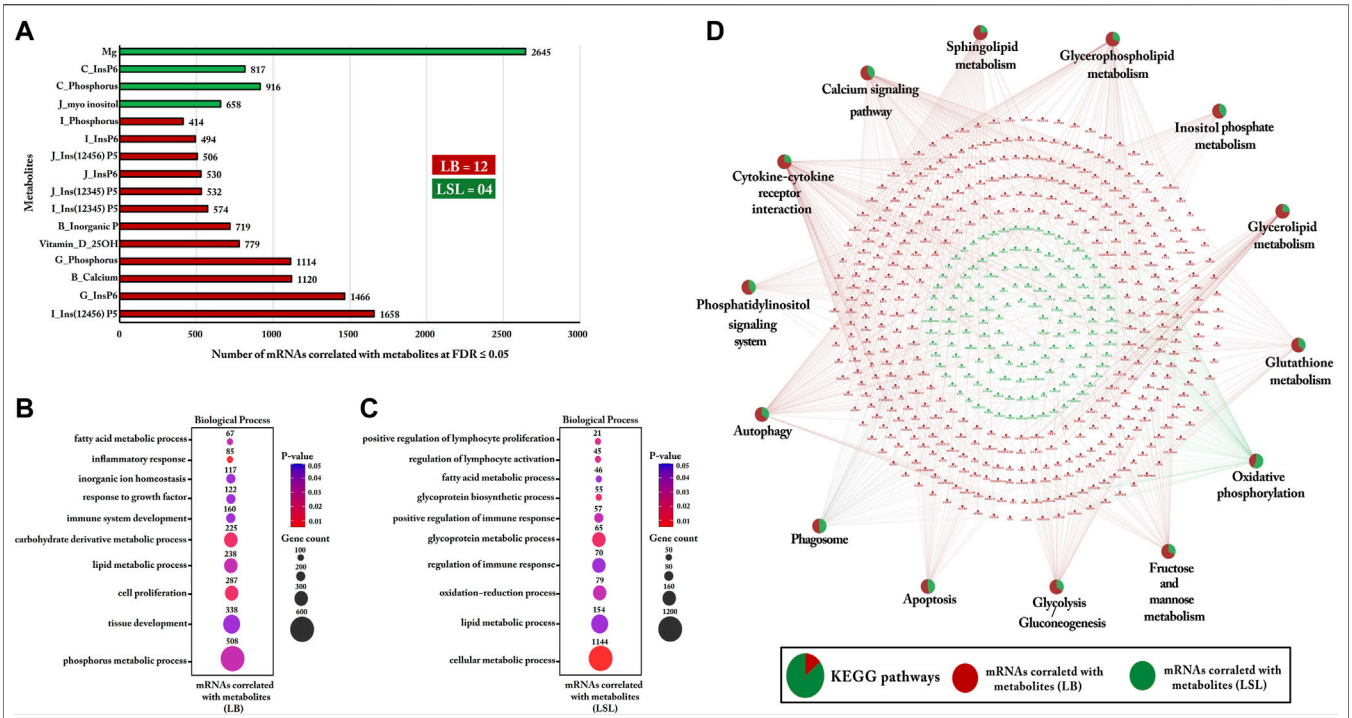


FIGURE 10 | Gene Ontology and KEGG pathways enrichment analysis of mRNAs correlated with blood/digesta parameters within LB and LSL. **(A)** The bar chart indicates the number of mRNAs correlated with blood/digesta parameters at FDR ≤ 5% within LB and LSL ($r \geq 0.5$). The green bar shows the blood/digesta parameters that are more abundant in LSL, while the red bar depicts blood/digesta parameters that are more abundant in LB. **(B)** The transcripts correlated with blood/digesta parameters in LB were subjected to DAVID (version 6.8) for Gene Ontology (biological processes) enrichment analysis. **(C)** Gene Ontology (biological processes) enrichment analysis of transcripts correlated with blood/digesta parameters in LSL. The dot size represents the number of transcripts involved in each biological process, while the dot's color indicates the *p*-value. **(D)** KEGG pathway enrichment analysis of mRNAs correlated with blood/digesta parameters within LB and LSL. The pie charts indicate the strain-specific proportions of mRNAs correlated with blood/digesta parameters to the KEGG pathways. The red ellipse shows mRNAs correlated with blood/digesta parameters in LB, and the green ellipse depicts mRNAs correlated with blood/digesta parameters in LSL. KEGG pathways and biological processes with $p \leq 0.05$ were considered significant.

development, differentiation, and function and regulates the cytokine–cytokine receptor interaction pathway. Therefore, the higher abundance of *CSF1/CSF1R* in LSL might contribute to the more pronounced acquired immunity compared with the LB strain. We also identified interleukin-cytokine receptor (*IL1R2, IL1RAP*, and *IL1RL1*) upregulation in LSL. Earlier studies have indicated that the immune response and inflammation are triggered by IL-1, secreted by macrophages, fibroblasts, B cells, granular lymphocytes, endothelium, and astrocytes (Carmi et al., 2009; Sims and Smith, 2010). Overall, significant differences were found in the miRNA- and mRNA-transcript profiles of the jejunum of the two strains. In particular, we shortlisted some essential biomarkers, which are the regulators of immune-related and developmental pathways. Given the high genetic differentiation of the two strains (Heumann-Kiesler et al., 2021), the data provide important insights into the interplay between miRNA and mRNA, which represent different strategies adopted at the molecular level to achieve optimal performance.

Previously, the genetic selection of poultry aimed to improve feed conversion and/or egg-laying performance. In addition, chickens vary in weight gain along a productive period. Recently, data have shown that the immune system varies during the hens' production period due to genetic differences (Koenen et al., 2002; Kjærup et al., 2017; Hofmann et al., 2021). One of the data sets included in the present study comprised T-lymphocytes, B-lymphocytes, heterophils, monocytes, and thrombocytes from blood, the spleen, and cecal tonsils of LB and LSL strains. Our feature selection results indicated that most of the T-lymphocytes and B-lymphocytes from the blood and spleen were more abundant in LSL, while only thrombocytes and heterophils from the blood were higher in LB than in LSL. Higher proportions of thrombocytes and heterophils have been previously detected in the blood of LB than LSL (Schmucker et al., 2021). Also, there is evidence that chicken thrombocytes may have an immunological function similar to that of mammalian platelets (St. Paul et al., 2012; Ferdous and Scott, 2015).

Compared to other cells, thrombocytes are the primary carrier of TGF-β in the body and contain 40 to 100 times more TGF-β (Karolczak and Watala, 2021), and platelet-thrombocyte number and the TGF-β concentration are positively correlated in peripheral blood (Weibrich et al., 2002; Lu et al., 2017; Guo et al., 2019). Interestingly, our results suggested that genes correlated with thrombocytes and heterophils in the LB strain were primarily involved in immune-related pathways, including TGF-β receptor signaling pathways, the apoptotic signaling

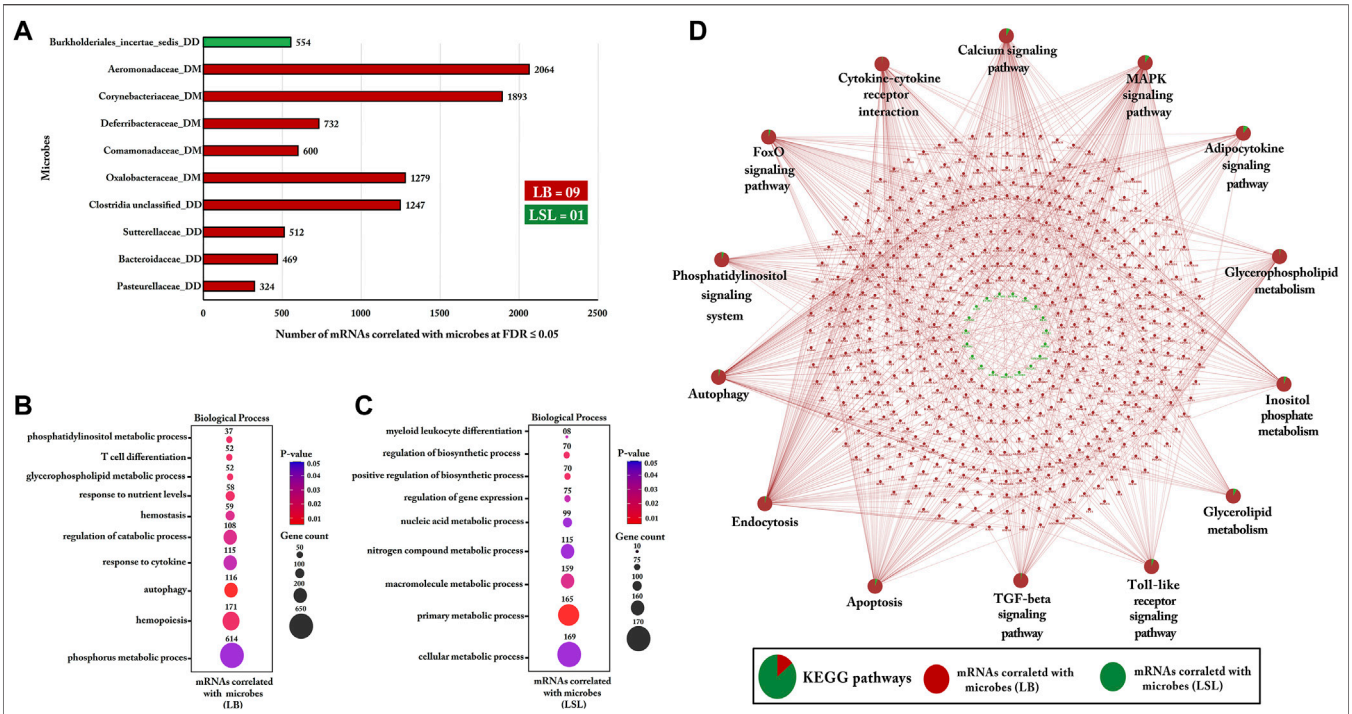


FIGURE 11 | Gene Ontology and KEGG pathways enrichment analysis of mRNAs correlated with duodenal microbiota within LB and LSL. **(A)** The bar chart indicates the number of mRNAs correlated with microbes at FDR ≤ 5% within LB and LSL ($r \geq 0.5$). The green bar shows more abundant microbes in LSL, while the red bar depicts more abundant microbes in LB. **(B)** The transcripts correlated with microbes in LB were subject to DAVID (version 6.8) for Gene Ontology (biological processes) enrichment analysis. **(C)** Gene Ontology (biological processes) enrichment analysis of transcripts correlated with microbes in LSL. The dot size represents the number of transcripts involved in each biological process, while the dot's color indicates the *p*-value. **(D)** KEGG pathway enrichment analysis of mRNAs correlated with microbes within LB and LSL. The pie charts indicate the strain-specific proportions of mRNAs correlated with microbiota to the KEGG pathways. The red ellipse shows mRNAs correlated with microbes in LB, and the green ellipse depicts mRNAs correlated with microbes in LSL. KEGG pathways and biological processes with $p \leq 0.05$ were considered significant.

pathway, regulation of MAPK cascade, and regulation of the immune system process (**Figure 9B**). We also detected a positive correlation between thrombocytes in blood with the *TGFB1*, *TGFB2*, *TGFB3*, and *TGFBR1* genes with essential roles in regulating the TGF-β signaling pathway. These findings suggest that thrombocytes may be active carriers of the TGF-β molecules. However, thrombocytes contribute to various functions within the immune system, cell differentiation, apoptosis, and cellular homeostasis, but the contribution of thrombocytes is not well documented in birds yet, thus highlighting the relevance of studying the connections between thrombocytes and TGF-β as a molecular mechanism in bird immunology. In avian species, three major subsets of lymphocytes compose the adaptive immune system: T and B lymphocytes and natural killer cells. Birds have a significantly higher number of circulating γδ T cells than humans and rodents. We detected a higher number of γδ T cells in the blood and spleen of the LSL strain than the LB strain (**Figure 9A**). We additionally identified that the genes correlated with γδ T cells significantly contributed to immune-related pathways in LSL strain, including cytokine–cytokine receptor interaction, the MAPK signaling pathway, PPAR signaling, apoptosis, autophagy, and mitophagy.

The gastrointestinal tract (GIT) harbors a complex and diverse microbiota in chicken, which plays a significant role in host health, metabolism, and immunity (Kers et al., 2018; Shang et al., 2018). Previously, studies demonstrated that gut microbiota profoundly affected chicken immune system development (Broom and Kogut, 2018; Diaz Carrasco et al., 2019). In accordance with these findings, our data suggest that genes correlated with the microbiota of mucosa and digesta were shown to be more abundant in LB than in LSL strain and were involved in the immune-related and metabolic pathways. Recently, significant correlations between cytokine gene expression (*IL-10*, *IL4*, and *IFN-γ*) and microbiota communities were observed at the early stages of chicken growth (Diaz Carrasco et al., 2019). In the intestinal tract, commensal microbes modulate cytokine production, essential for host innate and adaptive immune responses (Corthay, 2006). For instance, Clostridia was identified as a critical factor in regulating immune function (Schirmer et al., 2016). Similarly, our data from digesta highlighted that Clostridia was positively correlated with *IL20RA*, *IL22*, *IL2RB*, and *IL4R* expression in the LB strain. The *IL22* encoded protein is involved in host antimicrobial defense at the mucosal surface and is beneficial to the host intestinal inflammatory responses during infectious diseases, as shown in *IL22*−/− mice, which displayed reduced microbial diversity and slightly amplified vulnerability to host infectious diseases (Keir et al., 2020). We postulate that dynamic

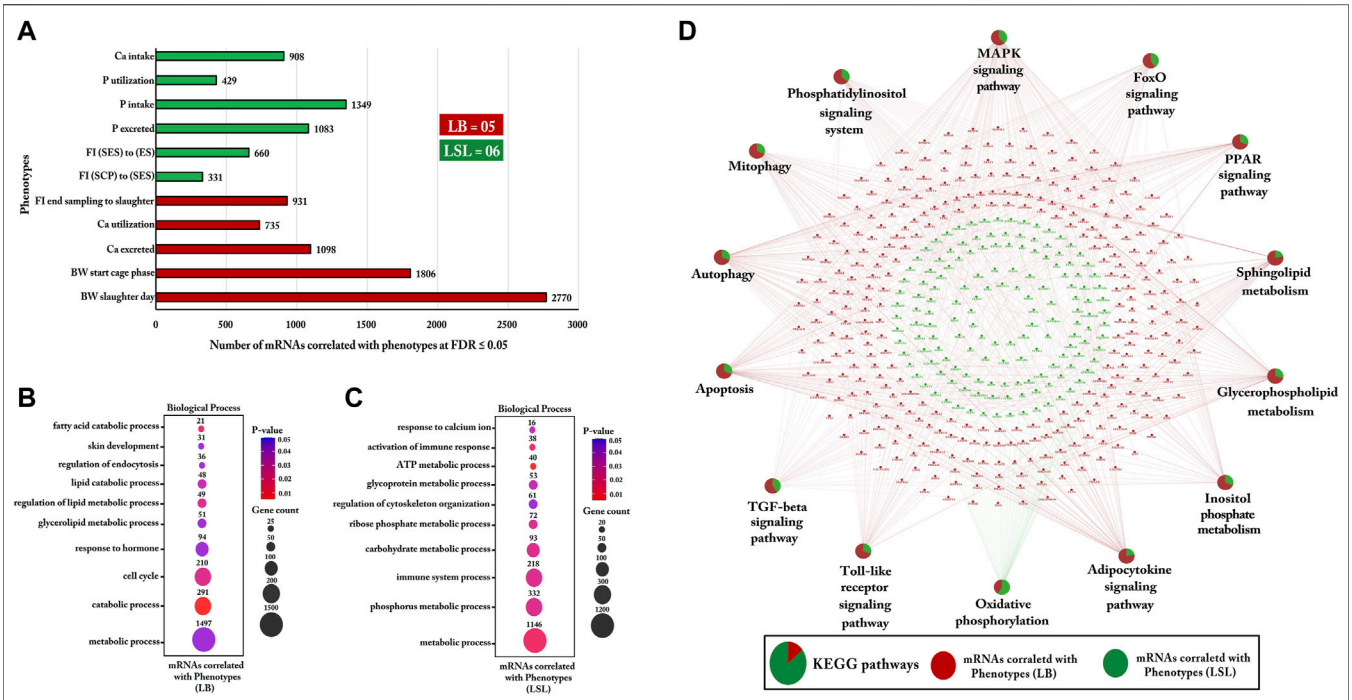


FIGURE 12 | Gene Ontology and KEGG pathways enrichment analysis of mRNAs correlated with phenotypic traits within LB and LSL. **(A)** The bar chart indicates the number of mRNAs correlated with phenotypic traits at FDR ≤5% within LB and LSL ($r \geq 0.5$). The green bar shows more abundant phenotypic traits in LSL, while the red bar depicts more abundant phenotypic traits in LB. **(B)** The transcripts correlated with phenotypic traits in LB were subjected to DAVID (version 6.8) for Gene Ontology (biological processes) enrichment analysis. **(C)** Gene Ontology (biological processes) enrichment analysis of transcripts correlated with phenotypic traits in LSL. The dot size represents the number of transcripts involved in each biological process, while the dot's color indicates the *p*-value. **(D)** KEGG pathway enrichment analysis of mRNAs correlated with phenotypic traits within LB and LSL. The pie charts indicate the strain-specific proportions of mRNAs correlated with phenotypic traits to the KEGG pathways. The red ellipse shows mRNAs correlated with phenotypic traits in LB, and the green ellipse depicts mRNAs correlated with phenotypic traits in LSL. KEGG pathways and biological processes with $p \leq 0.05$ were considered significant.

crosstalk between *IL22* and Clostridia may be relevant for achieving and maintaining the gut microbiota and host immunity balance. Studies have shown that different gut microbiota species are involved in the host's defense against harmful microorganisms, such as Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria (Gevers et al., 2012; Lloyd-Price et al., 2016).

We also identified Corynebacteriaceae (an Actinobacteria phylum member) and linked host genes enriched in autophagy and endocytosis. In addition, the autophagy-related 5 (*ATG5*) gene was positively correlated with Corynebacteriaceae and played a critical role in regulating innate and adaptive immunity. Furthermore, we identified five bacteria, Aeromonadaceae, Oxalobacteraceae, Comamonadaceae, Sutterellaceae, and Pasteurellaceae, which belong to the phylum Proteobacteria. Studies have shown that γ-proteobacteria are typical hallmarks of acute mucosal infections because of their pathogenic properties (Molloy et al., 2013; Raetz et al., 2013). Consistent with these findings, our results showed that Aeromonadaceae (a γ-proteobacteria) and genes negatively correlated with the abundance of this bacteria were involved in the Toll-like receptor signaling and apoptosis pathways. For instance, the *STAT1* transcription factor controls the responses to acute microbial infections through canonical interferon (IFN) signaling (Marié et al., 2021). We suggest a scenario where the expression of critical immune switches, for example, *STAT1*, *IFNAR1*, and *IFNAR2*, is downregulated by the pathogenic mechanisms of Aeromonadaceae and may lead to differential immune responses on the mucosal surface.

Furthermore, our results indicate that the abundance of other microbes such as Oxalobacteraceae, Comamonadaceae, Sutterellaceae, and Pasteurellaceae were correlated with immune pathways and metabolic pathways, including Ca signaling, adipocytokine signaling, Toll-like receptor signaling, FoxO signaling, glycerolipid metabolism, inositol phosphate metabolism, apoptosis, and TGF-β signaling pathway. However, these signaling and metabolic pathways are well known for their role in the immune system and metabolism. Compared to the results of the host gene expression profiles, where the immune system processes were predominantly highlighted in LSL, the duodenal microbiota component was predominantly on the side of the LB strain. This might underline the use of different strategies of these two strains but certainly highlights the importance of the gut microbiota for host immunity and metabolic activity.

Recently, a study indicated that magnesium (Mg) plays a pivotal role in energy production metabolic processes such as glycolysis, gluconeogenesis, and oxidative phosphorylation

(Pilchova et al., 2017). In line with these findings, our results revealed that genes correlated with Mg in LSL were enriched in oxidative phosphorylation, glycolysis, and gluconeogenesis pathways. For phytate metabolism, the degradation process of $InsP_6$ produces MI and lower inositol phosphates, which are involved in various immune cell functions, including proliferation, cytokine production, and cytotoxicity (Hofmann et al., 2021). Similarly, our study revealed that genes correlated with MI and $InsP_6$ in LSL were enriched in lymphocyte proliferation and activation. Several studies reported that numerous transcellular and paracellular mechanisms absorb nutrients from the GIT, including the vitamin D system as a critical player in maintaining Ca and P homeostasis in the body (Delezie et al., 2015; Sommerfeld et al., 2020; Hofmann et al., 2021; Iqbal et al., 2021; Reyer et al., 2021). Correspondingly, our results stipulate that transcripts correlated with vitamin-D-25OH were enriched in P metabolic processes, inositol phosphate metabolism, and Ca signaling pathways in the LB strain. Recently, a study reported that P homeostasis is controlled by the *ITPK1* gene in Arabidopsis (Whitfield et al., 2020). Likewise, we identified Inositol Tetrakisphosphate 1-Kinase (*ITPK1*), which was positively correlated with vitamin-D-25OH and involved in the phosphate metabolism, revealing the role of *ITPK1* in phosphate homeostasis. Interestingly, we found that genes linked to $InsP_6$, Ca, and P in the LB strain were involved in Ca signaling, phosphatidylinositol signaling, autophagy, apoptosis, cytokine–cytokine receptor interaction, sphingolipid metabolism, glycerophospholipid metabolism, inositol metabolism, and glycerolipid metabolism. These pathways are crucial in terms of immunity and metabolism and enlighten the importance of minerals, MI, and vitamin D in the chicken immune system and metabolic activity.

CONCLUSION

Together, we provided a bio-signature feature list containing 20 miRNAs, 20 mRNAs, 16 immune parameters, 10 microbes, 11 phenotypic traits, and 16 digesta/blood parameters, which discriminate between LB and LSL. This clearly shows that in addition to a zootechnical characterization, several molecular phenotypes can be inferred, providing unique strain-specific biosignatures that reliably distinguish these two contrasting high-yielding laying hen strains. Many of these strain-specific identified features were associated with many molecular pathways. We found the gut microbiota–specific LB strain to be associated with most immune-related pathways, whereas the host miRNA– and immune cell–specific LSL strain was enriched in immune-related pathways. Integration of extensive biological datasets, including deep sequencing mRNA and miRNA expression data in jejunum mucosa, immune cells, metabolites and hormones in the blood, the microbiota in both duodenum digesta and mucosa, and physiological data, revealed host–microbiota interactions and changes in immune and metabolic systems. Our results suggest that both strains implement different intrinsic approaches, shaping their immunity and metabolic activity.

To the best of our knowledge, this is the first study to simultaneously compare the two strains of laying hens, LB and LSL, in terms of their immune system, gene expression, and microbial activity in the gastrointestinal tract, metabolism, and phenotypic traits such as body weight, mineral utilization, and response to external stimuli. Even though we used a large number of laying hens per strain (LB, $n = 36$; LSL, $n = 35$), batch effects cannot be excluded here, as the animals originate from one batch each. Our results provide the basic information that contributes to the understanding of the mechanisms underlying the immune system, metabolism, and host–microbiome interaction in laying hens.

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.ebi.ac.uk/metagenomics/, E-MTAB-9136 and E-MTAB-9109.

ETHICS STATEMENT

The animal study was reviewed and approved by the Regierungspräsidium Tübingen, Germany (Project no. HOH50/17TE).

AUTHOR CONTRIBUTIONS

Conceptualization: KW and SP. Data curation: NT, SS, VS, CR and AC. Formal analysis: MI. Investigation: MO and HR. Methodology: FH, AP-S, and NT. Resources: MR and VS. Software: MI. Supervision: SP. Writing—original draft: MI. Writing—review and editing: MO, HR, NT, AP-S, SS, VS, CR, AC, KH,VS, MR, KW and SP. All authors reviewed and approved the manuscript.

FUNDING

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project numbers WI 3719/8-1, WI 1754/16-1, CA 1708/2-1, and RO 1217/10-1—as part of the research unit P-FOWL (FOR 2601).

ACKNOWLEDGMENTS

The authors thank Nicole Gentz, Annette Jugert, and Joana Bittner for their excellent technical assistance.

REFERENCES

Abbas, A. O., Alaqil, A. A., El-Beltagi, H. S., Abd El-Atty, H. K., Kamel, N. N., and Kamel, N. N. (2020). Modulating Laying Hens Productivity and Immune Performance in Response to Oxidative Stress Induced by E. Coli challenge Using Dietary Propolis Supplementation. *Antioxidants* 9 (9), 893. doi:10.3390/antiox9090893

Ahmadi, H., and Rodehutscord, M. (2012). A Meta-Analysis of Responses to Dietary Nonphytate Phosphorus and Phytase in Laying Hens. *Poult. Sci.* 91 (8), 2072–2078. doi:10.3382/ps.2012-02193

Arango Duque, G., and Descoteaux, A. (2014). Macrophage Cytokines: Involvement in Immunity and Infectious Diseases. *Front. Immunol.* 5, 491. doi:10.3389/fimmu.2014.00491

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 (8), 1091–1093. doi:10.1093/bioinformatics/btp101

Bindea, G., Galon, J., and Mlecnik, B. (2013). CluePedia Cytoscape Plugin: Pathway Insights Using Integrated Experimental and In Silico Data. *Bioinformatics* 29 (5), 661–663. doi:10.1093/bioinformatics/btt019

Borda-Molina, D., Iffland, H., Schmid, M., Müller, R., Schad, S., Seifert, J., et al. (2021). Gut Microbial Composition and Predicted Functions Are Not Associated with Feather Pecking and Antagonistic Behavior in Laying Hens. *Life* 11 (3), 235. doi:10.3390/life11030235

Broom, L. J., and Kogut, M. H. (2018). The Role of the Gut Microbiome in Shaping the Immune System of Chickens. *Vet. Immunol. Immunopathol.* 204, 44–51. doi:10.1016/j.vetimm.2018.10.002

Carmi, Y., Voronov, E., Dotan, S., Lahat, N., Rahat, M. A., Fogel, M., et al. (2009). The Role of Macrophage-Derived IL-1 in Induction and Maintenance of Angiogenesis. *J. Immunol.* 183 (7), 4705–4714. doi:10.4049/jimmunol.0901511

Corthay, A. (2006). A Three-Cell Model for Activation of Naïve T Helper Cells. *Scand. J. Immunol.* 64 (2), 93–96. doi:10.1111/j.1365-3083.2006.01782.x

Delezie, E., Bierman, K., Nollet, L., and Maertens, L. (2015). Impacts of Calcium and Phosphorus Concentration, Their Ratio, and Phytase Supplementation Level on Growth Performance, Foot Pad Lesions, and Hock Burn of Broiler Chickens. *J. Appl. Poult. Res.* 24 (2), 115–126. doi:10.3382/japr/pfv011

Diaz Carrasco, J. M., Casanova, N. A., and Fernández Miyakawa, M. E. (2019). Microbiota, Gut Health and Chicken Productivity: what Is the Connection? *Microorganisms* 7 (10), 374. doi:10.3390/microorganisms7100374

Elhelu, M. A. (1983). The Role of Macrophages in Immunology. *J. Natl. Med. Assoc.* 75 (3), 314–317.

Feng, J.-M. (2007). Minireview: Expression and Function of Golli Protein in Immune System. *Neurochem. Res.* 32 (2), 273–278. doi:10.1007/s11064-006-9164-1

Ferdous, F., and Scott, T. R. (2015). A Comparative Examination of Thrombocyte/platelet Immunity. *Immunol. Lett.* 163 (1), 32–39. doi:10.1016/j.imlet.2014.11.010

Ferrer, G., Bosch, R., Hodgson, K., Tejero, R., Roué, G., Colomer, D., et al. (2014). B Cell Activation through CD40 and IL4R Ligation Modulates the Response of Chronic Lymphocytic Leukaemia Cells to BAFF and APRIL. *Br. J. Haematol.* 164 (4), 570–578. doi:10.1111/bjh.12645

Gevers, D., Knight, R., Petrosino, J. F., Huang, K., McGuire, A. L., Birren, B. W., et al. (2012). The Human Microbiome Project: a Community Resource for the Healthy Human Microbiome. *Plos Biol.* 10 (8), e1001377. doi:10.1371/journal.pbio.1001377

Gley, K., Hadlich, F., Trakooljul, N., Haack, F., Murani, E., Gimsa, U., et al. (2021). Multi-Transcript Level Profiling Revealed Distinct mRNA, miRNA, and tRNA-Derived Fragment Bio-Signatures for Coping Behavior Linked Haplotypes in HPA Axis and Limbic System. *Front. Genet.* 12, 635794. doi:10.3389/fgene.2021.635794

Gligorijević, V., and Pržulj, N. (2015). Methods for Biological Data Integration: Perspectives and Challenges. *J. R. Soc. Interf.* 12 (112), 20150571. doi:10.1098/rsif.2015.0571

Goel, A., Ncho, C. M., and Choi, Y. H. (2021). Regulation of Gene Expression in Chickens by Heat Stress. *J. Anim. Sci. Biotechnol.* 12 (1), 11–13. doi:10.1186/s40104-020-00523-5

Gonzalez-Uarquin, F., Sommerfeld, V., Rodehutscord, M., and Huber, K. (2021). Interrelationship of Myo-Inositol Pathways with Systemic Metabolic Conditions in Two Strains of High-Performance Laying Hens during Their Productive Life Span. *Sci. Rep.* 11 (1), 4641–4714. doi:10.1038/s41598-021-84169-x

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., et al. (2021). Multi-omics Data Integration Considerations and Study Design for Biological Systems and Disease. *Mol. Omics* 17 (2), 170–185. doi:10.1039/d0mo00041h

Gross, W. B., and Siegel, H. S. (1983). Evaluation of the Heterophil/Lymphocyte Ratio as a Measure of Stress in Chickens. *Avian Dis.* 27, 972–979. doi:10.2307/1590198

Guo, Y., Cui, W., Pei, Y., and Xu, D. (2019). Platelets Promote Invasion and Induce Epithelial to Mesenchymal Transition in Ovarian Cancer Cells by TGF-β Signaling Pathway. *Gynecol. Oncol.* 153 (3), 639–650. doi:10.1016/j.ygyno.2019.02.026

Habig, C., Geffers, R., and Distl, O. (2012). Differential Gene Expression from Genome-wide Microarray Analyses Distinguishes Lohmann Selected Leghorn and Lohmann Brown Layers. *PLoS one* 7 (10), e46787. doi:10.1371/journal.pone.0046787

Habig, C., Geffers, R., and Distl, O. (2014). A Replication Study for Genome-wide Gene Expression Levels in Two Layer Lines Elucidates Differentially Expressed Genes of Pathways Involved in Bone Remodeling and Immune Responsiveness. *PloS one* 9 (6), e98350. doi:10.1371/journal.pone.0098350

Han, Z., Willer, T., Pielsticker, C., Gerzova, L., Rychlik, I., and Rautenschlein, S. (2016). Differences in Host Breed and Diet Influence Colonization by Campylobacter Jejuni and Induction of Local Immune Responses in Chicken. *Gut Pathog.* 8 (1), 56–14. doi:10.1186/s13099-016-0133-1

Hao, D., Wang, X., Thomsen, B., Kadarmideen, H. N., Wang, X., Lan, X., et al. (2020). Copy Number Variations and Expression Levels of Guanylate-Binding Protein 6 Gene Associated with Growth Traits of Chinese Cattle. *Animals* 10 (4), 566. doi:10.3390/ani10040566

Heumann-Kiesler, C., Sommerfeld, V., Iffland, H., Bennewitz, J., Rodehutscord, M., and Hasselmann, M. (2021). Insights into the Mitochondrial and Nuclear Genome Diversity of Two High Yielding Strains of Laying Hens. *Animals* 11 (3), 825. doi:10.3390/ani11030825

Hofmann, T., Schmucker, S., Sommerfeld, V., Huber, K., Rodehutscord, M., and Stefanski, V. (2021). Immunomodulatory Effects of Dietary Phosphorus and Calcium in Two Strains of Laying Hens. *Animals* 11 (1), 129. doi:10.3390/ani11010129

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., et al. (2007). DAVID Bioinformatics Resources: Expanded Annotation Database and Novel Algorithms to Better Extract Biology from Large Gene Lists. *Nucleic Acids Res.* 35 (Suppl. l_2), W169–W175. doi:10.1093/nar/gkm415

Hung, Y.-H., and Sethupathy, P. (2018). MicroRNAs in the Mammalian Gut Endocrine Lineage. *Endocrinology* 159 (2), 866–868. doi:10.1210/en.2017-03117

Iqbal, M. A., Ali, A., Hadlich, F., Oster, M., Reyer, H., Trakooljul, N., et al. (2021). Dietary Phosphorus and Calcium in Feed Affects miRNA Profiles and Their mRNA Targets in Jejunum of Two Strains of Laying Hens. *Sci. Rep.* 11 (1), 13534–13619. doi:10.1038/s41598-021-92932-3

Joat, N., Van, T. T. H., Stanley, D., Moore, R. J., and Chousalkar, K. (2021). Temporal Dynamics of Gut Microbiota in Caged Laying Hens: a Field Observation from Hatching to End of Lay. *Appl. Microbiol. Biotechnol.* 105 (11), 4719–4730. doi:10.1007/s00253-021-11333-8

Karolczak, K., and Watala, C. (2021). Blood Platelets as an Important but Underrated Circulating Source of TGFβ. *Ijms* 22 (9), 4492. doi:10.3390/ijms22094492

Kaufmann, F., Daş, G., Preisinger, R., Schmutz, M., König, S., and Gauly, M. (2011). Genetic Resistance to Natural Helminth Infections in Two Chicken Layer Lines. *Vet. Parasitol.* 176 (2-3), 250–257. doi:10.1016/j.vetpar.2010.11.007

Ke, Q. H., Chen, H. Y., He, Z. L., Lv, Z., Xu, X. F., Qian, Y. G., et al. (2019). Silencing of microRNA-375 Affects Immune Function in Mice with Liver Failure by Upregulating Astrocyte Elevated Gene-1 through Reducing Apoptosis of Kupffer Cells. *J. Cel Biochem.* 120 (1), 253–263. doi:10.1002/jcb.27338

Keir, M., Yi, Y., Lu, T., and Ghilardi, N. (2020). The Role of IL-22 in Intestinal Health and Disease. *J. Exp. Med.* 217 (3), e20192195. doi:10.1084/jem.20192195

Kers, J. G., Velkers, F. C., Fischer, E. A. J., Hermes, G. D. A., Stegeman, J. A., and Smidt, H. (2018). Host and Environmental Factors Affecting the Intestinal Microbiota in Chickens. *Front. Microbiol.* 9, 235. doi:10.3389/fmicb.2018.00235

Kidd, M. (2004). Nutritional Modulation of Immune Function in Broilers. *Poult. Sci.* 83 (4), 650–657. doi:10.1093/ps/83.4.650

Kjærup, R. B., Juul-Madsen, H. R., Norup, L. R., Sørensen, P., and Dalgaard, T. S. (2017). Comparison of Growth Performance and Immune Parameters of Three Commercial Chicken Lines Used in Organic Production. *Vet. Immunol. Immunopathol.* 187, 69–79. doi:10.1016/j.vetimm.2017.04.007

Koenen, M. E., Boonstra-Blom, A. G., and Jeurissen, S. H. (2002). Immunological Differences between Layer- and Broiler-type Chickens. *Vet. Immunol. Immunopathol.* 89 (1-2), 47–56. doi:10.1016/s0165-2427(02)00169-1

Kommadath, A., Bao, H., Choi, I., Reecy, J. M., Koltes, J. E., Fritz-Waters, E., et al. (2017). Genetic Architecture of Gene Expression Underlying Variation in Host Response to Porcine Reproductive and Respiratory Syndrome Virus Infection. *Sci. Rep.* 7 (1), 46203–46211. doi:10.1038/srep46203

Kumar, M., Sahu, S. K., Kumar, R., Subuddhi, A., Maji, R. K., Jana, K., et al. (2015). MicroRNA Let-7 Modulates the Immune Response to *Mycobacterium tuberculosis* Infection via Control of A20, an Inhibitor of the NF-κB Pathway. *Cell Host Microbe* 17 (3), 345–356. doi:10.1016/j.chom.2015.01.007

Lê Cao, K. A., Boitard, S., and Besse, P. (2011). Sparse PLS Discriminant Analysis: Biologically Relevant Feature Selection and Graphical Displays for Multiclass Problems. *BMC Bioinformatics* 12 (1), 253–317. doi:10.1186/1471-2105-12-253

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The Healthy Human Microbiome. *Genome Med.* 8 (1), 51–11. doi:10.1186/s13073-016-0307-y

Love, M., Anders, S., and Huber, W. (2014). Differential Analysis of Count Data–The DESeq2 Package. *Genome Biol.* 15 (550), 10–1186. doi:10.1186/s13059-014-0550-8

Lu, R.-B., Lee, S.-Y., Wang, T.-Y., Chang, Y.-H., Chen, P.-S., Yang, Y.-K., et al. (2017). Long-term Heroin Use Was Associated with the Downregulation of Systemic Platelets, BDNF, and TGF-β1, and it Contributed to the Disruption of Executive Function in Taiwanese Han Chinese. *Drug Alcohol Depend.* 179, 139–145. doi:10.1016/j.drugalcdep.2017.06.035

Maeda, S., Ohno, K., Fujiwara-Igarashi, A., Tomiyasu, H., Fujino, Y., and Tsujimoto, H. (2014). Methylation of TNFRSF13B and TNFRSF13C in Duodenal Mucosa in Canine Inflammatory Bowel Disease and its Association with Decreased Mucosal IgA Expression. *Vet. Immunol. Immunopathol.* 160 (1-2), 97–106. doi:10.1016/j.vetimm.2014.04.005

Manochantr, S., Marupanthorn, K., Tantrawatpan, C., Kheolamai, P., Tantikanlayaporn, D., and Sanguanjit, P. (2017). The Effects of BMP-2, miR-31, miR-106a, and miR-148a on Osteogenic Differentiation of MSCs Derived from Amnion in Comparison with MSCs Derived from the Bone Marrow. *Stem Cell Int.* 2017, 7257628. doi:10.1155/2017/7257628

Marié, I. J., Brambilla, L., Azzouz, D., Chen, Z., Baracho, G., Arnett, A., et al. (2021). Tonic Interferon Restricts Pathogenic IL-17-driven Inflammatory Disease via Balancing the Microbiome. *Elife* 10. doi:10.7554/eLife.68371

Marty, M. C., Alliot, F., Rutin, J., Fritz, R., Trisler, D., and Pessac, B. (2002). The Myelin Basic Protein Gene Is Expressed in Differentiated Blood Cell Lineages and in Hemopoietic Progenitors. *Proc. Natl. Acad. Sci. U.S.A.* 99 (13), 8856–8861. doi:10.1073/pnas.122079599

Mattick, J. S., Taft, R. J., and Faulkner, G. J. (2010). A Global View of Genomic Information - Moving beyond the Gene and the Master Regulator. *Trends Genetics* 26 (1), 21–28. doi:10.1016/j.tig.2009.11.002

Molloy, M. J., Grainger, J. R., Bouladoux, N., Hand, T. W., Koo, L. Y., Naik, S., et al. (2013). Intraluminal Containment of Commensal Outgrowth in the Gut during Infection-Induced Dysbiosis. *Cell host Microbe* 14 (3), 318–328. doi:10.1016/j.chom.2013.08.003

Monson, M. S., Van Goor, A. G., Ashwell, C. M., Persia, M. E., Rothschild, M. F., Schmidt, C. J., et al. (2018). Immunomodulatory Effects of Heat Stress and Lipopolysaccharide on the Bursal Transcriptome in Two Distinct Chicken Lines. *BMC Genomics* 19 (1), 643–715. doi:10.1186/s12864-018-5033-y

Nie, W., Wang, B., Gao, J., Guo, Y., and Wang, Z. (2018). Effects of Dietary Phosphorous Supplementation on Laying Performance, Egg Quality, Bone Health and Immune Responses of Laying Hens Challenged with *Escherichia coli* Lipopolysaccharide. *J. Anim. Sci. Biotechnol.* 9 (1), 53–11. doi:10.1186/s40104-018-0271-z

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of microRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* 9, 402. doi:10.3389/fendo.2018.00402

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., et al. (2015). Genetic and Environmental Control of Host-Gut Microbiota Interactions. *Genome Res.* 25 (10), 1558–1569. doi:10.1101/gr.194118.115

Pilchova, I., Klacanova, K., Tatarkova, Z., Kaplan, P., and Racay, P. (2017). The Involvement of Mg^{2+} in Regulation of Cellular and Mitochondrial Functions. *Oxid. Med. Cel. Longev,* 2017, 6797460. doi:10.1155/2017/6797460

Ponsuksili, S., Reyer, H., Hadlich, F., Weber, F., Trakooljul, N., Oster, M., et al. (2020). Identification of the Key Molecular Drivers of Phosphorus Utilization Based on Host miRNA-mRNA and Gut Microbiome Interactions. *Ijms* 21 (8), 2818. doi:10.3390/ijms21082818

Ponsuksili, S., Hadlich, F., Reyer, H., Oster, M., Trakooljul, N., Iqbal, M. A., et al. (2021). Genetic Background and Production Periods Shape the microRNA Profiles of the Gut in Laying Hens. *Genomics* 113 (4), 1790–1801. doi:10.1016/j.ygeno.2021.04.018

Preisinger, R. (2018). Innovative Layer Genetics to Handle Global Challenges in Egg Production. *Br. Poult. Sci.* 59 (1), 1–6. doi:10.1080/00071668.2018.1401828

Raetz, M., Hwang, S.-h., Wilhelm, C. L., Kirkland, D., Benson, A., Sturge, C. R., et al. (2013). Parasite-induced TH1 Cells and Intestinal Dysbiosis Cooperate in IFN-γ-dependent Elimination of Paneth Cells. *Nat. Immunol.* 14 (2), 136–142. doi:10.1038/ni.2508

Rehmsmeier, M., Steffen, P., Höchsmann, M., and Giegerich, R. (2004). Fast and Effective Prediction of microRNA/target Duplexes. *Rna* 10 (10), 1507–1517. doi:10.1261/rna.5248604

Reyer, H., Oster, M., Ponsuksili, S., Trakooljul, N., Omotoso, A. O., Iqbal, M. A., et al. (2021). Transcriptional Responses in Jejunum of Two Layer Chicken Strains Following Variations in Dietary Calcium and Phosphorus Levels. *BMC Genomics* 22 (1), 485–512. doi:10.1186/s12864-021-07814-9

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R Package for 'omics Feature Selection and Multiple Data Integration. *Plos Comput. Biol.* 13 (11), e1005752. doi:10.1371/journal.pcbi.1005752

Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., Lotia, S., et al. (2012). A Travel Guide to Cytoscape Plugins. *Nat. Methods* 9 (11), 1069–1076. doi:10.1038/nmeth.2212

Schirmer, M., Smeekens, S. P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E. A., et al. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* 167 (4), 1125–1136.e8. doi:10.1016/j.cell.2016.10.020

Schmucker, S., Hofmann, T., Sommerfeld, V., Huber, K., Rodehutscord, M., and Stefanski, V. (2021). Immune Parameters in Two Different Laying Hen Strains during Five Production Periods. *Poult. Sci.* 100 (11), 101408. doi:10.1016/j.psj.2021.101408

Schokker, D., Veninga, G., Vastenhouw, S. A., Bossers, A., de Bree, F. M., Kaal-Lansbergen, L. M., et al. (2015). Early Life Microbial Colonization of the Gut and Intestinal Development Differ between Genetically Divergent Broiler Lines. *BMC Genomics* 16 (1), 418–513. doi:10.1186/s12864-015-1646-6

Shang, Y., Kumar, S., Oakley, B., and Kim, W. K. (2018). Chicken Gut Microbiota: Importance and Detection Technology. *Front. Vet. Sci.* 5, 254. doi:10.3389/fvets.2018.00254

Shastak, Y., Zeller, E., Witzig, M., Schollenberger, M., and Rodehutscord, M. (2014). Effects of the Composition of the Basal Diet on the Evaluation of mineral Phosphorus Sources and Interactions with Phytate Hydrolysis in Broilers. *Poult. Sci.* 93 (10), 2548–2559. doi:10.3382/ps.2014-03961

Shenoy, A. R., Wellington, D. A., Kumar, P., Kassa, H., Booth, C. J., Cresswell, P., et al. (2012). GBP5 Promotes NLRP3 Inflammasome Assembly and Immunity in Mammals. *Science* 336 (6080), 481–485. doi:10.1126/science.1217141

Silversides, F. G., Singh, R., Cheng, K. M., and Korver, D. R. (2012). Comparison of Bones of 4 Strains of Laying Hens Kept in Conventional Cages and Floor Pens ,. *Poult. Sci.* 91 (1), 1–7. doi:10.3382/ps.2011-01453

Simon, K., Verwoolde, M. B., Zhang, J., Smidt, H., de Vries Reilingh, G., Kemp, B., et al. (2016). Long-term Effects of Early Life Microbiota Disturbance on Adaptive Immunity in Laying Hens. *Poult. Sci.* 95 (7), 1543–1554. doi:10.3382/ps/pew088

Sims, J. E., and Smith, D. E. (2010). The IL-1 Family: Regulators of Immunity. *Nat. Rev. Immunol.* 10 (2), 89–102. doi:10.1038/nri2691

Singh, R., Cheng, K. M., and Silversides, F. G. (2009). Production Performance and Egg Quality of Four Strains of Laying Hens Kept in Conventional Cages and Floor Pens. *Poult. Sci.* 88 (2), 256–264. doi:10.3382/ps.2008-00237

Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., et al. (2019). DIABLO: an Integrative Approach for Identifying Key Molecular Drivers from Multi-Omics Assays. *Bioinformatics* 35 (17), 3055–3062. doi:10.1093/bioinformatics/bty1054

Sohrabi, Y., Volkova, V., Kobets, T., Havelková, H., Krayem, I., Slapničková, M., et al. (2018). Genetic Regulation of Guanylate-Binding Proteins 2b and 5 during Leishmaniasis in Mice. *Front. Immunol.* 9, 130. doi:10.3389/fimmu.2018.00130

Sommerfeld, V., Omotoso, A., Oster, M., Reyer, H., Camarinha-Silva, A., Hasselmann, M., et al. (2020). Phytate Degradation, Transcellular mineral Transporters, and mineral Utilization by Two Strains of Laying Hens as Affected by Dietary Phosphorus and Calcium. *Animals* 10 (10), 1736. doi:10.3390/ani10101736

St. Paul, M., Paolucci, S., Barjesteh, N., Wood, R. D., Schat, K. A., and Sharif, S. (2012). Characterization of Chicken Thrombocyte Responses to Toll-like Receptor Ligands. *PLoS one* 7 (8), e43381. doi:10.1371/journal.pone.0043381

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and its Application. *Bioinform Biol. Insights* 14, 11779322219899051. doi:10.1177/1177932219899051

Swaggerty, C., Callaway, T., Kogut, M., Piva, A., and Grilli, E. (2019). Modulation of the Immune Response to Improve Health and Reduce Foodborne Pathogens in Poultry. *Microorganisms* 7 (3), 65. doi:10.3390/microorganisms7030065

Tamim, N. M., Angel, R., and Christman, M. (2004). Influence of Dietary Calcium and Phytase on Phytate Phosphorus Hydrolysis in Broiler Chickens. *Poult. Sci.* 83 (8), 1358–1367. doi:10.1093/ps/83.8.1358

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: an R Package for Visually Combining Expression Data with Functional Analysis: Fig. 1. *Bioinformatics* 31 (17), 2912–2914. doi:10.1093/bioinformatics/btv300

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). "Package 'gplots," in *Various R Programming Tools for Plotting Data*.

Weibrich, G., Kleis, W. K. G., Hafner, G., and Hitzler, W. E. (2002). Growth Factor Levels in Platelet-Rich Plasma and Correlations with Donor Age, Sex, and Platelet Count. *J. Craniomaxillofac. Surg.* 30 (2), 97–102. doi:10.1054/jcms.2002.0285

Whitfield, H., White, G., Sprigg, C., Riley, A. M., Potter, B. V. L., Hemmings, A. M., et al. (2020). An ATP-Responsive Metabolic Cassette Comprised of Inositol Tris/tetrakisphosphate Kinase 1 (ITPK1) and Inositol Pentakisphosphate 2-kinase (IPK1) Buffers Diphosphosphoinositol Phosphate Levels. *Biochem. J.* 477 (14), 2621–2638. doi:10.1042/bcj20200423

Wickham, H., Chang, W., and Wickham, M. H. (2016). "Package 'ggplot2," in *Create Elegant Data Visualisations Using the Grammar of Graphics*. 2, 1–189. Version

Wu, M., Chen, G., and Li, Y. P. (2016). TGF-β and BMP Signaling in Osteoblast, Skeletal Development, and Bone Formation, Homeostasis and Disease. *Bone Res.* 4 (1), 16009–16021. doi:10.1038/boneres.2016.9

Xu, J., Zhu, D., Shan, J., and Fan, Y. (2016). Changes of Gene Expression in T Lymphocytes Following Golli-MBP Gene RNA Interference. *Mol. Med. Rep.* 14 (5), 4575–4580. doi:10.3892/mmr.2016.5850

Zeller, E., Schollenberger, M., Witzig, M., Shastak, Y., Kühn, I., Hoelzle, L. E., et al. (2015). Interactions between Supplemented mineral Phosphorus and Phytase on Phytate Hydrolysis and Inositol Phosphates in the Small Intestine of Broilers ,. *Poult. Sci.* 94 (5), 1018–1029. doi:10.3382/ps/pev087

Zhang, L., Hou, D., Chen, X., Li, D., Zhu, L., Zhang, Y., et al. (2012). Exogenous Plant MIR168a Specifically Targets Mammalian LDLRAP1: Evidence of Cross-Kingdom Regulation by microRNA. *Cell Res.* 22 (1), 107–126. doi:10.1038/cr.2011.158

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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fgene.2022.831599

Identification of *SSTR5* Gene Polymorphisms and Their Association With Growth Traits in Hulun Buir Sheep

Xue Li[1,2], *Ning Ding*[1,2], *Zhichao Zhang*[1,2], *Dehong Tian*[1], *Buying Han*[1,2], *Dehui Liu*[1,2], *Sijia Liu*[1], *Fei Tian*[1], *Dejun Fu*[3], *Xiaoliang Song*[3] *and Kai Zhao*[1]*

[1]*Key Laboratory of Adaptation and Evolution of Plateau Biota, Qinghai Provincial Key Laboratory of Animal Ecological Genomics, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China,* [2]*University of Chinese Academy of Sciences, Beijing, China,* [3]*Inner Mongolia Daxing 'anling Agricultural Reclamation Group Co. LTD., Hulun Buir, China*

OPEN ACCESS

Edited by:
Natalia A. Zinovieva, L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

Reviewed by:
Fenghua Lyu, China Agricultural University, China
Abdulmojeed Yakubu, Nasarawa State University, Nigeria
Ke Wang, Northwest A&F University, China

***Correspondence:**
Kai Zhao
zhaokai@nwipb.cas.cn

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

Received: *08 December 2021*
Accepted: *10 March 2022*
Published: *26 April 2022*

Citation:
Li X, Ding N, Zhang Z, Tian D, Han B, Liu D, Liu S, Tian F, Fu D, Song X and Zhao K (2022) Identification of SSTR5 Gene Polymorphisms and Their Association With Growth Traits in Hulun Buir Sheep. Front. Genet. 13:831599. doi: 10.3389/fgene.2022.831599

The aim of this study was to locate *SSTR5* polymorphisms and evaluate their association with growth traits in Hulun Buir sheep. The study followed up 884 Hulun Buir sheep from birth to 16 months of age, which were born in the same pasture and the same year, and a consistent grazing management strategy was maintained. The birth weight (BRW) was recorded at birth, and body weight (BW), body height (BH), body length (BL), chest circumference (ChC), chest depth (ChD), chest width (ChW), hip width (HW), and cannon circumference (CaC) were measured at 4 and 9 months of age. BW, BH, BL, ChD, HW, and CaC were also recorded at 16 months of age. Based on the growth traits, 233 sheep were selected as experimental animals. Sanger sequencing was performed, and seven single-nucleotide polymorphisms (SNPs) were identified. Association analyses of the SNPs and the growth traits were then conducted. Seven SNPs of the *SSTR5* exhibited moderate polymorphism ($0.25 < PIC < 0.5$) and were consistent with the Hardy–Weinberg equilibrium. SNP7 (T989C, rs601836309) caused a change in amino acid sequences, while others did not cause any change. The genotypes of SNP1 (C186T, s400914340) were significantly associated with BW, ChW, and ChC at 4 months of age and with HW at 9 months of age ($p < 0.05$). These genotypes also showed extremely significant association with CaC at 4 months of age ($p < 0.01$). The genotypes of SNP7 exhibited a significant association with ChW and CaC at 4 and 9 months of age, respectively. Moreover, the genotypes of SNP3 (T384C, rs413380618)) and SNP4 (T537C, rs605867745) were significantly associated with CaC at 9 months of age ($p < 0.05$). Linkage disequilibrium was observed among the seven SNPs with five haplotypes. However, these haplotypes were not associated with growth traits at different ages. In conclusion, SNP1, SNP3, SNP4, and SNP7 may serve as molecular markers for the growth traits of Hulun Buir sheep.

Keywords: SSTR5, association, growth traits, Hulun Buir sheep, haplotypes

1 INTRODUCTION

Hulun Buir sheep is an esteemed local mutton breed in Hulun Buir, Inner Mongolia, China. This breed exhibits outstanding stress resistance, strong adaptability, stable heredity and provides high-quality, low-fat meat with a variety of amino acids. As a traditional mutton sheep breed, Hulun Buir sheep are not selected via advanced breeding methods; therefore, the breed exhibits low productivity, a slow growth rate, and a low slaughter rate. Many candidate genes have been reported to regulate metabolism and control the growth rate of domestic animals (Al-Mamun et al., 2015; Wang et al., 2015; La et al., 2019). Genetic variations in the candidate genes have been widely used as molecular markers, accelerating the breeding process and improving productivity. For example, new breeds of beef cattle with a myostatin mutation have been established; this molecular marker is also used in pig breeding (Cyranoski, 2015). Additionally, the ovine *VRTN* gene may be a new candidate gene for breeding sheep with more thoracic vertebrae (Li C et al., 2019).

In our previous experiments on the liver transcriptome of Hulun Buir sheep with different growth traits, somatostatin receptor subtype 5 (*SSTR5*) was identified as a differentially expressed gene. In the present study, we explored whether this gene would affect the growth traits of Hulun Buir sheep.

As a somatostatin (*SST*) receptor, *SSTR5* has functions inseparable from those of *SST*. *SST*, also known as growth hormone (GH)-inhibiting hormone or somatotropin release-inhibiting factor, is considered as a hypothalamic factor that inhibits the secretion of GH (Brazeau et al., 1973). In mammals, there are five somatostatin receptor subtypes (SSTR5). SST and SSTR are widely distributed in the central nervous system, pancreas, intestines, stomach, kidney, liver, pancreas, lungs, and placenta and has a variety of biological functions (Finley et al., 1981; Maecke and Reubi, 2011; Quan et al., 2020). SST employs diverse mechanisms to regulate growth, but its activity depends on the binding of G-protein-coupled somatostatin receptors (Anzola et al., 2019). In addition to exerting an inhibitory effect on GH release (Luque et al., 2006), SSTR also represses the secretion of prolactin, thyroid-stimulating hormone (TSH) (Colturi et al., 1984), stomach hormones, GH-releasing hormone (GHRH), secretin, glucagon, insulin, and SST in the pancreas (Lloyd et al., 1997). Furthermore, SSTR decreases the nutrient absorption rate in the gastrointestinal tract by inhibiting the secretion of gastrointestinal hormones and digestive enzymes (Tulassay, 1998). In addition, SSTR controls digestion and absorption rates by reducing gastrointestinal motility, gallbladder contraction, and blood flow, which negatively affects feed conversion and growth characteristics (Strowski et al., 2000).

TABLE 1 | Primer information of *SSTR5* of Hulun Buir sheep.

Primer name	Primer sequences (5–3′)	Size (bp)	Tm (°C)
E1-2	F: CCTCGGCTCAGTCGCTC R: TAGCACAGGCAGATGACCAG	761	60
E3-4	F: TGGAACACCTGCAACCTCAG R: GTCTCCTCTTCTGCTCCAGC	759	60

The nucleotide sequence of *SSTR5* is highly conserved among species; ovine *SSTR5* shares 85% sequence homology with humans and rats and 87% homology with mice. The ovine *SSTR5* transcript has four exons and three introns (ENSOARG00000014478.1) located on chromosome 24 (GenBank, Gene ID: 443,210), encoding 1,044-bp base (rs 812,728–813866) and 347-amino acid residues. As a receptor for somatostatin, SSTR5 plays an important role in many physiological processes, such as GH release, cell anti-proliferation, and regulating a variety of signal transduction pathways (Cattaneo et al., 1996; Melmed, 2003). *SSTR5* is also one of the major *SSTRs* expressed in the islets of Langerhans and plays an essential role in mediating the inhibitory effect of SST on insulin expression, secretion, and cell proliferation (Fagan et al., 1998).

Based on our previous experiments, and considering the importance of *SSTR5* in controlling the growth hormone axis and the lack of research on the effect of the *SSTR5* gene on growth traits in sheep, we conducted a molecular characterization of *SSTR5*, identified polymorphisms, and analyzed associations between different genotypes and growth traits in Hulun Buir sheep. The current study generated novel information about the genetic resources of Hulun Buir sheep, serving as a foundation for future applications of molecular markers in breeding.

2 MATERIALS AND METHODS

2.1 Experimental Animals and Growth Trait Data Acquisition

The study included 884 Hulun Buir sheep, the progeny of unrelated rams, born in the same pasture in March 2019 at the Hulun Buir Sheep Breeding Farm in Hulun Buir, Inner Mongolia, China. The growth traits of each sheep were recorded from birth to adulthood (16 months of age). During the experiment, all sheep were allowed to graze freely on natural pasture and had free access to water.

At birth, only birth weight (BRW) was recorded. Body weight (BW), body length (BL), body height (BH), chest circumference (ChC), chest depth (ChD), chest width (ChW), hip width (HW), and cannon circumference (CaC) were recorded at 4 and 9 months of age. BW, BL, BH, ChD, HW, and CaC were again recorded at 16 months of age. Size and weight were measured using a veterinary measuring tape and a sensitive platform balance, respectively (Zhang et al., 2016).

Based on the growth traits, a total of 233 healthy, disease-free sheep were selected as experimental animals, comprising 119 sheep with the fastest growth rate and 114 sheep with the slowest growth rate (124 females and 109 males). There were significant differences in growth traits between the two extreme populations (**Supplementary Table S1**). All animal experiments were conducted following the procedures described in the "Guidelines for animal care and use" manually approved by the Animal Care and Use Committee, Northwest Institute of Plateau Biology, Chinese Academy of Sciences (NWIPB2020302, 13 April 2020).

TABLE 2 | Correlations between growth traits of Hulun Buir sheep [a].

	BW	BL	BH	ChW	ChD	ChC	HW
4 months of age							
BL	**0.782****						
BH	0.537**	0.512**					
ChW	**0.736****	**0.724****	0.368**				
ChD	**0.803****	**0.729****	**0.756****	0.689**			
ChC	0.549**	0.510**	0.631**	0.454**	0.671**		
HW	0.611**	0.538**	0.653**	0.475**	**0.759****	0.521**	
CaC	0.447**	0.358**	0.331**	0.411**	0.495**	0.348**	0.456**
9 months of age							
BL	**0.757****						
BH	**0.855****	**0.727****					
ChW	0.647**	0.612**	0.568**				
ChD	0.663**	0.654**	0.605**	0.431**			
ChC	**0.863****	0.672**	**0.742****	0.620**	0.648**		
HW	**0.742****	0.561**	0.678**	0.336**	0.560**	**0.722****	
CaC	0.697**	0.545**	0.642**	0.602**	0.526**	0.565**	0.425**
16 months of age							
BL	0.479**						
BH	0.184*	0.346**					
ChD	0.343**	0.173*	0.363**				
HW	0.356**	0.381**	0.581**	0.532**			
CaC	0.498**	0.046	−0.424**	0.310**			−0.136

*BW= body weight; BL= body length; BH= body height; ChW = chest width; ChD = chest depth; ChC = chest circumference; HW= hip width; CaC = cannon circumference. Correlations with |r| > 0.7 are in bold, *p < 0.05, **p < 0.01.*

[a]*Data represent means ± SEM (n = 233).*

2.2 Primer Design and Sequencing

A 0.5-cm^2 ear tissue sample was collected from each sheep at birth for DNA extraction and preserved in 75% alcohol. DNA was purified using a DNA extraction kit (TIANGEN, Beijing, China), and the quality was evaluated by running samples on a gel. Primers were designed for all exons of the *SSTR5* using Primer3 v0.4.0 (1) (Koressaar and Remm, 2007). The *SSTR5* transcript has four exons (ENSOARG00000014478.1); one primer pair was designed to cover exons 1 to 2 and another to cover exons 3 and 4. Information about primers is presented in **Table 1**.

PCR amplifications were performed in a 30 μl reaction volume consisting of 1.0 μl of DNA, 15 μl of 2×Taq PCR Master Mix (Sangon, Shanghai, China), 1.0 μl of each primer, and double-distilled water (dH_2O) to make up the volume. Amplifications were performed using Bio-Rad S1000 thermal cyclers (Bio-Rad, Hercules, CA, United States). The thermal profile was as follows: initial denaturation at 94°C for 2 min, followed by 35 cycles at 94°C for 10 s (denaturation), 60°C for 30 s (annealing), 72°C for 60 s (elongation), with a final extension step at 72°C for 5 min. The PCR products were visualized using 1.0% agarose gel electrophoresis to determine amplicon quality and quantity. The sequencing was performed using Sanger sequencing (Agilent 3,730, United States). Sequence alignment and SNP identification were conducted via MEGA (version 5.0) (Electronics Ltd., Kuopio, Finland). DNAMAN software (version 5.2.10) (Lynnon BioSoft, Vaudreuil, Canada) was used to conduct sequence analyses.

2.3 Bioinformatics Analysis of Non-Synonymous Mutations

Protein analyses were conducted with ExPASy tools (http://expasy.org/tools/), and parameters including molecular weight, isoelectric point, instability index, aliphatic index, and grand average of hydropathicity were computed. SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/) was used to predict the presence of signal peptides. NetOGlyc 3.1 (http://www. cbs. dtu.dk/services/NetOGlyc/) and NetNGlyc 1.0 (http://www.cbs. dtu. dk/services/NetNGlyc/) were used to predict potential O– and N–glycosylation sites, respectively. NetPhos2.0 (http://www.cbs.dtu.dk/services/NetPhos/) was used to predict phosphorylation sites.

2.4 Population Genetic Analyses

Population genetic indices including allele frequency, heterozygosity (He), observed heterozygosity (Ho), effective allele numbers (Ne), and the polymorphism information content (PIC) were analyzed as previously reported (Nei and Roychoudhury, 1974). Genotypes of SNPs were tested for the Hardy–Weinberg equilibrium (HWE) (Ortega et al., 2016). Linkage disequilibrium (LD) and haplotypes analysis were conducted using Haploview (v.4.2) (Barrett et al., 2005).

2.5 Statistical Analysis

Measured traits were tested for normality by using the Shapiro–Wilk test in SPSS Statistics (V.19, IBM, Armonk, NY, United States). Pearson's correlation coefficients were calculated to determine the correlation among the following measured traits at 4, 9, and 16 months of age: BW, BL, BH, ChC, ChD, ChW, HW, and CaC. SPSS was used to perform all analyses, and values are expressed as mean ± standard error. General linear mixed models were established to examine the associations between the genotypes and individual growth traits, and statistical significance was defined at $p < 0.05$. In this model, genotype and gender were fixed factors, and their interaction was tested. If an interaction between genotype and gender was identified, the following statistical model was used:

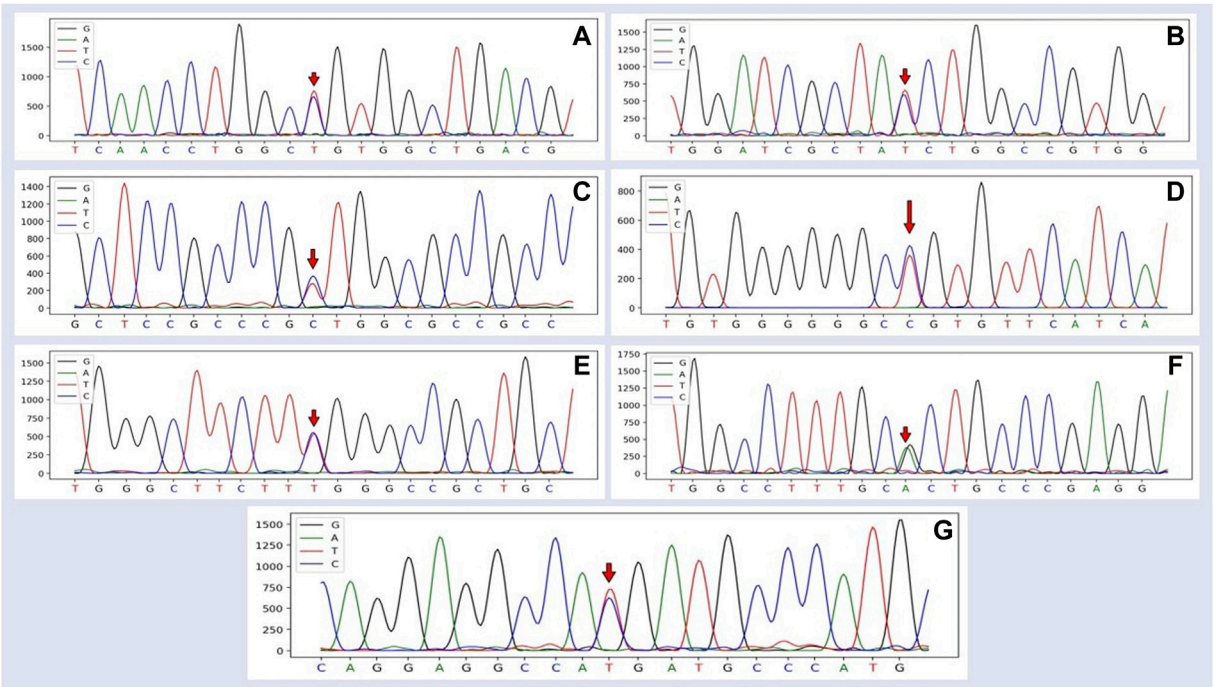


FIGURE 1 | Sequencing peak map of the *SSTR5* and the mutated SNP1–7 site of Hulun Buir sheep. The sequences were analyzed using DNAMAN software. **(A)** The site marked by the red arrow was the SNP1 mutation, which was found and identified in exon 2 of *SSTR5*, C186T (rs400944340). **(B)** The site marked by the red arrow was the SNP2 mutation, which was found and identified in exon 2 of *SSTR5*, C351T (rs404123088). **(C)**The site marked by the red arrow was the SNP3 mutation, which was found and identified in exon 2 of *SSTR5*, T384C (rs413380618). **(D)** The site marked by the red arrow was the SNP4 mutation, which was found and identified in exon 2 of *SSTR5*, T537C (rs605867745). **(E)** The site marked by the red arrow was the SNP5 mutation, which was found and identified in exon 3 of *SSTR5*, C576T (rs593868112). **(F)** The site marked by the red arrow was the SNP6 mutation, which was found and identified in exon 3 of *SSTR5*, G768A (rs403055255). **(G)** The site marked by the red arrow was the SNP7 mutation, which was found and identified in exon 4 of *SSTR5*, T989C (rs601836309).

TABLE 3 | Physicochemical properties of protein between wild type and mutant type in SNP7.

Characteristic	Wild type	Mutant type
Number of amino acids	347	347
Molecular weight (kDa)	37828.63	37858.71
Theoretical isoelectric point	9.44	9.44
Instability index	47.68	47.60
Aliphatic index	107.55	107.55
Grand average of hydropathicity	0.532	0.539
Signal peptide	0.2088	0.2088
O-glycosylation site	2	1
N-glycosylation site	3	3
Phosphorylation site	46	46

$Y = \mu + \text{Genotype} + \text{Gender} + \text{Combination} + \varepsilon$, where Y is the trait measured for each animal (BW, BL, BH, ChW, ChD, ChW, HW, and CaC), μ is the mean value of Y, Genotype is the genotype effect, Gender is the gender effect, Combination is the combined effect of the gender and genotype, and ε is the random error, assumed to be independent and normally distributed; $N(0, \sigma 2)$. If no interaction between genotype and gender was identified, the following statistical model was used: $Y = \mu + \text{Genotype} + \varepsilon$, where Y is the trait measured for each animal (BW, BL, BH, ChW, ChD, ChW, HW, and CaC), μ is the mean value of Y, Genotype is the genotype effect, and ε is the random error, assumed to be independent and normally distributed; $N(0, \sigma 2)$.

3 RESULTS

3.1 Correlations Between Growth Traits

All data conformed to a normal distribution, indicating suitability for subsequent analyses. At 4 months of age, BW exhibited strong correlations ($|r| > 0.7$) with BL, ChW, and ChD; BL exhibited strong correlations with ChW and ChD; BH exhibited a strong correlation with ChD; ChD exhibited a strong correlation with HW. At 9 months, BW exhibited strong correlations with BL, BH, ChC, and HW; BL exhibited a strong correlation with BH; BH exhibited a strong correlation with ChC; and ChC exhibited a strong correlation with HW. At 16 months, BW exhibited a negligible correlation ($|r| \leq 0.3$) with BH, BL exhibited negligible correlations with CaC and ChD, and HW exhibited a negligible correlation with CaC. Moderate correlations ($0.3 < |r| \leq 0.7$) were observed among the other traits (**Table 2**).

3.2 Polymorphism in *SSTR5*

Seven SNPs were identified by sequencing: C186T (rs400914340), C351T (rs404123088), T384C

TABLE 4 | Population genetics analyses of *SSTR5* of in Hulun Buir sheep[a]

SNP	Gene frequency		Ho	He	PIC	Ne	HW
	A	B					
SNP1 (C/T)	0.2756	0.7244	0.3993	0.3991	0.3196	1.6648	0.5888
SNP2 (C/T)	0.6197	0.3803	0.4714	0.4807	0.3603	1.8917	0.6642
SNP3 (T/C)	0.2035	0.7965	0.3241	0.3391	0.2716	1.4796	0.5052
SNP4 (T/C)	0.2137	0.7863	0.3360	0.3519	0.2796	1.5061	0.5188
SNP5 (C/T)	0.2756	0.7244	0.3993	0.3991	0.3196	1.6648	0.5888
SNP6 (G/A)	0.6923	0.3077	0.4260	0.4893	0.3353	1.7423	0.6172
SNP7 (T/C)	0.2158	0.7842	0.3385	0.3562	0.2812	1.5117	0.5216

He = heterozygosity; Ho = homozygosity; PIC= polymorphism information content; Ne = effective allele numbers; HW= Hardy–Weinberg equilibrium.
[a]*Group size of population genetics analyses was* n = *233.*

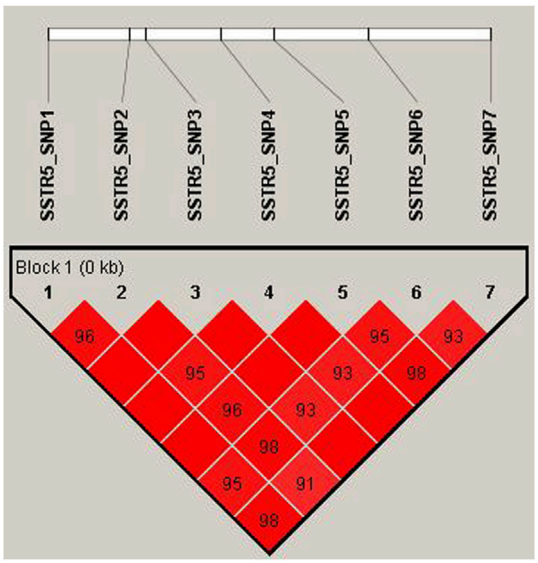


FIGURE 2 | Linkage disequilibrium plot of SNPs in *SSTR5* of in Hulun Buir sheep. The linkage disequilibrium (LD) was estimated among *SSTR5* variations in Hulun Buir sheep. The R^2 values indicated the group of the SNP1, SNP2, SNP3, SNP4, SNP5, SNP6, and SNP7.

(rs413380618), T537C (rs605867745), C576T (rs593868112), G768A (rs403055255), and T989C (rs601836309). The first four SNPs were located in exon 2, SNP5 and SNP6 in exon 3, and SNP7 in exon 4. The genetic map of the mutated sites in *SSTR5* based on the sequencing results is illustrated in **Figure 1**.

3.3 Analysis of Physicochemical Properties of Protein With Non-Synonymous Mutations

SNP1–6 were synonymous mutations. SNP7 was a non-synonymous mutation that caused the amino acid at the 330 position to change from methionine (Met) to threonine (Thr). According to analyses of the physicochemical properties of the mutant and wild-type proteins, the molecular weight of the wild-type was less than that of the mutant, and the wild-type had two O-glycosylation sites, but the mutant had only one (**Table 3**).

3.4 Population Genetics and the Linkage Disequilibrium Analysis

The Ne (effective allele numbers), calculated for each SNP, ranged from 1 to 2. The allele frequency of SNPs was in Hardy–Weinberg equilibrium ($p > 0.05$). Based on the PIC, SNP1–7 were classified as moderate polymorphic loci (**Table 4**). LD analysis revealed a strong LD (D' > 0.85) among SNP1–7 (**Figure 2**), and five common haplotypes were identified in this LD region.

3.5 Association Analysis of Genetic Variants and Haplotypes in *SSTR5* With Growth Traits of Hulun Buir Sheep

The results of association analyses of the *SSTR5* SNP genotypes and the growth traits at birth, 4, 9, and 16 months of age are shown in **Supplementary Tables S2–S4**.

3.5.1 Association Analysis of *SSTR5* With Growth Traits

The genotypes of SNP1 were significantly associated with BW, ChW, and ChC at 4 months of age and HW at 9 months of age ($p < 0.05$). They also showed extremely significant association with CaC at 4 months of age ($p < 0.01$). CaC at 9 months of age was significantly associated with the genotypes of SNP3, SNP4, and SNP7, and the genotypes of SNP7 were significantly associated with ChW at 4 months of age ($p < 0.05$, **Figure 3**). No significant differences were observed among the rest of the SNPs with other growth traits ($p > 0.05$).

3.5.2 Haplotype Association Analysis With Growth Traits

The results of association analyses between the haplotypes and growth traits at birth, 4, 9, and 16 months of age are shown in **Supplementary Table S5**. There were no detectable differences among the five haplotypes of the seven SNPs in strong LD ($p > 0.05$).

4 DISCUSSION

This is the first report of associations between SNPs of *SSTR5* and growth traits in sheep. Growth is one of the most important

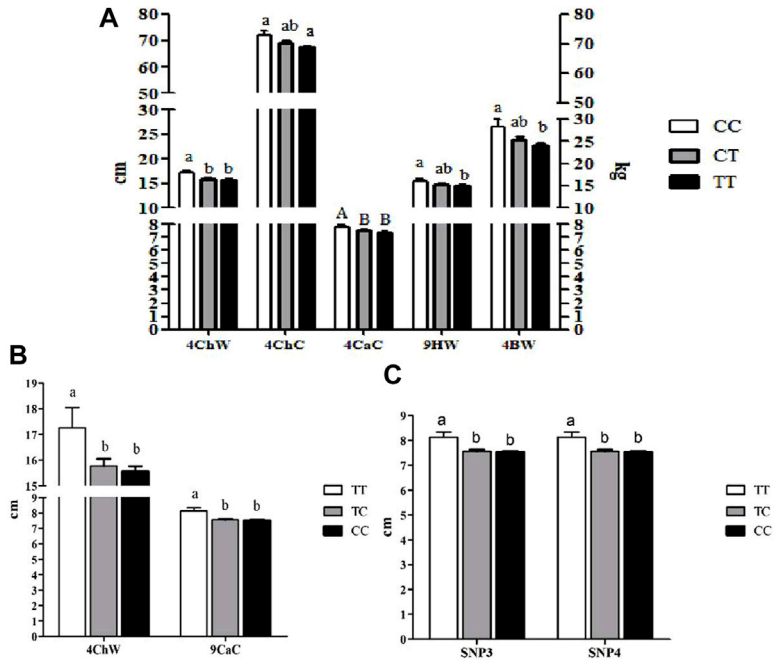


FIGURE 3 | Association analyses of SNPs in *SSTR5* with growth traits of Hulun Buir sheep[1]. **(A)** Association analyses of SNP1 in *SSTR5* with growth traits of Hulun Buir sheep; 4ChW = chest width at 4 months of age; 4ChC = chest circumference at 4 months of age; 4CaC = cannon circumference at 4 months of age; 4HW = hip width at 4 months of age; 4BW = body weight at 4 months of age. **(B)** Association analyses of SNP7 in *SSTR5* with growth traits of Hulun Buir sheep; 4ChW = chest width at 4 months of age; 9CaC = cannon circumference at 9 months of age. **(C)** Association analyses of SNP3 and SNP4 in *SSTR5* with cannon circumference at 9 months of age of Hulun Buir sheep. ^{a,b} means with different superscript letters are significantly different ($p < 0.05$). ^{A,B} means with different superscript letters are very significantly different ($p < 0.01$). [1]A group size of Population genetics analyses was $n = 233$.

economical traits monitored in domestic animals; therefore, it is a significant research topic in most genetic selection programs (Koller et al., 2020). As the main measure of growth traits, body weight and size have important impacts on the production of meat and wool (Luo, et al., 2021). In this study, we recorded these production traits from birth to adulthood in Hulun Buir sheep. We measured body weight at 4, 9, and 16 months of age, using seven body size indicators at 4 and 9 months of age, and using five body size indicators at 16 months of age. The phenotypic data contained the main growth traits of the sheep, providing a comprehensive image of growth trends in the Hulun Buir sheep. Our study was more comprehensive than most other publications and featured a longer period.

Body size indicators are important tools for accurately measuring livestock. These measurements are used to study the appearance, characteristics of the breed, and production performance. The growth rate of body size of ovine used for meat varies with age (Chao et al., 2016). In Hulun Buir sheep, we identified significant positive correlations between body size indicators and body weight, and individual size indicators affected body weight directly or indirectly. At 4 months, ChD, BL, and ChW had the strongest correlations with BW, successively. At 9 months, ChC, BH, and BL had the strongest correlation with BW, successively. At 16 months, CaC, BL, and HW had the strongest correlation with BW, successively. Moreover, with an increase in age, the correlation between body size and body weight weakened.

SNPs, defined as a substitution, insertion, or deletion of a single nucleotide, are important genetic sources for animal breeding. Gene expression and protein functions are related to the location of the SNPs in regulatory sequences or coding regions (Stevenson, 2015). Exons are protein-coding regions consisting of only 1%–2% of the genome, and the mutation rate in exons is approximately one-fifth of that in non-coding regions (Komar, 2009). However, almost 85% of reported disease-causing genes harbor mutations in their exons, which is of great significance for the study of genetic diseases (Li H et al., 2019). Thus, exome sequencing is the most efficient approach for identifying potentially functional mutations related to phenotypes in domestic animals (Li et al., 2013). Exome sequencing is also the most cost-efficient sequencing approach for conducting genome research and animal phenotyping (Guo et al., 2014), and this technology can be used to more effectively identify SNPs affecting growth traits.

Linked loci are a particular concern as there is substantial LD among causal SNPs (Koch, 2019). Studies have shown that body size is affected by the buildup of interpopulation LD among loci, caused by selection (Routtu et al., 2014). In this study, we observed LD among SNP1–7 and identified five common haplotypes in this LD region. However, there was no correlation between these haplotypes and growth traits possibly because of the small sample size or because of interactions between other genes and the SNPs in *SSTR5*.

SSTR5 reduces the activity of MAPK, which is considered a key molecule in the transduction of growth factor proliferation signals (Cattaneo et al., 1996; Cordelier et al., 1997). Thus, we speculated that SSTR5 plays an important role in reducing GH secretion (Franck et al., 2017). In addition, because these SSTR5-specific agonists are 1,000 times more powerful than other SSTR5-specific agonists in inhibiting GHRH-stimulated GH release in the primary culture of rat pituitary cells, SSTR5 is considered one of the main mediators of SST-induced inhibition of GH release (Shimon et al., 1997). Finally, *SSTR5* is involved in the regulation of ACTH release and may mediate SST-induced inhibition of insulin expression/secretion and cell proliferation by down-regulating the duodenal homolog box-1 (PDX-1) (Park et al., 2003; Zhou et al., 2012); both ACTH and PDX-1 are key hormones related to animal growth traits. Therefore, *SSTR5* may regulate the growth traits of Hulun Buir sheep. However, studies on *SSTR5* have mainly focused on tumors; there has been no prior study on the association between *SSTR5* and growth traits of livestock.

To discover the potential functional mutation related to the growth traits of Hulun Buir sheep, we conducted exome sequencing of *SSTR5* and then performed association analyses of the phenotypic records. Seven SNPs (C186T, C351T, T384C, T537C, C576T, G768A, and T989C) were all in Hardy–Weinberg equilibrium, which indicated that the population size under random mating conditions (without selection) was adequate for the experiment (Liu et al., 2018).

Synonymous mutations can affect translation dynamics and protein folding, leading to phenotypic changes (McCarthy et al., 2017; Chu and Wei, 2019). This was consistent with our findings, while SNP1, SNP3, and SNP4 did not cause synonymous mutations in amino acid sequences, SNP1 genotypes were significantly associated with BW at 4 months of age and with ChC, Chw, CaC, and HW at 9 months of age; SNP3 and SNP4 were significantly associated with CaC at 4 months of age.

These types of mutations are often deleterious (Saavedra-Rodriguez et al., 2021). In this study, SNP7 was identified as a non-synonymous mutation significantly associated with CaC at 9 months of age. The CaC of the sheep with a wild-type genotype (TT) was significantly larger than that of the sheep with two mutants (TC and CC). These results were also observed in other association analyses of SNPs and growth traits, suggesting that the SNPs identified in this study may very likely be deleterious. These findings indicate that Hulun Buir sheep may exhibit a tendency for breed deterioration; therefore, further studies on the effect of *SSTR5* on the growth traits of Hulun Buir sheep are necessary.

Analyses of the physicochemical properties of SNP7 indicated that the wild-type SSTR5 protein had two O-glycosylation sites, whereas the mutant had only one. The O-glycosylated stalk domain serves as a functional element for delivering proteins to the apical plasma membrane (Yeaman et al., 1997) and plays diverse, highly specific roles in fine-tuning protein functions (Schjoldager and Clausen, 2012). Therefore, SNP7 mutation may affect the function of the SSTR5 protein.

Association analyses of SNPs and growth traits in different ages revealed that the genotypes of SNP1, SNP3, SNP4, and SNP7 were associated with various growth indexes at 4 and 9 months of age. We inferred that the *SSTR5* might affect the early growth and development of Hulun Buir sheep. These four SNPs were significantly correlated with CaC and may serve as molecular markers to determine CaC.

5 CONCLUSION

In the current study, we used exon sequencing technology to screen *SSTR5* and discovered seven SNPs in Hulun Buir sheep. SNP1, SNP3, SNP4, and SNP7 were associated with CaC, demonstrating potential as molecular markers for the selection of CaC in Hulun Buir sheep. The genotypes of SNP1 were also associated with BW and ChC at 4 months of age, and HW at 9 months of age, indicating that SNP1 could be used as molecular markers for the selection of growth traits in Hulun Buir sheep. These molecular markers may provide a theoretical basis for improving the growth traits of Hulun Buir sheep.

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 in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Experimental Animal Welfare, Northwest Institute of Plateau Biology, Chinese Academy of Sciences.

AUTHOR CONTRIBUTIONS

XL, DT, FT and KZ provided the study concept and design. XL wrote and revised the manuscript. ND, ZZ, DL, BH, SL, DF and XS collected and analyzed the data. All authors have read and approved the final manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA (26040303-01)); "Wang Kuancheng Leading Talents Program for Industrial and Research Talents support Project" of Chinese Academy of Sciences; the Project of Ecological Grassland Animal Husbandry Engineering Laboratory, Chinese Academy of Sciences (KFJ-PTXM-007).

ACKNOWLEDGMENTS

We sincerely acknowledge the public databases available to us that we used to complete this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.831599/full#supplementary-material

REFERENCES

Al-Mamun, H. A., Kwan, P., Clark, S. A., Ferdosi, M. H., Tellam, R., and Gondro, C. (2015). Genome-wide Association Study of Body Weight in Australian Merino Sheep Reveals an Orthologous Region on OAR6 to Human and Bovine Genomic Regions Affecting Height and Weight. *Genet. Sel Evol.* 47 (1), 66. doi:10.1186/s12711-015-0142-4

Anzola, L. K., Rivera, J. N., Dierckx, R. A., Lauri, C., Valabrega, S., Galli, F., et al. (2019). Value of Somatostatin Receptor Scintigraphy with 99mTc-HYNIC-TOC in Patients with Primary Sjögren Syndrome. *Jcm* 8 (6), 763. doi:10.3390/jcm8060763

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: Analysis and Visualization of LD and Haplotype Maps. *Bioinformatics* 21 (2), 263–265. doi:10.1093/bioinformatics/bth457

Brazeau, P., Vale, W., Burgus, R., Ling, N., Butcher, M., Rivier, J., et al. (1973). Hypothalamic Polypeptide that Inhibits the Secretion of Immunoreactive Pituitary Growth Hormone. *Science* 179 (4068), 77–79. doi:10.1126/science.179.4068.77

Cattaneo, M. G., Amoroso, D., Gussoni, G., Sanguini, A. M., and Vicentini, L. M. (1996). A Somatostatin Analogue Inhibits MAP Kinase Activation and Cell Proliferation in Human Neuroblastoma and in Human Small Cell Lung Carcinoma Cell Lines. *FEBS Lett.* 397 (2-3), 164–168. doi:10.1016/s0014-5793(96)01159-3

Chao, T., Wang, G., Wang, J., Liu, Z., Ji, Z., Hou, L., et al. (2016). Identification and Classification of New Transcripts in Dorper and Small-Tailed Han Sheep Skeletal Muscle Transcriptomes. *PloS one* 11 (7), e0159638. doi:10.1371/journal.pone.0159638

Chu, D., and Wei, L. (2019). Nonsynonymous, Synonymous and Nonsense Mutations in Human Cancer-Related Genes Undergo Stronger Purifying Selections Than Expectation. *BMC cancer* 19 (1), 359. doi:10.1186/s12885-019-5572-x

Colturi, T. J., Unger, R. H., and Feldman, M. (1984). Role of Circulating Somatostatin in Regulation of Gastric Acid Secretion, Gastrin Release, and Islet Cell Function. Studies in Healthy Subjects and Duodenal Ulcer patientsStudies in Healthy Subjects and Duodenal Ulcer Patients. *J. Clin. Invest.* 74 (2), 417–423. doi:10.1172/jci111437

Cordelier, P., Esteve, J.-P., Bousquet, C., Delesque, N., O'Carroll, A.-M., Schally, A. V., et al. (1997). Characterization of the Antiproliferative Signal Mediated by the Somatostatin Receptor Subtype Sst5. *Proc. Natl. Acad. Sci.* 94 (17), 9343–9348. doi:10.1073/pnas.94.17.9343

Cyranoski, D. (2015). Super-muscly Pigs Created by Small Genetic Tweak. *Nature* 523 (7558), 13–14. doi:10.1038/523013a

Fagan, S. P., Azizzadeh, A., Moldovan, S., Ray, M. K., Adrian, T. E., Ding, X., et al. (1998). Insulin Secretion Is Inhibited by Subtype Five Somatostatin Receptor in the Mouse. *Surgery* 124 (2), 254–259. doi:10.1016/S0039-6060(98)70128-X

Finley, J. C. W., Maderdrut, J. L., Roger, L. J., and Petrusz, P. (1981). The Immunocytochemical Localization of Somatostatin-Containing Neurons in the Rat central Nervous System. *Neuroscience* 6 (11), 2173–2192. doi:10.1016/0306-4522(81)90006-3

Franck, S. E., Gatto, F., van der Lely, A. J., Janssen, J. A. M. J. L., Dallenga, A. H. G., Nagtegaal, A. P., et al. (2017). Somatostatin Receptor Expression in GH-Secreting Pituitary Adenomas Treated with Long-Acting Somatostatin Analogues in Combination with Pegvisomant. *Neuroendocrinology* 105 (1), 44–53. doi:10.1159/000448429

Guo, Y., Zhao, S., Lehmann, B. D., Sheng, Q., Shaver, T. M., Stricker, T. P., et al. (2014). Detection of Internal Exon Deletion with Exon Del. *BMC bioinformatics* 15 (1), 332. doi:10.1186/1471-2105-15-332

Keller, T. E., Mis, S. D., Jia, K. E., and Wilke, C. O. (2012). Reduced mRNA Secondary-Structure Stability Near the Start Codon Indicates Functional Genes in Prokaryotes. *Genome Biol. Evol.* 4 (2), 80–88. doi:10.1093/gbe/evr12910.1093/gbe/evr129

Koch, E. M. (2019). The Effects of Demography and Genetics on the Neutral Distribution of Quantitative Traits. *Genetics* 211 (4), 1371–1394. doi:10.1534/genetics.118.301839

Koller, D., Saiz-Rodríguez, M., Zubiaur, P., Ochoa, D., Almenara, S., Román, M., et al. (2020). The Effects of Aripiprazole and Olanzapine on Pupillary Light Reflex and its Relationship with Pharmacogenetics in a Randomized Multiple-dose Trial. *Br. J. Clin. Pharmacol.* 86 (10), 2051–2062. doi:10.1111/bcp.14300

Komar, A. (2009). *Single Nucleotide Polymorphisms : Methods and Protocols*. Humana Press. doi:10.1007/978-1-60327-411-1 Single Nucleotide Polymorphisms

Koressaar, T., and Remm, M. (2007). Enhancements and Modifications of Primer Design Program Primer3. *Bioinformatics* 23 (10), 1289–1291. doi:10.1093/bioinformatics/btm091

La, Y., Zhang, X., Li, F., Zhang, D., Li, C., Mo, F., et al. (2019). Molecular Characterization and Expression of SPP1, LAP3 and LCORL and Their Association with Growth Traits in Sheep. *Genes* 10 (8), 616. doi:10.3390/genes10080616

Li, C., Li, M., Li, X., Ni, W., Xu, Y., Yao, R., et al. (2019). Whole-Genome Resequencing Reveals Loci Associated with Thoracic Vertebrae Number in Sheep. *Front. Genet.* 10, 674. doi:10.3389/fgene.2019.00674

Li, H., Yang, H., Lv, N., Ma, C., Li, J., and Shang, Q. (2019). Whole Exome Sequencing and Methylation-specific M-ultiplex L-igation-dependent P-robe A-mplification A-pplied to I-dentify Angelman S-yndrome D-ue to P-aternal U-niparental D-isomy in T-wo U-nrelated P-atients. *Mol. Med. Rep.* 20 (2), 1178–1186. doi:10.3892/mmr.2019.10339

Li, X., Luo, R., Mo, X., Jiang, R., Kong, H., Hua, W., et al. (2013). Polymorphism of ZBTB17 Gene Is Associated with Idiopathic Dilated Cardiomyopathy: a Case Control Study in a Han Chinese Population. *Eur. J. Med. Res.* 18 (1), 10. doi:10.1186/2047-783X-18-10

Liu, B., An, T., Li, M., Yi, Z., Li, C., Sun, X., et al. (2018). The Association between Early-Onset Cardiac Events Caused by Neoadjuvant or Adjuvant Chemotherapy in Triple-Negative Breast Cancer Patients and Some Novel Autophagy-Related Polymorphisms in Their Genomic DNA: a Real-World Study. *Cancer Commun.* 38 (1), 71. doi:10.1186/s40880-018-0343-7

Lloyd, K. C., Amirmoazzami, S., Friedik, F., Chew, P., and Walsh, J. H. (1997). Somatostatin Inhibits Gastrin Release and Acid Secretion by Activating Sst2 in Dogs. *Am. J. Physiol.* 272 (6 Pt 1), G1481–G1488. doi:10.1152/ajpgi.1997.272.6.G1481

Luo, W., Zhou, Y., Wang, J., Yu, X., and Tong, J. (2021). Identifying Candidate Genes Involved in the Regulation of Early Growth Using Full-Length Transcriptome and RNA-Seq Analyses of Frontal and Parietal Bones and Vertebral Bones in Bighead Carp (Hypophthalmichthys Nobilis). *Front. Genet.* 11, 603454. doi:10.3389/fgene.2020.603454

Luque, R. M., Gahete, M. D., Hochgeschwender, U., and Kineman, R. D. (2006). Evidence that Endogenous SST Inhibits ACTH and Ghrelin Expression by Independent Pathways. *Am. J. Physiology-Endocrinology MetabolismEndocrinology Metab.* 291 (2), E395–E403. doi:10.1152/ajpendo.00038.2006

Maecke, H. R., and Reubi, J. C. (2011). Somatostatin Receptors as Targets for Nuclear Medicine Imaging and Radionuclide Treatment. *J. Nucl. Med.* 52 (6), 841–844. doi:10.2967/jnumed.110.084236

McCarthy, C., Carrea, A., and Diambra, L. (2017). Bicodon Bias Can Determine the Role of Synonymous SNPs in Human Diseases. *BMC genomics* 18 (1), 227. doi:10.1186/s12864-017-3609-6

Melmed, S. (2003). Mechanisms for Pituitary Tumorigenesis: the Plastic Pituitary. *J. Clin. Invest.* 112 (11), 1603–1618. doi:10.1172/jci20401

Nei, M., and Roychoudhury, A. K. (1974). Sampling Variances of Heterozygosity and Genetic Distance. *Genetics* 76 (2), 379–390. doi:10.1093/genetics/76.2.379

Ortega, M. S., Denicol, A. C., Cole, J. B., Null, D. J., and Hansen, P. J. (2016). Use of Single Nucleotide Polymorphisms in Candidate Genes Associated with Daughter Pregnancy Rate for Prediction of Genetic merit for Reproduction in Holstein Cows. *Anim. Genet.* 47 (3), 288–297. doi:10.1111/age.12420

Park, S., Kamegai, J., and Kineman, R. D. (2003). Role of Glucocorticoids in the Regulation of Pituitary Somatostatin Receptor Subtype (Sst1-sst5) mRNA Levels: Evidence for Direct and Somatostatin-Mediated Effects. *Neuroendocrinology* 78 (3), 163–175. doi:10.1159/000072798

Quan, F. B., Desban, L., Mirat, O., Kermarquer, M., Roussel, J., Koëth, F., et al. (2020). Somatostatin 1.1 Contributes to the Innate Exploration of Zebrafish Larva. *Sci. Rep.* 10 (1), 15235. doi:10.1038/s41598-020-72039-x

Routtu, J., Hall, M. D., Albere, B., Beisel, C., Bergeron, R. D., Chaturvedi, A., et al. (2014). An SNP-Based Second-Generation Genetic Map of Daphnia magna and its Application to QTL Analysis of Phenotypic Traits. *BMC genomics* 15 (1), 1033. doi:10.1186/1471-2164-15-1033

Saavedra-Rodriguez, K., Campbell, C. L., Lozano, S., Penilla-Navarro, P., Lopez-Solis, A., Solis-Santoyo, F., et al. (2021). Permethrin Resistance in *Aedes aegypti*: Genomic Variants that Confer Knockdown Resistance, Recovery, and Death. *Plos Genet.* 17 (6), e1009606. doi:10.1371/journal.pgen.1009606

Schjoldager, K. T.-B. G., and Clausen, H. (2012). Site-specific Protein O-Glycosylation Modulates Proprotein Processing - Deciphering Specific Functions of the Large Polypeptide GalNAc-Transferase Gene Family. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1820, 2079–2094. doi:10.1016/j.bbagen.2012.09.014

Shimon, I., Taylor, J. E., Dong, J. Z., Bitonte, R. A., Kim, S., Morgan, B., et al. (1997). Somatostatin Receptor Subtype Specificity in Human Fetal Pituitary Cultures. Differential Role of SSTR2 and SSTR5 for Growth Hormone, Thyroid-Stimulating Hormone, and Prolactin Regulation. *J. Clin. Invest.* 99 (4), 789–798. doi:10.1172/JCI119225

Stevenson, K. (2015). Genetic Diversity of *Mycobacterium avium* Subspecies Paratuberculosis and the Influence of Strain Type on Infection and Pathogenesis: a Review. *Vet. Res.* 46 (1), 64. doi:10.1186/s13567-015-0203-2

Strowski, M. Z., Parmar, R. M., Blake, A. D., and Schaeffer, J. M. (2000). Somatostatin Inhibits Insulin and Glucagon Secretion via Two Receptor Subtypes: An *In Vitro* Study of Pancreatic Islets from Somatostatin Receptor 2 Knockout Mice*. *Endocrinology* 141 (1), 111–117. doi:10.1210/endo.141.1.7263

Tulassay, Z. (1998). Somatostatin and the Gastrointestinal Tract. *Scand. J. Gastroenterol.* 33, 115–121. Supplement. doi:10.1080/003655298750026642

Wang, H., Zhang, L., Cao, J., Wu, M., Ma, X., Liu, Z., et al. (2015). Genome-Wide Specific Selection in Three Domestic Sheep Breeds. *PloS one* 10 (6), e0128688. doi:10.1371/journal.pone.0128688

Yeaman, C., Gall, A. H. L., Baldwin, A. N., Monlauzeur, L., Bivic, A. L., and Rodriguez-Boulan, E. (1997). The O-Glycosylated Stalk Domain Is Required for Apical Sorting of Neurotrophin Receptors in Polarized MDCK Cells. *J. Cel. Biol.* 139 (4), 929–940. doi:10.1083/jcb.139.4.929

Zhang, L. N., Wu, P., Xuan, C. Z., Liu, Y. Q., and Wu, J. (2016). Advance in Body Size Measurement and Conformation Appraisal for Sheep. *Trans. Chin. Soc. Agric. Eng.* 32 (S1), 190–197.

Zhou, G., Liu, S.-H., Shahi, K. M., Wang, H., Duan, X., Lin, X., et al. (2012). Negative Regulation of Pancreatic and Duodenal Homeobox-1 by Somatostatin Receptor Subtype 5. *Mol. Endocrinol. (Baltimore, Md* 26 (7), 1225–1234. doi:10.1210/me.2012-1095

Conflict of Interest: DF and XS were employed by the Inner Mongolia Daxing 'anling Agricultural Reclamation Group Co. LTD. They only help with sample collection and have no business cooperation or conflict of interest with their 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.

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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fgene.2022.849399

Genome-Wide Identification and Characterization of Long Non-Coding RNAs in Longissimus dorsi Skeletal Muscle of Shandong Black Cattle and Luxi Cattle

Ruili Liu[1†], *Mingxuan Han*[1†], *Xianxun Liu*[1], *Kun Yu*[1], *Xuejin Bai*[1,2] *and Yajuan Dong*[1,2*]

[1]College of Animal Science and Technology, Qingdao Agricultural University, Qingdao, China, [2]Laboratory of Animal Molecular Shandong Black Cattle Breeding Engineering Technology Center, College of Animal Science, Qingdao Agricultural University, Qingdao, China

OPEN ACCESS

Edited by:
Natalia A Zinovieva,
L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

Reviewed by:
Juliana Afonso,
Animal Biotechnology Laboratory. Embrapa Pecuária Sudeste, Brazil
Ali Esmailizadeh,
Shahid Bahonar University of Kerman, Iran

***Correspondence:**
Yajuan Dong
etcenter@126.com

[†]These authors have contributed equally to this work

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

Received: 06 January 2022
Accepted: 08 April 2022
Published: 16 May 2022

Citation:
Liu R, Han M, Liu X, Yu K, Bai X and Dong Y (2022) Genome-Wide Identification and Characterization of Long Non-Coding RNAs in Longissimus dorsi Skeletal Muscle of Shandong Black Cattle and Luxi Cattle. Front. Genet. 13:849399. doi: 10.3389/fgene.2022.849399

There is an increasing understanding of the possible regulatory role of long non-coding RNAs (LncRNA). Studies on livestock have mainly focused on the regulation of cell differentiation, fat synthesis, and embryonic development. However, there has been little study of skeletal muscle of domestic animals and the potential role of lncRNA. In this study, the transcriptome numbers of longissimus muscle of different beef cattle (Shandong black catle and Luxi catlle) were used to construct muscle related lncRNAs-miRNA-mRNA interaction network through bioinformatics analysis. This is helpful to clarify the molecular mechanism of bovine muscle development, and can be used to promote animal husbandry and improve animal husbandry production. According to the screening criteria of $|FC| \geqq 2$ and $q < 0.05$, a total of 1,415 transcripts (of which 480 were LncRNAs) were differentially expressed ($q < 0.05$) in the different breeds. Further, we found that the most differentially expressed LncRNAs were found on chromosome 9, in which the differentially expressed LncRNAs targeted 1,164 protein coding genes (*MYORG*, *Wnt4*, *PAK1*, *ADCY7,etc*) (upstream and downstream<50 Kb). In addition, Pearson's correlation coefficients of co-expression levels indicated a potential trans regulatory relationship between the differentially expressed LncRNAs and 43844 mRNAs ($r > 0.9$). The identified co-expressed mRNAs (*MYORG*, *Dll1*, *EFNB2*, *SOX6*, *MYOCD*, and *MYLK*3) are related to the formation of muscle structure, and enriched in muscle system process, strained muscle cell differentiation, muscle cell development, striated muscle tissue development, calcium signaling, and AMPK signaling. Additionally, we also found that some LncRNAs (*LOC112444238*, *LOC101903367*, *LOC104975788*, *LOC112441863*, *LOC112449549*, and *LOC101907194*) may interact with miRNAs related to cattle muscle growth and development. Based on this, we constructed a LncRNAs-miRNA-mRNA interaction network as the putative basis for biological regulation in cattle skeletal muscle.

Abbreviations: DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, The Gene Ontology; LncRNA, Long noncoding RNA; miRNA, MicroRNA (miRNA) is a class of noncoding single stranded RNA molecules with a length of about 22 nucleotides encoded by endogenous genes; mRNA, Messenger RNA (mRNA) that encodes proteins.

Interestingly, a candidate differential LncRNA (*LOC104975788*) and a protein-coding gene (*Pax7*) contain miR-133a binding sites and binding was confirmed by luciferase reporter assay. *LOC104975788* may combined miR-133a competitively with *Pax7*, thus relieving the inhibitory effect of miR-133a on *Pax7* to regulate skeletal muscle development. These results will provide the theoretical basis for further study of LncRNA regulation and activity in different cattle breeds.

Keywords: shandong black cattle, luxi cattle, lncRNA, longissimus dorsi, identification

INTRODUCTION

As an important economic trait that affects the production efficiency of beef cattle, meat production traits are a research focus in the field of beef cattle genetics and breeding. The analysis of the molecular regulation mechanisms of muscle growth can facilitate cattle breeding. Local cattle breeds in China typically present reduced growth, but increased meat quality when compared to imported cattle. These differences may be related to differences in skeletal muscle development among different breeds (Xu et al., 2018). The growth and development of beef cattle skeletal muscle can be affected by a variety of regulatory factors. Previous work mainly focused on contributions from DNA, mRNA, and miRNA, but regulatory effects of long non-coding RNA (LncRNA) on the growth and development of beef skeletal muscle remain poorly understood (Zhang et al., 2007; Dylan et al., 2008; Xu et al., 2020).

Previous studies reported contributions of LncRNA in the process of skeletal muscle proliferation and differentiation. Liu identified, mapped, and determined the global and skeletal muscle expression patterns of 7,188 LncRNAs, finding that these LncRNAs had similar open reading frames and expression levels as those of other mammalian LncRNAs (Sun et al., 2016; Yu-ying et al., 2017). Subsequently, Yue identified a highly expressed *LncRNA-YYW* in muscle tissue (Yue et al., 2017). Microarray analysis showed that *LncRNA-YYW* positively regulated the expression of growth hormone-1 and its downstream genes *AKT1* and *PIK3CD* in bovine myoblasts, and promoted myoblast proliferation (Dylan et al., 2008; Billerey et al., 2014). Cai found that lnc-ORA interacted with IGF2BP2 to inhibit the PI3K/AKT signaling pathway, thereby inhibiting muscle production and inducing skeletal muscle atrophy (Cai et al., 2021). Studies had shown that LncMD could be used as a competitive endogenous RNA (ceRNA) to compete with IGF2 and bound miR-125b to weaken its inhibitory effect on IGF2, and promoted the differentiation of bovine skeletal muscle satellite cells (Sun et al., 2016). Linc-MD1 had been shown to be a competitive endogenous RNA for miR-133 and miR-135 targets during myoblast differentiation. These transcription factors activated muscle-specific gene expression to control the time of muscle differentiation (Marcella Davide et al., 2011). Lnc-31 maintained the proliferation of myoblasts and counteracts differentiation (Dacia et al., 2018). The active regulation of Rock1 translation by lnc-31 was of great significance to the control of myogenesis (Ballarino et al., 2015). These results suggest that LncRNA may regulate the growth and development of beef skeletal muscle and meat quality. Although LncRNAs lack coding capacity, many LncRNAs act in various biological processes, serving as an additional level of genomic regulation.

The first embryo transfer calf in China was obtained from a vitrified frozen somatic cell cloned embryo, which combined Japan black cattle, Luxi cattle, and Bohai black cattle. The new generation core breeding group of this new breed, Shandong black cattle, was established by the filial generation. The new generation core breeding group hybridized with Shandong black cattle, and the second generation group of Japan black cattle (3/4) was crossed with Luxi cattle (1/4) to produce the new generation (3/4) × Bohai black cattle (1/4). Ideal black cattle and Shandong black cattle were selected from the second-generation group for cross-cross fixation. In this stage, several good lines were cross-bred with distant relatives, and then selected and the excellent lines were preserved. After four generations, this line was finally bred into a new Shandong black cattle variety matching line and used as a bull. In 2015, the Shandong black cattle variety was approved by the National Animal and Poultry Genetic Resources Committee as a new population and successfully established as a new cultivated variety in China (Lu, 2015; Liu et al., 2018).

In recent years, there has been growing research of LncRNA, but most work has mainly focused on the effects of neurologic diseases, tumors, embryonic development, and cell differentiation in human and mouse, with little research on domestic animals. The growth and development of beef cattle skeletal muscle are highly regulated processes. Although the contributions of DNA, mRNA, and miRNA to this regulation have been studied, the extent to which LncRNA regulates beef skeletal muscle growth and development remains poorly understood. To investigate the potential contribution of LncRNA, RNA samples were prepared from longissimus dorsi muscle tissues of Shandong black cattle (hybrid offspring) and Luxi Cattle (the first maternal generation). RNA-seq technology was used to identify LncRNA transcripts and their genomic characteristics, and LncRNAs differentially expressed in skeletal muscle tissues in Shandong black cattle and Luxi cattle were determined. Using these data, we constructed an interaction network of LncRNA-miRNA-mRNA and muscle development which can guide future breeding efforts.

MATERIALS AND METHODS

Ethics Statement

The methods used in this study were performed in accordance with the guidelines of Good Experimental Practices adopted by the Institute of Animal Science (Qingdao Agricultural University,

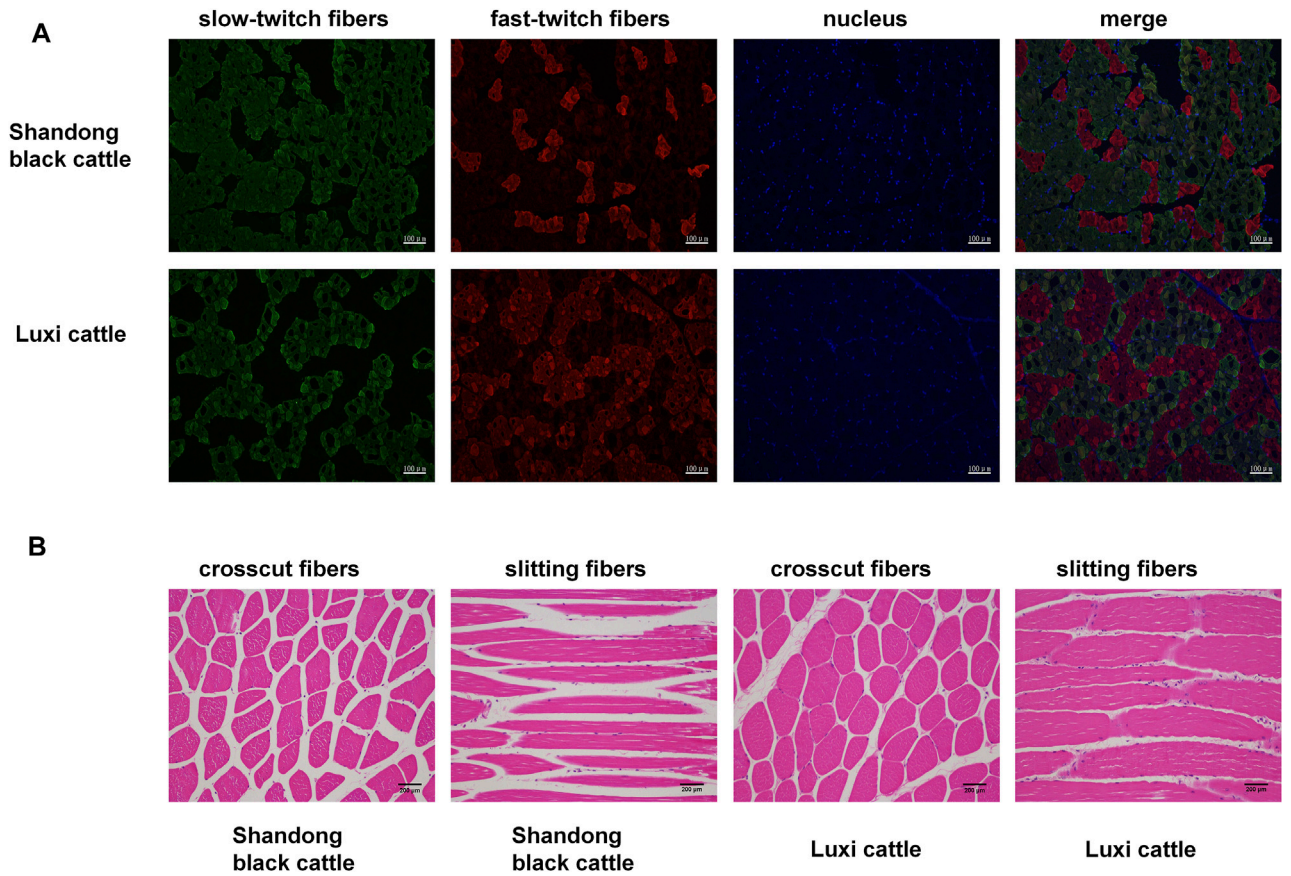


FIGURE 1 | Fast/slow muscle fluorescence staining and paraffin section HE staining of muscle tissue. **(A)** Green represents slow muscle fiber, red represents fast muscle fiber, and blue represents the nucleus of the muscle cell. **(B)** Red represents a single muscle fibre, Blue represents a nucleus.

TABLE 1 | Comparison of muscle fibre characteristics and growth characteristics of different breeds of cattle.

Characteristics	Shandong black cattle	Luxi cattle
Area [μm Zhang et al. (2007)]	5,490.222 ± 184.649*	4,869.008 ± 69.596
Diameter (μm)	106.837 ± 12.537	120.491 ± 4.324
Length (μm)	174.220 ± 7.395	142.435 ± 0.968
Density (Number of muscle fibres/Muscle fiber area, EA/μm Zhang et al. (2007))	4,887.848 ± 373.586**	7,172.966 ± 319.501
Number of muscle fibres (EA)	57.667 ± 4.333	39 ± 1.732
Fast-twitch fibers/Slow-twitch fibers	0.412 ± 0.096**	3.280 ± 1.082
Fast-twitch fibers/Muscle fiber area	0.200 ± 0.0342	0.652 ± 0.110
Slow-twitch fibers/Muscle fiber area	0.508 ± 0.046*	0.246 ± 0.072
Weight (kg)	509.667 ± 2.026	489.333 ± 1.764

*Note: In the table, * indicates a significant difference (p < 0.05); ** indicates a extremely significant difference (p < 0.01).*

Qingdao, China). All surgical procedures involving cattle were performed according to the approved protocols of the Biological Studies Animal Care and Use Committee, Shandong Province, China (Cao et al., 2014).

Animal and Tissue Preparation

Shandong black cattle(three)and Luxi cattle(three)were used in this study. Cattle were fed three times and received approximately equal amounts of green coarse feed and concentrate feed everyday according to standard NY5127—2002 pollution-free feeding management of beef cattle. They were raised in the same environment. At the age of 18 months, three healthy male beef cattle in each group were randomly selected for this study. The selected cattle had no scratches, scars, and scabs on their bodies, with no fat deposits in the internal organs or abdomen. No disease was found during examination, and all physiological and biochemical indexes were normal. *Longissimus dorsi* muscle samples were collected and immediately frozen in

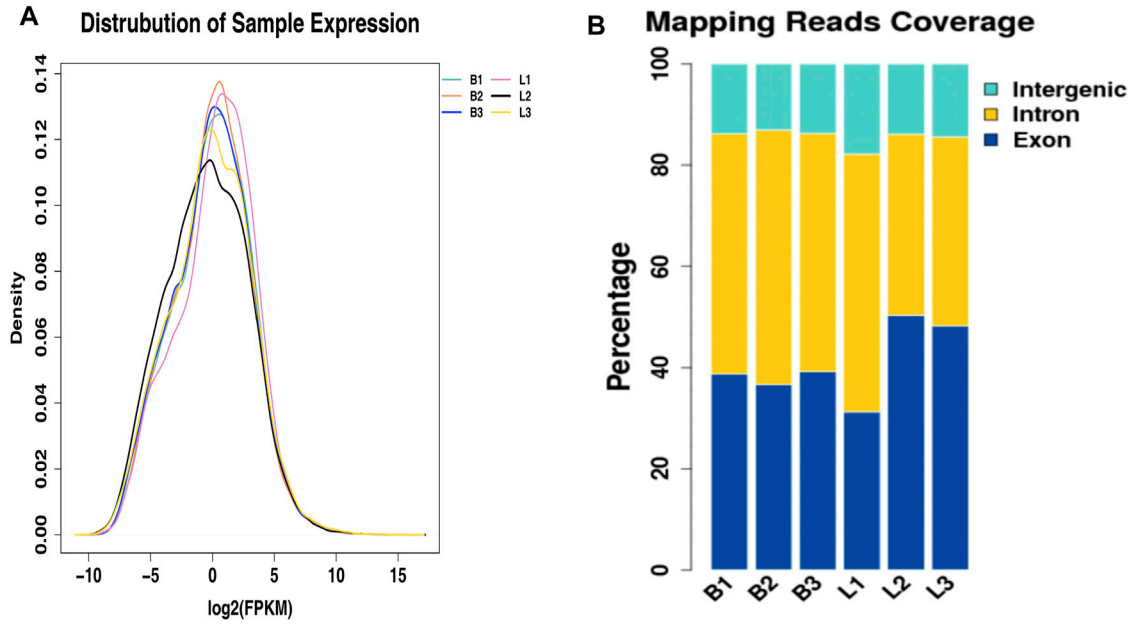


FIGURE 2 | The relative expression and the genome distribution of potential transcripts and clean reads. **(A)** The new transcripts were relatively expressed in longissimus dorsi muscle of Shandong black cattle and Luxi cattle, and the density distribution map was made after taking logarithm of 2 as the base. The abscissa was log2 (FPKM +0.0001), and the ordinate was gene density. Different colors represent different samples. **(B)** The results showed that the genomic distribution of clean data of longissimus dorsi muscle tissue of Shandong black cattle and Luxi cattle was genome distribution; the abscissa was each sample, and the ordinate was the ratio of the sequence number of exon, intron and intergenic region between the reference genome.

liquid nitrogen for RNA extraction. The cattle used in the experiment were euthanized as follows: first, a 1–3% sodium pentobarbital solution was prepared with physiological saline, and then intravenously injected. The injection dose is 90–135 mg/kg.

Hematoxylin and Eosin Staining of Muscle Tissue and Fast/Slow Muscle Fiber Fluorescence Staining

In order to better observe the histological morphology of muscle, we performed HE staining and fast/slow muscle staining. Paraffin sections were prepared from muscle tissue fixed with 4% paraformaldehyde. The HE staining protocol was performed as described previously (Liu et al., 2020). Briefly, dewaxing, covering with water, Hematoxylin staining, washing with water, 5% acetic acid differentiation, eosin staining, dehydration, natural drying, sealing, and image acquisition were performed.

Tissue sections were placed in a box filled with EDTA antigen repair buffer (Purchased from Shanghai Beyotime Biotechnology Co., Ltd) (ph 8.0) for antigen repair. After natural cooling, the slides were washed in PBS (pH7.4) with three ashes of 5 minutes each. BSA was added for blocking, and then the antibody was added, followed by DAPI to stain the nucleus. An autofluorescence quenching agent was added to the slices for 5 min and the samples were then washed with running water for 10 min. After natural drying, the film was sealed with an anti-fluorescence quenching sealing agent. Finally, the slices were observed under the fluorescence microscope and images were collected.

Image-Pro Plus software was used to count the images and measure the surface area. SPSS software was used for statistical analysis to determine significant differences.

RNA Extraction, Library Construction, and Sequencing

Total RNA samples were isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was checked using a NanoPhotometer® spectrophotometer (IMPLEN, Los Angeles, CA, United States). RNA concentration was measured using a Qubit® RNA Assay Kit in a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, United States). RNA integrity was assessed using a RNA Nano 6000 Assay Kit in a Bioanalyzer 2,100 system (Agilent Technologies, Santa Clara, CA, United States). Only samples with RNA Integrity Number (RIN) scores >8 were used for sequencing, which is different from the previous studies (Liu et al., 2020). The different RIN values are due to the different requirements for RNA quality when constructing different data sets. A total of 3 μg RNA per sample was used as the input material for RNA library preparation. This study used Illumina hiseq xten sequencing platform.

Transcriptome Assembly

The original sequencing data were processed by Fastp software (v0.23.2) (Chen et al., 2018) with parameters "-Q 20 -P90" with disjointing sequence and low-quality sequence. Clean data were obtained by removing reads containing

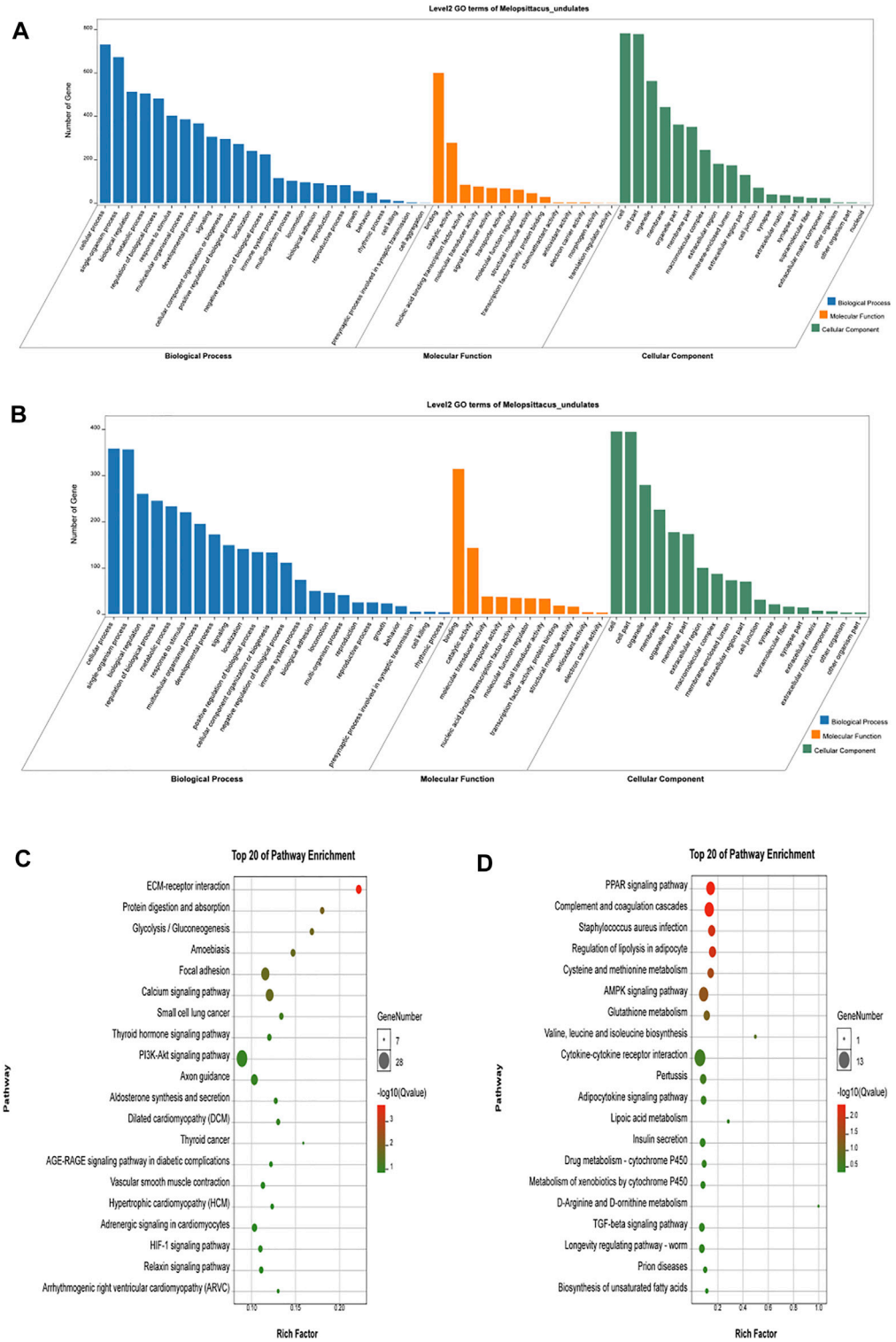


FIGURE 3 | Functional analysis of differential genes. **(A,B)** The results of Go cluster enrichment analysis for up-regulated and down-regulated genes: the horizontal axis represents the name of Go entry, and the vertical axis represents the number of genes enriched in the Go entry. **(C,D)**. Up-regulated genes and down-regulated genes were analyzed by Go cluster analysis. The horizontal axis represents the name of the sample group, and the vertical axis represents the KEGG biological pathway. Each point indicates the enrichment degree of the KEGG entry, and the closer the color to green, the higher the enrichment degree. The size of each dot indicates the number of genes enriched in the KEGG entry. The larger the dot, the more genes are enriched in the KEGG entry, and vice versa.

TABLE 2 | Partial GO term related muscle.

Term_type	GO Accession	*p*-value	q-value	Description	Up-regulated genes	Down-regulated genes
Biological_process	GO: 0060173	0.02064757	0.211430212	Limb development.	*LMOD1,TNNI2,FGF10,NOTCH1, ARMH2,HOXA11,MYH3,NOTCH4,NRARP,*	*AHI1,AQP2,CEBPA, CITED2,CSRP3,CXCL 10,EN1,FGF9,FHOD3,FRAS1,GDF11,GSC,ID2, KLB,KRT19,LMOD2,LOC107133268,LOC112445823 , MST1,MYBPH,MYH10MYOM3,MYOZ2,PAX2,PERP, POSTN,PROX1, PRR5LPRRX1,RFFL, RUNX2,SFRP5,SMAD9,SMOC1TNNI1,TP63*
	GO: 0010761	0.275749234	0.592245428	Myoblast migration. Fast	*OSR1,PODXL,WNT4,AARD, ACVRL1, ADAMTS2,ADGRG1,AGTR2,BMPER, ERRFI1,EYA1,FOXF2,GATA6,HIF1AN,*	
	GO: 0014883	0.02191239	0.211430212	muscle to slow muscle	*HOXA10,HOXC10,HOXC11,HOXD9,HYAL2,IGFBP5,ITGA6,JAG2, LAMA5,LMOD1,MDK,MYBPC2,*	
	GO: 0030239	3.42E-06	0.0003717	conversion. Assembly of	*MYH11,MYLK3,MYOZ3,NEBL, NEURL1,NFKBIA, NOTCH3, NR4A3,OSR2,PAK1,PDPN, PEAR1,PITX1,*	
	GO: 0035295	0.000424469	0.016857115	myofibrillar cones. Hippo	*PP3CA,PPP3R1,PRKD2,PTP4A3,RAB3A*	
	GO: 0008543	0.045842966	0.325955497	cascade signaling		
	GO: 0007219	0.009548728	0.13700442	pathway. Myotube		
	GO: 0035148	0.218341028	0.546196246	development. Fibroblast growth factor receptor signaling pathway. Notch signaling pathway. Myotube formation		
Molecular_function	GO: 0031432	0.015693382	0.275878444	Actin binding. Protomyosin	*DYRK3,MYH11,PLK1,TNK2,ALK,BMX,CCDC88C,CLK1,CNTN3,EFNB2, EML1,EML6,EPHB1,KDR,KIF20A,KIF22,KIF26A,MYH1,MYH3,*	*ANKRD2,AXL,DYNC1I1,FGF9,GAS2L2,KIF1A,KIF5C, MST1R,MYH10,MY O 1G,TRIM54,ALPK1, ANKRD1,CSRP3,FAM221A,KRT19, LIG3,LMOD2,MYBPH,MYL6B,MYOM3,NEXN, PREX2,STK32A,TSSK3,VRK2,WISP1*
	GO: 0005523	0.028051936	0.362260165	binding. Protein kinase	*MYH4,MYH8,MY O 10,MY O 5B,NEBL, NTRK3,PRAG1,REEP1,RET, S100A9,STIM1,TIE1,ACVRL1,BEAN1,CAMK1D,CDKL5,*	
	GO: 0004672	0.380205354	0.649365154	activity. Muscle	*EEF2K,GRK5,HTRA1,IGFBP7,IRAK2,ITGA6,LMOD1,MYBPC2,MYLK3, MYLK4,MYLPF,NEK7,NRK,PAK1,PHKA1,PIM1,PKN3,PNCK, PRKD2,PROX1,*	
	GO: 0008307	1.45E-05	0.006860596	structural components.	*RPS6KA3,SBK2,SPEG,TPM1,TRIM63,TRPM7,TSSK1B,TSSK2,TSSK6,WISP2,WNK2*	
	GO: 0004713	0.076243247	0.395407495	Protein tyrosine kinase		
	GO: 0008017	0.342175561	0.624173906	activity. Microtubule		
	GO: 0004712	0.656228503	0.835574106	binding. Protein serine/		
	GO: 0005520	0.015268829	0.275878444	threonine. Tyrosine		
	GO: 0015631	0.710643065	0.863835582	kinase activity. Insulin like growth factor binding. Microtubule binding		
Cellular_component	GO: 0005865	0.000202554	0.005408948	Myofibril	*TNNT2,TNNT3,TPM1,CACNA1S,FBXL22,FHL3,IGFN1, MYBPC2,MYH1,MYH3,MYH4,MYH8,MYL3,MYOZ3,NEBL, NOS1,PDE4B,PPP3CA,TRIM32,TRIM63*	*FHOD3,LMOD2,TNNI1,ANKRD1,ANKRD2,CRYAB, CSRP3,KRT19,LOC104969184,MYBPH, MYOM3,MYOZ2,NEXN, PDLIM1,SCN3B,SCN5A,SYNP O 2L,TRIM54,TWF2*

TABLE 3 | Partial KEGG pathway related muscle.

KEGG term (level 2)	*p*-value	q-value	UP-regulated genes	Down-regulated genes
Wnt signaling pathway	0.74905225	0.996316062	*NFATC2,NKD2,TCF7L1,CTNNB,IP1, WNT4 PPP3R1,PPP3CA*	*WNT5A,SFRP5*
Gap junction	0.8942074	0.996316062	*ADCY4,ADCY1,PDGFB*	*ADCY7*
TGF-beta signaling pathway	0.9865578	0.9998376	*LOC535280*	
Hedgehog signaling pathway	0.68703532	0.996316062	*GRK5*	*PTCH2*
FoxO signaling pathway	0.5736576	0.996316062	*S1PR1,FOX O 6,PIK3R3,GADD45B,GADD45G,PLK1*	
Hippo signaling pathway	0.5058804	0.974598156	*NKD2,TCF7L1,WNT4,TEAD1*	*WNT5A,ITGB2,DLG2,ID2,BMP8B*
Thyroid hormone synthesis	0.20608885	0.7960525	*ATP1A1,ADCY4,ADCY1,TSHR, CREB3L1*	*GPX3,ADCY7*
Ras signaling pathway	0.3897899	0.87757151	*RAPGEF5,PLA2G5,EFNA4,ANGPT2,PIK3R3,KDR, CALM3,ANGPT1,PAK1,FGF10,LOC521224, PLCE1,IGF2,PDGFB, VEGFA*	*PLA2G2D1,RASGRP1,RRAS2,FGF9,RASGRP4*
TNF signaling pathway	0.78064365	0.996317071	*NFKBIA,PIK3R3,SOCS3,CREB3L1,GR O 1*	*CXCL10,CASP3*
AMPK signaling pathway	0.4665018	0.9998376	*PFKFB2,PIK3R3,EEF2K, PFKFB3PFKM,PPARGC1A,CREB3L1*	
PI3K-Akt signaling pathway	0.959319	0.995760863	*COL6A2,COL6A1,COL4A5,COL4A6,LOC530102,COL4A2,ITGA6, COL1A1,LAMA5,NR4A1,ITGB8,COL1A2,EFNA4,ANGPT2,PIK3R3, KDR,LAMA3,DDIT4,ANGPT1,PKN3,FGF10,RXRA,COL6A3, CREB3L1,IGF2,PDGFB, VEGFA,LAMB2*	*PPP2R2C,HSP90AB1,FGF9,G6PC,PCK1*
cGMP - PKG signaling pathway	0.677879	0.992794525		*PDE3B,ADRA1A,ADORA1,ADCY7*
Rap1 signaling pathway	0.6979082	0.992794525		*PFN2,LOC101903261,RAPGEF4,FGF9,ADCY7*
Calcium signaling pathway	0.5893919	0.992794525		*RYR3,ADRA1A,PTGER3,GNA15,ADCY7*

adapters, reads containing over 10% of poly(N), and low-quality reads (>50% of the bases had Phred quality scores ≤10) from the raw data. All downstream analyses were based on the high quality clean data. The *Bos taurus* genome reference genome and gene model annotation files were downloaded from the NCBI database (CHIR_1.0, NCBI) (Mahmoudi et al., 2020). An index of the reference genome was built using Bowtie v2.0.6 (Cai et al., 2015; Cai et al., 2018) and paired-end clean reads were aligned to the reference genome using HISAT v2-2.1.0 (Yuan et al., 2019). The mapped reads from each library were assembled with Cufflinks v2.2.1 (Andersson, 2009), using Cufflinks with 'min-frags-per-transfrag = 0' and'–library-typefr-firststrand', and other parameters set as default.

LncRNA Discovery

Stringtie (Pertea et al., 2015; Kovaka et al., 2019) was used to sort reads into different classes and then generate a map for each class. Based on the length of LncRNA, and the characteristics of non-protein coding sequences, we established strict screening conditions to screen LncRNA as follows:

(1) Transcripts equal to or longer than 200 bp in length, and containing two or more exons;
(2) Transcript read coverage of at least five reads;
(3) No transcripts of known mRNA or other specific non-coding RNAs (rRNA, tRNA, snoRNA, or snRNA). This screening was based on gffcompare (http://ccb.jhu.edu/software/stringtie/gff.shtml) using the same species annotation file;
(4) According to the class in the comparison result Code information ("U", "I", "X") was used to screen potential lincrna and intronic LncRNA anti-sense LncRNA Coding potential prediction screening:

To further assess if the identified transcripts are LncRNAs, a variety of coding potential analysis software was integrated, including CNCI analysis, CPC analysis, Pfam protein domain analysis, and CPAT analysis (only for animals) (Langfelder et al., 2008; Jing et al., 2010; Luo et al., 2017; Ito et al., 2018). The transcripts identified as non-coding by several methods were the final potential LncRNA dataset.

Differential Expression Analysis

Comparison of raw counts data (Wang et al., 2010) for different genes is a very effective tool for quantitative estimation of gene expression based on RNA-Seq data. This method can eliminate the influence of gene length and sequencing quantity on the determination of gene

TABLE 4 | LncRNA specially expressed in Shandong Black cattle and Luxi cattle.

Breed	Genes	Genomic position	*p*-value	q-value
Luxi cattle	*LOC112447810*	chrNC_037335.1:59843881-59851000	0.002587626	0.03702486
	LOC112448752	chrNC_037338.1:13047319-13056419	3.14E-06	0.000158716
	LOC101905635	chrNC_037340.1:76723459-76739979	0.000910843	0.016809347
	LOC112449550	chrNC_037341.1:76478493-76487345	5.40E-06	0.000252129
	LOC104974324	chrNC_037342.1:75777823-75780010	0.000204918	0.005162159
	LOC104974692	chrNC_037344.1:67130432-67132302	2.33E-05	0.000858982
	LOC112442200	chrNC_037345.1:45609576-45616675	0.001584248	0.025977108
	LOC112442383	chrNC_037345.1:62159119-62160399	0.000214833	0.005332407
	LOC112443129	chrNC_037348.1:25091115-25128295	0.003878832	0.049737964
	LOC101903141	chrNC_037353.1:47669900-47687837	7.72E-13	1.95E-10
	LOC112444607	chrNC_037354.1:36287211-36298269	9.39E-07	5.76E-05
	LOC112446394	chrNC_037331.1:68871351-68873779	1.62E-08	1.64E-06
	LOC101907194	chrNC_037332.1:26047494-26049444	3.59E-12	8.06E-10
Shandong Black cattle	*LOC112447868*	chrNC_037335.1:410273-419534	4.43E-06	0.000212223
	LOC112449245	chrNC_037340.1:3165193-3220562	7.90E-15	2.68E-12
	LOC112441643	chrNC_037342.1:54862373-54875026	4.23E-05	0.001397951
	LOC112441659	chrNC_037342.1:78782537-78785977	0.000658605	0.01318643
	LOC104975164	chrNC_037346.1:61277209-61278623	6.05E-05	0.001918059
	LOC112442997	chrNC_037347.1:23104372-23114203	0.000324072	0.007462557
	LOC100847412	chrNC_037348.1:32211463-32219891	1.13E-06	6.79E-05
	LOC100847415	chrNC_037357.1:18645452-18649324	2.19E-06	0.000118565

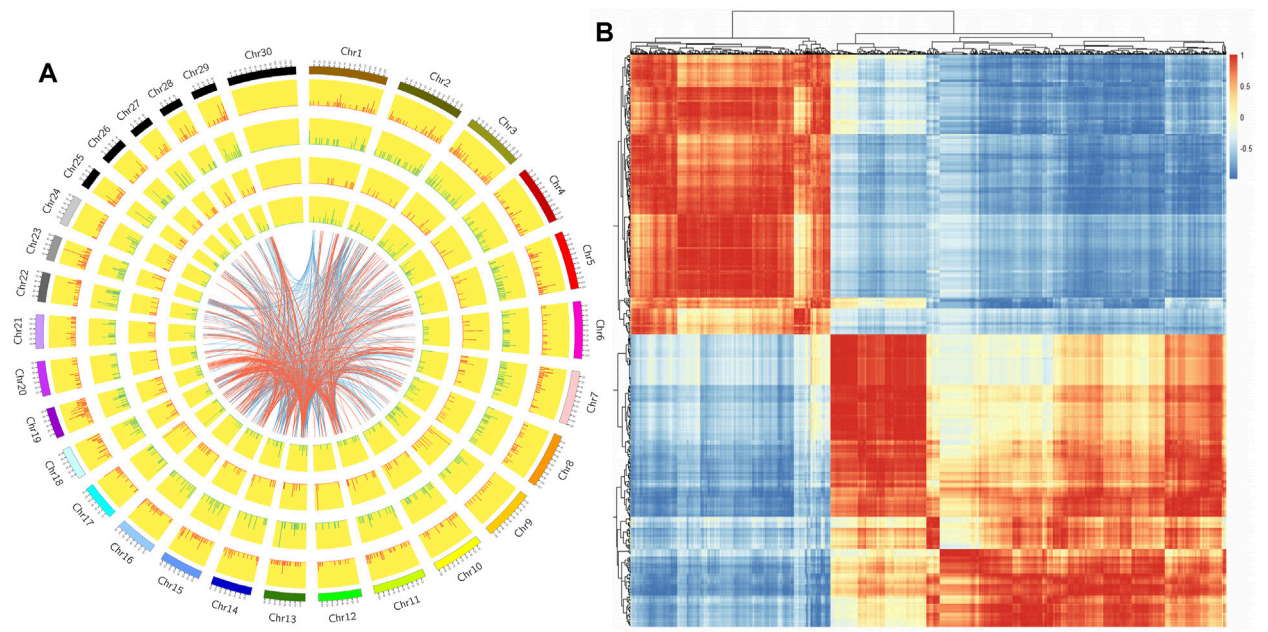


FIGURE 4 | (A). Circos map of differentially expressed genes and LncRNAs on cattle chromosomes. The outer ring is the diagram of autosomal distribution of the species; the first and second circles show the distribution of differentially expressed genes on chromosomes, with red lines indicating up-regulation, green lines indicating down-regulation, and the higher the column, the more differentially expressed genes in this region; the fourth and fifth circles show the distribution of LncRNAs on chromosomes, with the same expression pattern as genes. The internal lines indicate the corresponding relationship between LncRNAs and genes coexpressed by TOP 500. **(B)**. The Pearson correlation coefficient heat map of differential LncRNAs and mRNAs: Red represents positive correlation and blue represents negative correlation.

expression levels. The calculated gene expression values allow direct comparison of gene expression differences among different varieties. Here, we used DEseq (Anders and Huber, 2010) to analyze the differential expression between the treatment group with the reference group, and selected $|\text{Log2ratio}| \geq 2$ and $q < 0.05$ genes as indicative of significant differential expression. The numbers of up-regulated and down-regulated genes were also obtained.

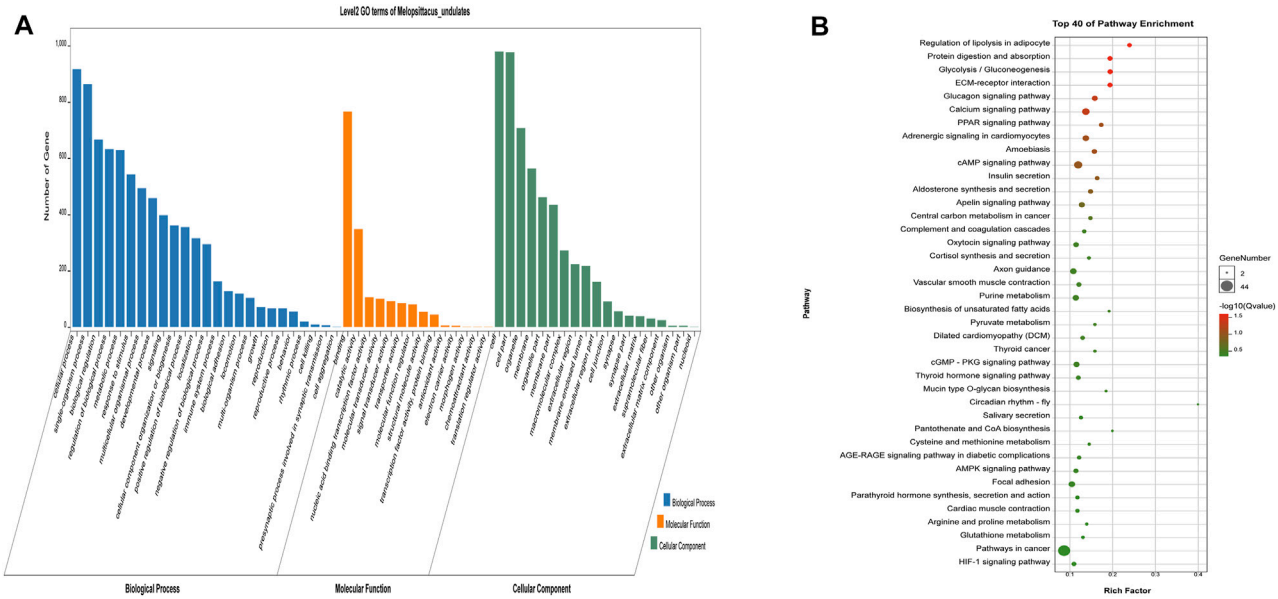


FIGURE 5 | Analysis results of potential target genes GO **(A)** and KEGG **(B)** of LncRNA differentially expressed in beef cattle muscle growth and development. The legend is the same as **Figure 3**.

Prediction and Analysis of Target Genes of Differential LncRNA

LncRNA can act on target genes by Cis or Trans. Cis effect refers to LncRNA acting on adjacent target genes (Cabili et al., 2011). To predict the cis-regulatory genes of LncRNA, we screened the 50 kb of sequence upstream and downstream of LncRNA and looked for potential target genes. The coexpression relationship between LncRNA and mRNA was described by Trans. The basic principle of transaction is that the function of LncRNA has nothing to do with the position of the coding gene,but it is related to the protein-coding genes it coexpressed. According to the correlation coefficient between LncRNA and mRNA expression (correlation coefficient cor ≥ 0.9); WGCNA(Weighted correlation network analysis) was used to predict the target genes of lncRNA for cluster analysis and functional enrichment analysis of LncRNA target genes. p-value < 0.05 was set as the significance threshold.

Analysis of Enrichment Pathway of GO and KEGG

GO (gene ontology) (Young et al., 2010) can be used for enrichment analysis of target genes with differential expression and Goseq (Young et al., 2010) was used to analyze the target genes of differentially expressed LncRNA. A value of $p < 0.05$ is considered as significant enrichment of differentially expressed genes.

The KEGG database (Kanehisa and Goto, 2000) describes advanced functions and utility of biological systems such as cells, organisms, and ecosystems (http://www.genome.jp/kegg/). We used KOBAS V3.0 software (Bu et al., 2021) to detect the enrichment of LncRNA target genes differentially expressed in the KEGG pathway.

Prediction of LncRNA-miRNA-mRNA Interaction

As competing endogenous RNAs (ceRNA), LncRNA can act as a sponge to adsorb miRNA and affect mRNA expression. Therefore, to further understand the function of LncRNA in the growth and development of skeletal muscle, we used TargetScan (http://www.targetscan.org/vert_71/) miRanda ((http://www.microrna.org/microrna/home.do) and PITA qtar (https://genie.weizmann.ac.il/pubs/mir07/mir07_dyn_data.html) software to predict the differentially expressed miRNAs(Sequencing has been completed) adsorbed by LncRNA and downstream-regulated mRNA. Online miRNA binding site prediction software (RNA22: http://cbcsrv.watson.ibm.com/rna22.html and RNAhybird ()) predicted potential interaction of miRNA with LncRNA. We then constructed an LncRNA-miRNA-mRNA interaction network map using Cytoscape version 3.5.1 (Shannon et al., 2003).

Luciferase Reporter Assay

Cells were seeded in 96-well plates at a density of 5 × 103 cells (HEK-293T) per well, 24 h before transfection. The cells were co-transfected with a mixture of 50 ng Firefly luciferase (FL) reporter vectors, 5 ng Renilla luciferase (RL) reporter vectors (pRL-TK), and miRNA mimics at the indicated concentration. The miRNA mimics were obtained from Life Technologies. After 48 h, the luciferase activity was measured with a dual luciferase reporter assay system using the psiCHECK-2 vector (Promega, Madison, WI). The *LOC104975788* and *LOC536229* (*Pax7*) sequences were separately cloned into the reporter gene vector (psiCHECK-2) to synthesize the predicted miRNA mimics and control. The potential binding target of each miRNA was cloned into the 3′UTR region of r-luciferase (hrluc), and then co-transfected with the miRNA to

TABLE 5 | Partial GO term related muscle.

GO term (level 2)	*p*-value	q-value	Tagert genes
Muscle structure development	4.89E-16	3.99E-13	*PITX1,MYOZ3,EPHB1,KY,MYORG, TRIM32,VGLL2,BVES, CITED2,DLL1,FHL2,LOC101903367,COL3A1,TRIM54,PLPP7,NOTCH1,RXRA,FGF9,HOXD9,CXADR, EFNB2,NEBL, GPCPD1,LAMA5,TCF15,JPH2,NFATC2,ADGRB1,CHD7,CRYAB,USP2,SOX6,MYBPH, ZBTB18,TBX1,LMOD1,TNNI1,NOS1,TBX1,MYLK3,IGFBP5,SPEG, MYBPC2,HLF,CHRND,MYH10,MYOCD,FGF10,MYOM3,WNT4,CACNA2D2,MYL3,MNF1,PIM1,JARID2,FHOD3,GATA6,MYH11,MYLPF, ANKRD1,ANKRD2,HIF1AN,NEURL1,MEF2D,CSRP3,IGF2,MBNL3,LOC112445030,S1PR1,NEXN,COL6A3,HDAC9,LMOD2,MYF5,MYL6B,PKP2,MKL1,MYOZ2,PPP3CA,UCHL1,CXCL10,SHOX2,OLFM2*
Muscle tissue development	1.11E-13	5.55E-11	*PITX1,EPHB1,MYORG, VGLL2,BVES, CITED2,DLL1,FHL2,LOC101903367,COL3A1,OSR1,ID2,NOTCH1,RXRA,FGF9,HOXD9,CXADR, EFNB2,NEBL, GPCPD1,JPH2,CHD7,EYA1,SOX6,ZBTB18,PROX1,TNNI1,TBX1,MYLK3,IGFBP5,HLF,CHRND,MYH10,MYOCD, ADAMTS9,MYL3,MNF1,PIM1,JARID2,FHOD3,GATA6,MYH11,MYLPF, ANKRD1,ANKRD2,NEURL1,MEF2D,CSRP3,S1PR1,NEXN, HDAC9,TP63,MYF5,MYL6B,PKP2,PPP3CA,SHOX2*
Striated muscle tissue development	3.06E-12	1.18E-09	*PITX1,EPHB1,MYORG, VGLL2,BVES, CITED2,DLL1,FHL2,LOC101903367,ID2,NOTCH1,RXRA,FGF9,HOXD9,CXADR, EFNB2,NEBL, GPCPD1,JPH2,CHD7,EYA1,SOX6,ZBTB18,PROX1,TNNI1,TBX1,MYLK3,HLF,CHRND,MYH10,MYOCD, ADAMTS9,MYL3,MNF1,PIM1,JARID2,FHOD3,GATA6,MYH11,MYLPF, ANKRD1,ANKRD2,NEURL1,MEF2D,CSRP3,S1PR1,NEXN, HDAC9,MYF5,MYL6B,PKP2,PPP3CA,SHOX2*
Muscle organ development	1.19E-11	4.08E-09	*PITX1,EPHB1,KY,MYORG, VGLL2,BVES, CITED2,DLL1,LOC101903367,COL3A1,NOTCH1,RXRA,FGF9,HOXD9,CXADR, EFNB2,GPCPD1,LAMA5,TCF15,JPH2,ADGRB1,CHD7,CRYAB,USP2,SOX6,ZBTB18,PROX1,TNNI1,TBX1,SPEG,HLF,CHRND, MYOCD,MYL3,MNF1,PIM1,JARID2,GATA6,MYLPF, ANKRD1,ANKRD2,NEURL1,MEF2D,CSRP3,LOC112445030,S1PR1,COL6A3,HDAC9,MYF5,MYL6B,PKP2,PPP3CA,SHOX2*
Striated muscle cell differentiation	1.88E-11	6.14E-09	*MYOZ3,MYORG, TRIM32,DLL1,FHL2,PLPP7,NOTCH1,RXRA, CXADR,EFNB2,NEBL, NFATC2,ADGRB1,SOX6,MYBPH, PROX1,LMOD1,NOS1,TBX1,MYLK3,IGFBP5,MYBPC2,MYH10,MYOCDMYOM3,CACNA2D2,FHOD3,GATA6,MYH11,ANKRD2,CSRP3,IGF2,NEXN, HDAC9,LMOD2,MYF5,MYOZ2,PPP3CA,UCHL1,CXCL10,SHOX2*

TABLE 6 | Partial KEGG pathway related muscle.

KEGG term (level 2)	*p*-value	q-value	Tagert genes
Calcium signaling pathway	0.00083234	0.04879693	*ATP2A1,RYR3,MYLK3,P2RX6,CACNA1S,MYLK4,ADRA1A,PTGER3,PHKB, PLCD3,NOS1,CALM3,PLCD4,GNA15,ADCY4,STIM1,ORAI1,ATP2B2,HRH2,PPP3CA,ADCY1,PHKA1,ADCY7,PLCE1*
PPAR signaling pathway	0.00244591	0.0812586	*PLIN2,ME3,SCD5,FABP3,FABP7,PLIN1,RXRA, FABP4,PCK1,GK,FADS2,ADIPOQ*
Vascular smooth muscle contraction	0.02665648	0.36304747	*PPP1R14A,MYLK3,CACNA1S,MYLK4,MYH11,ADRA1A,MYH10,CALM3,ADCY4,ADCY1,MYH4,MYH1,ADCY7,MYL6B*
cGMP - PKG signaling pathway	0.02036938	0.36304747	*ATP2A1,NFATC2,MYLK3,TRPC6,CACNA1S,MYLK4,PDE3B,ADRA1A,ERI1,ATP1A1,CALM3,ADORA1,ADCY4,ATP2B2, PPP3CA,ADCY1,ADCY7,CREB3L1*
Thyroid hormone signaling pathway	0.03464076	0.38992103	*ATP2A1,ESR1,RCAN1,PLCD3,ATP1A1,WNT4,DI O 2,PLCD4,PFKM, RXRA,SLC2A1,PLCE1,NOTCH1*
AMPK signaling pathway	0.04135674	0.3978691	*PPP2R2C,SCD5,CAB39,ADRA1A,LEPR,EEF2K,PFKFB3,PFKM,FBP1,LIPE, PPARGC1A,CREB3L1,PCK1,ADIPOQ*
Cardiac muscle contraction	0.05499784	0.444442	*ATP2A1,CACNA1S,CACNA2D2,ATP1A1,MYL3,COX7A1,ACTC1,UQCRB,TPM1,LOC101902754,CACNA2D3*

determine the activity of R-Luciferase. F-Luciferase (hluc +) was used as an internal reference to correct for differences in transfection efficiency between different samples. The miRNA mimics and psicheck-LOC104975788 or psicheck-LOC536229 (*Pax7*) were co-transfected into 293T cells. The expression level of reporter genes was detected using a multi-functional enzyme labeling instrument, and the miRNAs that exhibited down-regulated reporter gene expression were further screened.

Validation of Sequencing Data by Quantitative Reverse Transcription PCR (qRT-PCR)

The reaction system (20 μl) for the RT-PCR reaction consisted of the following: 1 μl of template cDNA, 10 μl each of the upstream and downstream primers, and 5 ml (5 × 1 ml vials) of RNase-free water. The thermal cycling procedure was as follows: 94°C for 10 min, 94°C for 30 s, 60°C for 30 s, and 72°C for 40 s, with 40 cycles12. The expression of GAPDH was calculated by the $2^{-\triangle\triangle CT}$ method. Primers used for qRT-PCR as shown in **Supplementary Table S1**.

RESULTS

To investigate the potential regulatory effects of LncRNA on the growth and development of beef skeletal muscle, we analyzed samples from muscles of different breeds of beef cattle.

Apparent Differences in Muscle Fibres in Different Breeds of Beef Cattle

First, we obtained samples from Shandong black cattle and Luxi cattle and performed fluorescence staining sections of fast and slow muscle fibers (**Figure 1A**) and HE staining (**Figure 1B**). The results showed significant differences in muscle fibers of *longissimus dorsi* muscles between Shandong black cattle and Luxi cattle. IPP(Image-Pro Plus) software analysis showed that the average area of muscle fibres of Shandong black cattle was significantly larger than that of Luxi cattle ($p < 0.05$), with significant differences in the muscle fibre density and the ratio of fast-twitch fibers to slow-twitch fibers and slow-twitch fibers to muscle fiber area ($p < 0.05$), but not in other muscle fibre properties ($p > 0.05$) (**Table 1**).

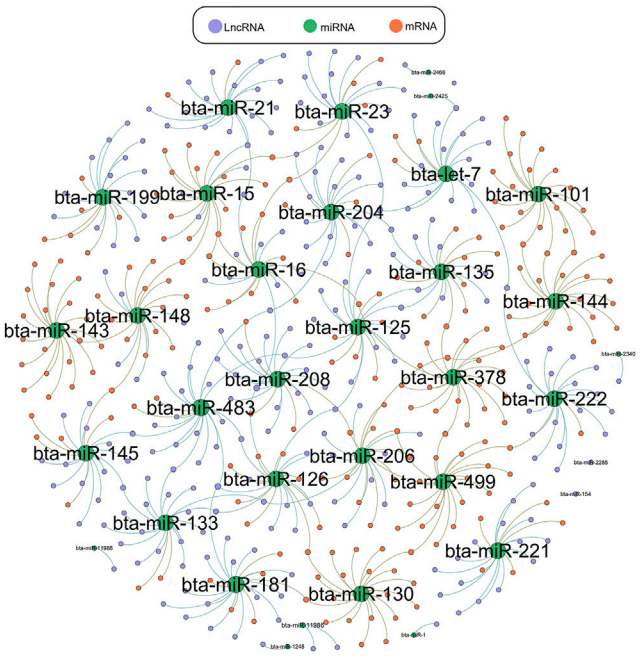


FIGURE 6 | LncRNA-miRNA-mRNA network interaction diagram. LncRNA-miRNA-mRNA interaction network diagram related to muscle, green represents miRNA, orange represents mRNA, purple represents LncRNA.

Overview of Sequencing Data of *Longissimus Dorsi* in Beef Cattle

RNA was extracted from longissimus dorsi muscle samples of Shandong black cattle and Luxi cattle and sequenced. Totals of 105345396 and 126735736 reads were obtained from Shandong black cattle and Luxi cattle, respectively, with 101145946 and 122463284 mapped reads. In the data for Luxi cattle *longissimus dorsi* muscle, the percentages of exons, introns, and genes were 43.26%, 41.34%, and 15.4%, respectively, with 38.16%, 48.3%, and 13.54%, respectively, for Shandong black cattle *longissimus dorsi* muscle tissue, as shown in **Figure 2**.

Analysis of Differential mRNA Expression

A total of 20104 transcripts were obtained from the two libraries, including 18859 and 18976 mRNAs for Luxi cattle and Shandong black cattle, respectively. According to the screening criteria of $|FC| \geqq 2$ and $q < 0.05$, 1,415 differentially expressed mRNA (**Supplementary Table S2**) were found, with 939 mRNAs significantly up-regulated and 476 down-regulated in Luxi cattle. The 92 mRNAs were specifically expressed in cattle, with 30 and 62 mRNAs specifically expressed in Shandong black cattle and Luxi Cattle, respectively (**Supplementary Table S2**). The top 40 genes with significant differential expression in the two breeds in cattle are listed in **Supplementary Table S3**, and suggest genes that could be used for the functional comparison and analysis of differences among varieties.

Functional Analysis of Differential mRNA

Gene Ontology (GO) is an international standardized classification system of gene function that includes three categories: biological process (BP), molecular function (MF), and cell composition (CC). We performed GO enrichment analysis on 1,415 differentially expressed genes (**Figures 3A,B**), and the results showed 2,662 items ($p < 0.05$), with 388 were enriched in cell composition (CC), 361 in molecular function (MF), and 1913 in biological process (BP) in up-regulated genes (Shandong black cattle as a reference group). The enrichment results of down-regulated genes showed 1883 items in total ($p < 0.05$), with 251 enriched in cell composition (CC), 247 in molecular function (MF), and 1,385 in biological process (BP). In the muscle regulation system, muscle contraction is significantly enriched in GO terms, indicating that some differential genes may participate in muscle development and function and other life activities. Muscle-related GO terms are shown in **Table 2**.

Different genes coordinate with each other to perform their biological functions in organisms. We next performed KEGG pathway analysis of the differentially expressed genes to identify the related signal transduction and metabolic pathways (**Figures 3C,D**). The results showed enrichment of 939 up-regulated genes in 243 pathways ($p < 0.05$), and enrichment of 476 down-regulated genes in 239 pathways, including metabolism, disease, cell communication and endocrine regulation (**Supplementary Table S4**). The enriched pathways ($p < 0.05$) included some signaling pathways related to skeletal muscle growth, development, and function (**Table 3**), such as the Wnt signaling pathway, the calcium signal transduction pathway, gap junction, the Hippo signaling pathway, the AMPK signaling pathway, and the thyroid hormone synthesis pathway.

TABLE 7 | LncRNA-miRNA-mRNA related to skeletal muscle development and growth.

LncRNA	microRNA (Targets-miR)	Targets (including potential targets)	Tissues and cells
LOC525506,LOC526226,LOC112444238,LOC101903367	*miR-135a/b*	*Smad5, JAK2 , MEF2C*	C2C12, Hodgkin lymphoma
LOC540051,LOC112448318,LOC112441863,LOC112449549,LOC101907194	*miR-143/145*	*SRF, myocardin, Nkx2-5*	smooth muscle, skeletal muscle
LOC112448962,LOC104975788,LOC112449549	*miR-133a/b*	*SRF,KLF15,Igf1R, Runx2, dynamin 2,Pax7, MAML-1*	C2C12, skeletal muscle, cardiac muscle
LOC112449031	*miR-365-3p*	*SelT*	skeletal muscle
LOC101904174	*miR-214*	*Nras*	C2C12, skeletal muscle, cardiac muscle
LOC104970868	*miR-449a*	*Cdk6, Cdc25a*	C2C12 , skeletal muscle
LOC112444607,LOC112441811, LOC101906545	*miR-423*	*SRF, SUFU*	skeletal muscle
LOC104975788,LOC112447073	*miR-103*	*Wnt3a,RAI14*	C2C12 , skeletal muscle
LOC100847415	*miR-17-5p*	*Mfn2, RB1, ENH1*	Skeletal muscle satellite cells
LOC525506,LOC540051,LOC514189,LOC112444238,LOC112446526,LOC112448972,LOC112446078	*miR-21*	*PTEN, PDCD4*	skeletal muscle, cardiac muscle, smooth muscle
LOC525506	*miR-1*	*KCNJ2,HSP60,HSP70,caspase-9,c-Met,Pax7,Pax3,IGF-1R,KLF4*	C2C12, skeletal muscle, cardiac muscle, smoothmuscle
LOC525506,LOC515772	*miR-125a/b*	*IGF-II, Cbx7, SP7*	skeletal muscle, smooth muscle,ESC
LOC112445239,LOC112441718,LOC112446120	*miR-126*	*Spred-1, VCAM-1, IRS-1*	cardiac muscle, ESC
LOC525506,LOC112449623	*miR-128a*	*PPARγ, Runx1, Pax3*	skeletal muscle
LOC112442443,LOC514189,LOC112448623,LOC100299745,LOC112443546,LOC112446047,LOC518869,LOC508468	*miR-130a*	*GAX, HOXA5*	vascular endothelial cells (ECs)
LOC525506,LOC112441673	*miR-144*	*IRS1*	Type II diabetes mellitus
LOC112441718,LOC112448623	*miR-148a*	*ROCK1*	skeletal muscle
LOC540051,LOC526226,LOC112444791,LOC112448967,LOC615901,LOC508180,LOC617406	*miR-15a*	*DLK1*	3T3-L1 preadipocytes
LOC540051,LOC112444791,LOC11244896,LOC615901,LOC508180,LOC617406	*miR-16a/b*	*cyclin D1*	skeletal muscle
LOC540051,LOC526226,LOC112446505,LOC112442443,LOC112444239,LOC788100	*miR-181*	*Hox-A11 Cbx7*	skeletal muscle, ESC
LOC525506,LOC112446505,LOC514189,LOC112446526	*miR-199a*	*Hif-1α, Sirt1*	cardiac muscl
LOC525506,LOC787554	*miR-204*	*Runx2*	smooth muscle
LOC525506,LOC112448970	*miR-208b/499*	*Sox6, Purβ, Sp3, HP-1β*	muscle
LOC540051,LOC514189,LOC112446078,LOC112441937	*miR-214*	*Ezh2 N-Ras*	C2C12, skeletal muscle, ESC
LOC112442455,LOC112446845	*miR-221*	*p27, Mdm2*	skeletal muscle, mesenchymal cells
LOC525506,LOC526226,LOC112442877,LOC112449155	*miR-23a*	*MAFbx/atrogin-1, MuRF1, Myh1/2/4*	skeletal muscle

Note: *Green represents up-regulated LncRNAs, red represents down-regulated LncRNAs, black represents LncRNAs, with no significant difference.*

Screening of Differential LncRNAs

A total of 8,427 LncRNAs were found in the small RNA Library of *longissimus dorsi* muscle of Shandong black cattle and Luxi cattle. Of these, 3,498 LncRNAs were annotated in non-coding regions and the remaining 4,929 LncRNAs were in regions that are not annotated. According to the screening criteria of $|FC| \geqq 2$ and $q < 0.05$, 480 differentially expressed LncRNAs were found, with 245 and 235 LncRNAs significantly up-regulated and down-regulated in Luxi Cattle compared to the levels in Shandong black cattle (**Supplementary Table S5**). A total of 92 LncRNAs were only expressed in one cattle breed, with 30 specific LncRNAs in Shandong black cattle and 62 specific LncRNAs in Luxi cattle (**Supplementary Table S5**). These breed-specific *longissimus dorsi* muscle LncRNAs included eight annotated LncRNAs in Shandong black cattle and 13 annotated LncRNAs in Luxi cattle, as shown in **Table 4**. The LncRNAs specifically expressed in Shandong black cattle and Luxi cattle, the first 40 unexplained LncRNAs and annotated LncRNAs are listed in **Supplementary Table S6**.

We constructed a Circos diagram to visually display the distribution of differential LncRNAs and differential mRNAs (**Figure 4A**). The results showed that most differentially

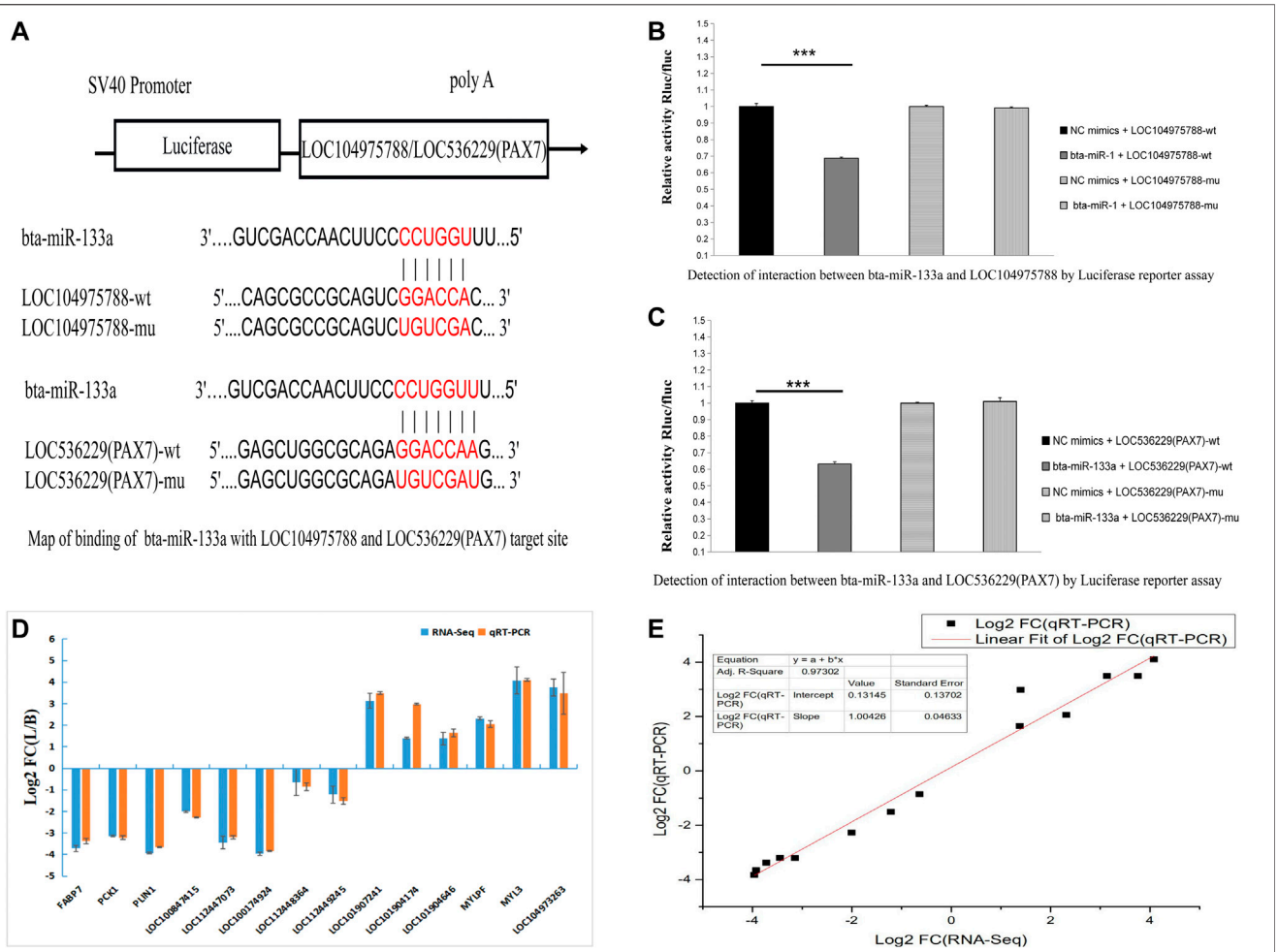


FIGURE 7 | Map of binding of bta-miR-133a with LOC104975788 and LOC536229 (Pax7) target site **(A)**. Luciferase reporter assay showed significantly lower luciferase activity of miR-133a co transfected with the psi-LOC112448162 vector compared to the control group ($p < 0.01$). The luciferase activity of miR-133a co-transfected with psi-LOC536229 (Pax7) vector was also significantly decreased relative to the control Q19 group ($p < 0.01$) **(B, C)**. These results indicate that both LOC104975788 and LOC536229 (Pax7) contain miR-133a binding sites. Nine differentially expressed LncRNAs and five differentially expressed mRNAs were verified by qRT-PCR **(D, E)**.

expressed LncRNAs mapped to chromosome 9, with others mapping to chromosomes 3, 1, 11, and 7. Pearson correlation coefficient analysis of differentially expressed LncRNA and differential mRNA genes (**Figure 4B**) showed 43844 pairs of co-expressed LncRNAs and differential mRNA genes with correlation coefficient >0.9 (**Supplementary Table S7**). Based on genomic location, 387 differentially expressed LncRNAs were selected for further study.

Prediction of Differentially Expressed LncRNAs Target Genes

LncRNAs can regulate the expression of adjacent mRNAs. We analyzed the LncRNAs and protein-coding genes by analyzing mRNAs within 50 kb of LncRNAs. The results showed that 387 of the differentially expressed LncRNAs targeted 1,164 genes. In these predictions, one LncRNA targeted multiple mRNAs and many LncRNAs targeted the same mRNA (**Supplementary Table S8**). Of these, *LOC112448071*, *LOC112444635*, *LOC112445963*, *LOC104975064*, *LOC101903261*, *LOC535280*, and *LOC521224* can target genes related to muscle development, including *MYORG*, *Wnt4*, *PAK1*, and *ADCY7*.

The potential target genes of LncRNA were next subjected to GO and KEGG analysis (**Figure 5**). We found 130 significantly enriched GO items ($p < 0.05$), with five related to the regulation of muscle development (**Table 5**). The 40 most enriched terms include multicellular organization development, single multicellular organization process, and multicellular organic process (**Supplementary Table S9**). The GO terms related to muscle development are listed in **Table 5**. KEGG analysis showed enrichment of 1,164 potential target genes in 299 pathways, with the 40 most significantly enriched pathways listed in **Supplementary Table S9**. Some of these are related to muscle development (**Table 6**), including the calcium signaling pathway, the AMPK signaling pathway, the cGMP-PKG signaling pathway, and the PPAR signaling pathway.

LncRNA-miRNA-mRNA Network Interaction Analysis

We used Targetscan (Agarwal et al., 2015), Miranda (Enright et al., 2003), and PITA qtar software (Li et al., 2018) to predict the target gene relationships between mRNA and miRNA, and used Miranda and PITA qtar to predict the target gene relationships between miRNA and LncRNA. LncRNA can function as ceRNA to regulate the expression of downstream genes at the post-transcriptional level by binding miRNAs. We selected several miRNAs related to muscle development (including *miR-1*, *miR-133*, *miR-206*, *miR-181*, *miR-21*, and *miR-23a*) to predict and screen downstream-regulated mRNA using Targetscan and Miranda software. We then constructed an LncRNA-miRNA-mRNA interaction network diagram using Cytoscape version 3.5.1 (**Figure 6**). LncRNAs that may be involved in the development of beef cattle muscle are summarized in **Table 7**.

Luciferase Reporter Assay

We next looked for potential miRNA binding sites of *LOC104975788* and found potential *miR-133a* and *miR-103* binding sites. Online miRNA binding site prediction software (RNA22 and RNAhybird) predicted potential interaction of *miR-133a* with *LOC104975788* and *LOC536229*(*Pax7*). The results showed that both *LOC104975788* and *LOC536229* (*Pax7*) contained *miR-133a* binding sites.

Verification of Sequencing Results

Nine differentially expressed LncRNAs and five differentially expressed mRNAs were verified by qRT-PCR. The qRT-PCR results (**Figures 7D,E**) confirmed differential expression of these LncRNAs and mRNAs.

DISCUSSION

The muscle fiber is the basic unit of muscle, and the type of muscle fiber greatly shapes muscle characteristics (Steinbacher et al., 2006). There were obvious differences in appearance in the muscle fibres isolated from *longissimus dorsi* muscles of Shandong black cattle and Luxi cattle. The average muscle fiber area of Shandong black cattle was significantly larger than that of Luxi cattle ($p < 0.05$), and the muscle fiber diameter was smaller than that of Luxi cattle. There was a significant positive correlation between muscle fiber area and carcass traits. Muscle fiber diameter is closely related to meat quality and taste and thicker muscle fiber diameter corresponds to decreased muscle tenderness. There is a negative correlation between muscle fiber diameter and muscle fiber density (Zhou, 2010). The amount of slow muscle fiber affects sarcomere length and has an important impact on meat quality (Hendricks et al., 1971). There were significant differences in the muscle fiber density, the ratio of fast-twitch fibers/slow-twitch fibers, and weight ($p < 0.05$) in this study. These differences may be key factors leading to the differences in meat production performance and meat quality of the two breeds of cattle after birth, which was also the research basis of this study to explore the underlying molecular regulatory mechanism.

With low conservation of LncRNA sequences among species, bioinformatics methods are required to screen and identify LncRNAs. This type of analysis is based on transcript length, the number of exons, and the coding potential. In this study, transcriptome data of skeletal muscle of Shandong black cattle and Luxi cattle were generated and analyzed. In our studies, 1,415 transcripts were found to be differentially expressed, with 939 and 476 transcripts were significantly up-regulated and down-regulated in Luxi cattle (Shandong black cattle as a reference group), respectively. Further, 19 GO items and 14 regulatory pathways related to muscle development were screened by hierarchical GO and KEGG cluster analysis. Many genes involved in the regulation of muscle development were identified, including myosin protein family genes (*MYH1*, *MYH3*, *MYH4*, *MYH8*, and *MYL3*), myogenic regulatory factor family genes (*MY O 10* and *MY O 5B*), Homeobox family genes (*HOXC10*, *HOXC11*, and *HOXD9*), Troponin T family genes (*TNNT2* and *TNNT3*) and some regulatory transcription factors (*WNT4* and *TNNT2*). It is worth noting that three LncRNAs (*LOC107133268*, *LOC112445823*, *LOC104969184*) are directly enriched into muscle regulation items. Further target gene prediction analysis shows that LncRNAs may participate in the regulation of muscle development through *bta-miR-2892* (*LOC107133268*), *bta-miR-2360* and *bta-miR-2449* (*LOC112445823*). The specific regulatory mechanism needs to be further verified and analyzed.

Compared with studies in human (Federica et al., 2019) and other model organisms (James et al., 2015), there has been limited identification and characterization of beef LncRNAs, especially ones related to skeletal muscle development, with most studies of cattly focused on the identification of genes and miRNAs (Xue et al., 2013; Peters et al., 2017). In this study, we identified 8,427 multiple exon LncRNAs in beef skeletal muscle, and 480 differentially expressed LncRNAs were identified. More LncRNAs were detected in this study than previously reported in goats (Sun et al., 2013; Ying-hui et al., 2019). Fifteen randomly selected differentially expressed transcripts were verified by qPCR, and the results were consistent with the results of RNA sequencing. In conclusion, these results confirm the reliability of the identification of LncRNAs (Langfelder et al., 2008).

Although LncRNAs can act on gene sites far from their chromosomal location (Cohen et al., 2000), genes in close proximity on a chromosome may participate in the same cellular metabolic pathways and have similar biological functions (Leonardo et al., 2011). Therefore, the distribution of differentially expressed LncRNAs on chromosomes and the linkage differential expression of nearby genes may have biological significance that can help us to determine the function of a gene. The most differentially expressed LncRNAs were found on chromosome 9, followed by chromosomes 3, 1, 11, and 7. In co-expression analysis, we detected 387 differentially expressed LncRNA transcripts related to protein-coding genes according to the expression correlation coefficient values ($r > 0.9$), and predicted 1,164 target genes. GO analysis showed 20 GO terms related to the regulation of genes involved in muscle

development. KEGG analysis showed enrichment of 1,164 potential target genes in 299 pathways, with some related to muscle development, such as the calcium signaling pathway, the AMPK signaling pathway, the cGMP - PKG signaling pathway, and the PPAR signaling pathway. Interestingly, we also found *MYORG*, *Dll1*, *EFNB2*, *SOX6*, *PROX1*, *MYOCD*, *NEBL* and *MYLK3* genes, annotated as related to muscle development. Overall, LncRNAs may play a regulatory role in skeletal muscle biological development through cis or trans mechanisms.

At the post transcriptional level, LncRNA binds miRNA competitively with mRNA for protein-coding genes, thus relieving the inhibitory effect of miRNA on protein-coding genes to promote expression of these genes (Kartha and Subramanian, 2014). We analyzed the expression patterns of LncRNAs differentially expressed among different breeds, and constructed an LncRNA-miRNA-mRNA interaction network related to skeletal muscle development. A total of 52 LncRNAs related to muscle development were identified, with nine up-regulated and five down-regulated. All of these LncRNAs contain one or more putative miRNA binding sites with 48 LncRNAs predicted to interact with one to two miRNAs related to muscle development, and other LncRNAs predicted to interact with more than three miRNAs, such as *LOC525506*, *LOC540051*, *LOC514189*. With the miRNA seed sequences, LncRNAs can bind to miRNA to act as a sponge, preventing miRNA from binding to its target mRNA. As a classic example, M. Cesana (Marcella Davide et al., 2011) confirmed that linc-MD1, a long non-coding RNA specifically expressed during myoblast differentiation, regulates the expression of muscle specific transcription regulators *MAML1* and *MEF2C* through *miR-133* and *miR-135*. In particular, *LOC525506* may be involved in the development of beef skeletal muscle by regulating *miR-1*, *miR-23a*, *miR-378*, *let-7*, *miR-483*, and *miR-21*. This is the most predicted LncRNA, but little is known about its expression and function. Both *LOC112447073* and *LOC104975788* are involved in the skeletal muscle development of beef cattle by interaction with *miR-103*. More interestingly, some target genes of *LOC112447073*, *LOC104975788*, and *LOC101903367* directly regulate the development of muscle fiber and maintain the stability and development of muscle, such as *miR-103*, *miR-133a*, *miR-145*, *MEF2C*, myocardin, and *Pax7* (Richard et al., 2008; Mathew et al., 2011). Choi identified 11 LncRNAs in bovine transcripts by studying skeletal muscle and adipose tissue of Korean cattle, with four related to muscle function (Jae et al., 2018). These data provide new insight into the role of LncRNA in muscle development and enrich the existing LncRNA mammalian skeletal muscle resources.

The second to eighth bases of the sequence at the 5 ′end of miRNA is called the seed sequence. Complete complementarity of this sequence and that of the target gene indicates the potential for binding of the target gene by miRNA (Lewis et al., 2005; Andrew et al., 2007). We predicted the miRNA binding sites of functional LncRNAs that may be related to skeletal muscle development. Using a double luciferase test, we found a recognition site of *bta-mir-133a* in the sequence of *LOC104975788* and a binding site for *bta-mir-133a* in the downstream target gene *Pax7*. We speculate that *LOC104975788* may be involved in the regulation of skeletal muscle development by competitively inhibiting the expression of the target gene *Pax7* through binding of *bta-miR-133a*. Future work should test this proposed regulatory mechanism.

CONCLUSION

The expression patterns of LncRNAs in skeletal muscle of Shandong black cattle and Luxi cattle were elucidated by RNA sequence analysis, and the LncRNAs that may be involved in the skeletal muscle development of beef cattle were identified. The results allowed construction of interaction networks of LncRNAs-miRNA-mRNA regulated by muscle biology. We speculate that LOC104975788 may be involved in the regulation of skeletal muscle development by competitively inhibiting the expression of the target gene Pax7 through binding of bta-miR-133a.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. The RNA sequencing data has been uploaded to the NCBI GEO database, and can be retrieved using the following accession numbers: GSM4904154, GSM4904155, 623 GSM4904156, GSM4904157, GSM4904158, GSM4904159.

ETHICS STATEMENT

The animal study was reviewed and approved by Biological Studies Animal Care and Use Committee, Shandong Province, China.

AUTHOR CONTRIBUTIONS

YD, XB, KY, RL and MH designed the study. RL and MH conducted the experiment. RL, MH, XL and XB performed and collected the data. RL analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by Shandong modern agricultural industry technology system cattle industry innovation team (668-2216030). The team participated in the collection and analysis of data.

ACKNOWLEDGMENTS

The authors would like to thank Shanghai Sangon Biotechnology Co., Ltd. and Beijing AnnoroadGene Co., Ltd. for their assistance in the original data processing and related bioinformatics analysis for RNA-sequencing and GC-MS. The

authors also thank Liu for her meticulous revision of language. More importance, special thanks to Yue Liu for his contribution to data processing. Additionally, the authors would like to particularly thank her parents and Dong, for their continuing love and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.849399/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting Effective microRNA Target Sites in Mammalian mRNAs. *eLife* 12, e05005. doi:10.7554/eLife.05005

Anders, S., and Huber, W. (2010). Differential Expression Analysis for Sequence Count Data. *Genome Biol.* 11, R106. doi:10.1186/gb-2010-11-10-r106

Andersson, L. (2009). Genome-wide Association Analysis in Domestic Animals: a Powerful Approach for Genetic Dissection of Trait Loci. *Genetica* 136 (2), 341–349. doi:10.1007/s10709-008-9312-4

Andrew, G., Kyle, K. F., Wendy, K. J., Philip, G., Lee, P. L., and David, P. B. (2007). Micro RNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cel* 27, 91–105. doi:10.1016/j.molcel.2007.06.017

Ballarino, M., Cazzella, V., D'Andrea, D., Grassi, L., Bisceglie, L., Cipriano, A., et al. (2015). Novel Long Noncoding RNAs (lncRNAs) in Myogenesis: a miR-31 Overlapping lncRNA Transcript Controls Myoblast Differentiation. *Mol. Cel Biol* 35 (4), 728–736. doi:10.1128/mcb.01394-14

Billerey, C., Boussaha, M., Esquerré, D., Rebours, E., Djari, A., Meersseman, C., et al. (2014). Identification of Large Intergenic Non-coding RNAs in Bovine Muscle Using Next-Generation Transcriptomic Sequencing. *BMC genomics* 15 (1), 499. doi:10.1186/1471-2164-15-499

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-i: Intelligent Prioritization and Exploratory Visualization of Biological Functions for Gene Enrichment Analysis. *Nucleic Acids Res.* 49, W317–W325. doi:10.1093/nar/gkab447

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative Annotation of Human Large Intergenic Noncoding RNAs Reveals Global Properties and Specific Subclasses. *Genes Dev.* 25, 1915–1927. doi:10.1101/gad.17446611

Cai, R., Zhang, Q., Wang, Y., Yong, W., Zhao, R., and Pang, W. (2021). Lnc-ORA Interacts with microRNA-532-3p and IGF2BP2 to Inhibit Skeletal Muscle Myogenesis. *J. Biol. Chem.* 296, 100376. doi:10.1016/j.jbc.2021.100376

Cai, W., Li, C., Liu, S., Zhou, C., Yin, H., Song, J., et al. (2018). Genome Wide Identification of Novel Long Non-coding RNAs and Their Potential Associations with Milk Proteins in Chinese Holstein Cows. *Front. Genet.* 9, 281. doi:10.3389/fgene.2018.00281

Cai, Y., He, J., and Zhang, D. (2015). Long Noncoding RNA CCAT2 Promotes Breast Tumor Growth by Regulating the Wnt Signaling Pathway. *Onco Targets Ther.* 8, 2657–2664. doi:10.2147/OTT.S90485

Cao, T., Shi, L. G., and Zhang, L. L. (2014). Comparative Study on Fetal Muscle Fiber of Wuzhishan Pig and Changbai Pig during 65d Gestation. *J. Anim. Ecol.* 35, 37–40.

Chen, S., Zhou., Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560

Cohen, B. A., Mitra, R. D., Hughes, J. D., and Church, G. M. (2000). A Computational Analysis of Whole-Genome Expression Data Reveals Chromosomal Domains of Gene Expression. *Nat. Genet.* 26, 183–186. doi:10.1038/79896

Dacia, D., Alessio, C., Monica, B., Julie, M., Davide, M., Johannes, D., et al. (2018). Irene, B. The Long Non-coding RNA Lnc-31 Interacts with Rock1 mRNA and Mediates its YB-1-dependent Translation. *Cel Rep.* 23 (3), 733–740. doi:10.1016/j.celrep.2018.03.101

Dylan, S., Katarzyna, G., Tina, R., Svetlana, O., Thomas, B., Tamas, D., et al. (2008). Specific Requirements of MRFs for the Expression of Muscle Specific microRNAs, miR-1, miR-206 and miR-133. *Dev. Biol.* 321 (2), 491–499. doi:10.1016/j.ydbio.2008.06.019

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA Targets in Drosophila. *Genome Biol.* 5 (1), R1. doi:10.1186/gb-2003-5-1-r1

Federica, A., Vijay, A. R., Andrei, E. M., Sivapriya, K. V., and Anthony, T. V. (2019). Long Noncoding RNAs in Host-Pathogen Interactions. *Trends Immunol.* 40, 492–510. doi:10.1016/j.it.2019.04.001

Hendricks, H. B., Lafferty, D. T., Aberle, E. D., Judge, M. D., and Forrest, J. C. (1971). Relation of Porcine Muscle Fiber Type and Size to Postmortem Shortening. *Anim. Sci.* 32, 57–61. doi:10.2527/jas1971.32157x

Ito, E. A., Katahira, I., Vicente, F. F. d. R., Pereira, L. F. P., and Lopes, F. M. (2018). BASiNET-BiologicAl Sequences NETwork: a Case Study on Coding and Non-coding RNAs Identification. *Nucleic Acids Res.* 46 (16), e96. doi:10.1093/nar/gky462

Jae, Y. C., Donghyun, S., Hyun, J. L., and Jae, D. O. (2018). Comparison of Long Noncoding RNA between Muscles and Adipose Tissues in Hanwoo Beef Cattle. *Anim. Cell Syst. official Publ. Zoolog. Soc. Korea* 23, 1–9. doi:10.1080/19768354.2018.1512522

James, A. B., Ernst, J. W., John, S. M., John, L. R., and Guy, B. (2015). Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron* 88, 861–877. doi:10.1016/j.neuron.2015.09.045

Jing, Z., Toshiro, K. O., Johnny, T. K., Yuya, O., Daniel, J. G., and Kavitha, S. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-Seq. *Mol. Cel* 40, 939–953. doi:10.1016/j.molcel.2010.12.011

Kanehisa, M., and Goto, S. K. E. G. G. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 128 (1), 27–30. doi:10.1093/nar/28.1.27

Kartha, R. V., and Subramanian, S. (2014). Competing Endogenous RNAs (ceRNAs): New Entrants to the Intricacies of Gene Regulation. *Front. Genet.* 5, 8. doi:10.3389/fgene.2014.00008

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2. *Genome Biol.* 20, 278. doi:10.1186/s13059-019-1910-1

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining Clusters from a Hierarchical Cluster Tree: The Dynamic Tree Cut Package for R. *Bioinform* 24, 719–720. doi:10.1093/bioinformatics/btm563

Leonardo, S., Laura, P., Yvonne, T., Lev, K., and Pier, P. P. (2011). A ceRNA Hypothesis: the Rosetta Stone of a Hidden RNA Language? *Cell* 146, 353–358. doi:10.1016/j.cell.2011.07.014

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes Are MicroRNA Targets. *Cell* 120, 15–20. doi:10.1016/j.cell.2004.12.035

Li, D., Bao, P., Yin, Z., Sun, L., Feng, J., He, Z., et al. (2018). Exploration of the Involvement of LncRNA in HIV-Associated Encephalitis Using Bioinformatics. *PeerJ* 6, e5721. doi:10.7717/peerj.5721

Liu, R., Liu, X., Bai, X., Xiao, C., and Dong, Y. (2020). Different Expression of Lipid Metabolism-Related Genes in Shandong Black Cattle and Luxi Cattle Based on Transcriptome Analysis[J]. *Scientific Rep.* 10, 1. doi:10.1038/s41598-020-79086-4

Liu, R., Liu, X., Bai, X., Xiao, C., and Dong, Y. (2018). Screening of Skeletal Muscle Differential Genes Based on Transcriptome. *North China Agric. J.* 33, 64–72. doi:10.3389/fvets.2022.831519

Lu, W. (2015). Feeding and Management Techniques of Luxi Cattle in Fattening Period. *Beijing Agric.* 25 (2015), 130–131.

Luo, H., Bu, D., Sun, L., Fang, S., Liu, Z., and Zhao, Y. (2017). Identification and Function Annotation of Long Intervening Noncoding RNAs. *Brief Bioinform* 18 (5), 789–797. doi:10.1093/bib/bbw046

Mahmoudi, B., Fayazi, J., Roshanfekr, H., Sari, M., and Bakhtiarizadeh, M. R. (2020). Genome-wide Identification and Characterization of Novel Long Non-coding RNA in Ruminal Tissue Affected with Sub-acute Ruminal Acidosis from Holstein Cattle. *Vet. Res. Commun.* 44 (1), 19–27. doi:10.1007/s11259-020-09769-w

Marcella, C., Davide, C., Ivano, L., Tiziana, S., Olga, S., Mauro, C., et al. (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endoge-Nous RNA. *Cell* 147 (2), 358–369. doi:10.1016/j.cell.2011.09.028

Mathew, S. J., Hansen, J. M., Merrell, A. J., Murphy, M. M., Lawson, J. A., Hutcheson, D. A., et al. (2011). Connective Tissue Fibroblasts and Tcf4 Regulate Myogenesis. *Dev. (Camb. Engl.* 138, 371–384. doi:10.1242/dev.057463

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33 (3), 290–295. doi:10.1038/nbt.3122

Peters, E., van der Linde, S., Vogel, I., Haroon, M., Offringa, C., de Wit, G., et al. (2017). IGF-1 Attenuates Hypoxia-Induced Atrophy but Inhibits Myoglobin Expression in C2C12 Skeletal Muscle Myotubes. *Ijms* 18, 1889. doi:10.3390/ijms18091889

Richard, D. B., Ariel, J. C., Amy, S. M., and Emanuel, E. S. (2008). Interaction with the IQ3 Motif of Myosin-10 Is Required for Calmodulin-like Protein-dependent Filopodial Extension. *Febs Lett.* 582, 2377–2381. doi:10.1016/j.febslet.2008.05.051

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Steinbacher, P., Haslett, J. R., Sänger, A. M., and Stoiber, W. (2006). Evolution of Myogenesis in Fish: a sturgeon View of the Mechanisms of Muscle Development. *Anat. Embryol.* 211, 311–322. doi:10.1007/s00429-006-0082-4

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing Sequence Intrinsic Composition to Classify Protein-Coding and Long Non-coding Transcripts. *Nucleic Acids Res.* 41, e166. doi:10.1093/nar/gkt646

Sun, X., Li, M., Sun, Y., Cai, H., Lan, X., Huang, Y., et al. (2016). The Developmental Transcriptome Sequencing of Bovine Skeletal Muscle Reveals a Long Noncoding RNA, lncMD, Promotes Muscle Differentiation by Sponging miR-125b. *Biochim Biophys Acta.* 1863 (11), 2835–2845. doi:10.1016/j.bbamcr.2016.08.014

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R Package for Identifying Differentially Expressed Genes from RNA-Seq Data. *Bioinformatics* 26, 136–138. doi:10.1093/bioinformatics/btp612

Xu, J. W., Zheng, L., Li, L. J., Yao, Y. F., Hua, H., Yang, S. Z., et al. (2018). Novel Copy Number Variation of the KLF3 Gene is Associated With Growth Traits in Beef Cattle. *Gene* 680, 99–104. doi:10.1016/j.gene.2018.08.040

Xu, M., Chen, X., Chen, D., Yu, B., Li, M., He, J., et al. (2020). Regulation of Skeletal Myogenesis by microRNAs. *J. Cel Physiol* 235 (1), 87–104. doi:10.1002/jcp.28986

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-y., Feng, Y., et al. (2013). Genetic Programs in Human and Mouse Early Embryos Revealed by Single-Cell RNA Sequencing. *Nature* 500, 593–597. doi:10.1038/nature12364

Ying-hui, L., Qi, Z., Meng-hua, S., Lu, Z., Li-na, X., Yun-hai, Z., et al. (2019). Comprehensive Analysis of LncRNA Reveals the Temporal-specific Module of Goat Skeletal Muscle Development. *Int. J. Mol. Sci.* 20 (16), 3950. doi:10.3390/ijms20163950

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biol.* 11, R14. doi:10.1186/gb-2010-11-2-r14

Yu-ying, L., Xiao-na, C., Hao, S., and Hua-ting, W. (2017). Long Non-coding RNAs in the Regulation of Skeletal Myogenesis and Muscle Diseases. *Cancer Lett.* 417, 58–64. doi:10.1016/j.canlet.2017.12.015

Yuan, G., Shi-peng, L., Zhen-yu, L., Zi-hui, Z., Fei, W., Yong-zhen, H., et al. (2019). Analysis of Long Non-coding RNA and mRNA Expression Profiling in Immature and Mature Bovine (*Bos taurus*) Testes. *Front. Genet.* 10, 646. doi:10.3389/fgene.2019.00646

Yue, Y., Jin, C., Chen, M., Zhang, L., Liu, X., Ma, W., et al. (2017). A lncRNA Promotes Myoblast Proliferation by Up-Regulating GH1. In Vitro *Cell.Dev.Biol.-Animal* 53 (8), 699–705. doi:10.1007/s11626-017-0180-z

Zhang, S., Knight, T. J., Stalder, K. J., Goodwin, R. N., Lonergan, S. M., and Beitz, D. C. (2007). Effects of Breed, Sex, and Halothane Genotype on Fatty Acid Composition of Pork Longissimus Muscle1. *J. Anim. Sci.* 85 (3), 583–591. doi:10.2527/jas.2006-239

Zhou, Z. J. (2010). Relationship between Muscle Histological Characteristics and Meat Quality. *Breed. Technol. consultant* 8, 57–58. doi:10.3390/nu401000

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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 07 June 2022
doi: 10.3389/fgene.2022.832702

Genetics of Base Coat Colour Variations and Coat Colour-Patterns of the South African Nguni Cattle Investigated Using High-Density SNP Genotypes

Langelihle Mbali Kunene[1], *Farai Catherine Muchadeyi*[2], *Khanyisile Hadebe*[2], *Gábor Mészáros*[3], *Johann Sölkner*[3], *Trevor Dugmore*[4] *and Edgar Farai Dzomba*[1]*

[1]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Scottsville, South Africa, [2]Agricultural Research Council, Biotechnology Platform, Onderstepoort, South Africa, [3]Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria, [4]KZN Department of Agriculture and Rural Development, Pietermaritzburg, South Africa

Nguni cattle are a Sanga type breed with mixed *B. taurus* and *B. indicus* ancestry and proven resistance to ticks, diseases and other harsh conditions of the African geographical landscape. The multi-coloured Nguni coats have found a niche market in the leather industry leading to breeding objectives towards the promotion of such diversity. However, there is limited studies on the genomic architecture underlying the coat colour and patterns hampering any potential breeding and improvement of such trait. This study investigated the genetics of base coat colour, colour-sidedness and the white forehead stripe in Nguni cattle using coat colour phenotyped Nguni cattle and Illumina Bovine HD (770K) genotypes. Base coat colour phenotypes were categorised into eumelanin (n = 45) and pheomelanin (n = 19). Animals were categorised into either colour-sided (n = 46) or non-colour-sided (n = 94) and similarly into presence (n = 15) or absence (n = 67) of white forehead stripe. Genome-wide association tests were conducted using 622,103 quality controlled SNPs and the Efficient Mixed Model Association eXpedited method (EMMAX) implemented in Golden Helix SNP Variation Suite. The genome-wide association studies for base coat colour (eumelanin vs. pheomelanin) resulted into four indicative SNPs on BTA18 and a well-known gene, MC1R, was observed within 1 MB from the indicative SNPs ($p < 0.00001$) and found to play a role in the melanogenesis (core pathway for melanin production) and the MAPK signalling pathway. GWAS for colour-sidedness resulted in four indicative SNPs, none of which were in close proximity to the KIT candidate gene known for colour-sidedness. GWAS for the white forehead stripe resulted in 17 indicative SNPs on BTA6. Four genes MAPK10, EFNA5, PPP2R3C and PAK1 were found to be associated with the white forehead stripe and were part of the MAPK, adrenergic and Wnt signalling pathways that are synergistically associated with the synthesis of melanin. Overall, our results prove prior knowledge of the role of MC1R in base coat colours in cattle and suggested a different genetic mechanism for forehead stripe phenotypes in Nguni cattle.

Keywords: genome-wide association study, nguni cattle, coat colour, colour-sidedness, white forehead stripe

OPEN ACCESS

Edited by:
Filippo Biscarini, National Research Council (CNR), Italy

Reviewed by:
Vincenzo Landi, University of Bari Aldo Moro, Italy
Salvatore Mastrangelo, University of Palermo, Italy

****Correspondence:***
Edgar Farai Dzomba
Dzomba@ukzn.ac.za

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

Received: *10 December 2021*
Accepted: *25 March 2022*
Published: *07 June 2022*

Citation:
Kunene LM, Muchadeyi FC, Hadebe K, Mészáros G, Sölkner J, Dugmore T and Dzomba EF (2022) Genetics of Base Coat Colour Variations and Coat Colour-Patterns of the South African Nguni Cattle Investigated Using High-Density SNP Genotypes. Front. Genet. 13:832702. doi: 10.3389/fgene.2022.832702

INTRODUCTION

The Nguni cattle are a Sanga breed native to South Africa (Makina et al., 2014) and farmed in other countries worldwide. This breed is of great importance to South Africa due to its ability to survive under very harsh climatic conditions coupled with ticks and tick borne diseases (Mapiye et al., 2007). Nguni cattle are thought to have survived through strong selection during their migration from North Africa to Southern Africa, leading to phenotypic changes in the animals and making them capable of withstanding the harsh environments of the country (Bester et al., 2003). Traits displayed by the Nguni cattle in order to survive in extreme climate with limited feed and exposure to tick infestations are relevant to resource limited smallholder farmers (Bester et al., 2003) who rely on Nguni cattle for food security (milk and meat) and by-products (hides, skin and horns) that they use to generate income (Musemwa et al., 2010). Nguni cattle have diverse coat colours with different patterns, which, in the previous years, led to the belief that they were admixed with other breeds and were thus considered inferior to exotic breeds of uniform and standard coat colour patterns (Bester et al., 2003). Three main coat colours exist in Nguni cattle which are black, brown and red and the other coat colours that are found in the breed are modifications of these (Olson 1999).

The unique and diverse Nguni cattle coat colours are of cultural significance particularly to the Nguni ethnic people (Oosthuizen 1996). Because of their cultural values, there is a strong bond between the Nguni cattle and Nguni people leading to instances such as poetic naming of the animals of different physical attributes with inspiration arising from the beautiful coat colour and patterns on their hides (Oosthuizen 1996). For example, the, white Nguni cattle with black spots reminds the Nguni people of flies in the buttermilk and is named "Inasenezimbukane" while black cattle with a white pattern resembling a spitting cobra are named "iFezi" (Abundant Herd, 2003). Black cattle for instance are used when a king is deceased and during celebrations such as rain ceremonies (Oosthuizen 1996). In addition to the socio-cultural roles, Nguni cattle hides have an economic value in the manufacturing of leather products for the automotive, furniture and fashion industries (Brits 2014). Cattle hides are exported overseas for car upholstery and are locally used for the manufacture of furniture, handbags, horse-riding equipment and clothing garments (footwear and belts) (Brits, 2014). Coupled to these socio-cultural and economic roles, coat colour in livestock including cattle is also important because as it has been observed to be associated with other traits of economic importance (Brits 2014). Different coat colours in Nguni cattle are associated with resistance and susceptibility to pests/tick infestation and diseases that can result in unwanted culling or death of certain animals (Mapholi et al., 2014). Because of these socio-cultural, economic and adaptive associations, farmers across commercial and smallholder sectors have been observed to have preferences for certain coat colour patterns. In a study on trait preferences for smallholder Nguni bulls, coat colour was the sixth trait of interest after aggression and mating behaviour, tick and disease resistance, body condition score, scrotal circumference and body size and conformation (Tada et al., 2013). Coat colour was also ranked as the fifth preferred trait after tick and disease resistance, reproductive efficiency, body condition score and body size and conformation in Nguni cows (Tada et al., 2013). In order to inform selection and breed improvement programs, understanding the genetics of coat colour becomes a prerequisite.

The mechanism of coat colouration (melanogenesis) is determined by the melanin found within the melanosomes in the melanocytes (Marks and Seabra 2001) and can either be categorised into either eumelanin and pheomelanin (Olson 1999). Eumelanin is entirely responsible for black to brown pigmentation of the skin and coat colours while pheomelanin is responsible for red to yellow colouration of mammal coat colours (Olson 1999). The production of eumelanin relies on the stimulation of a G-protein coupled receptor called melanocyte-specific melanocortin receptor (MC1R) by agonists α-melanocyte-stimulating hormone (α-MSH) and adrenocorticotropic hormone (ACTH) (Millington 2006). The antagonist of the MC1R, agouti signalling protein (ASIP) controls the production of eumelanin by stimulating pheomelanogenesis (Videira et al., 2013). Melanogenesis is a process where melanin is synthesized in the presence of an amino acid called tyrosine (TYR). Mutations of the TYR amino acid results in unpigmented coat colours and the mutation is termed albinism (Cieslak et al., 2011). When melanogenesis is disrupted, leucism can occur. Leucism is a condition where there is lack of pigmentation due to the absence of melanocytes on the animals' skin and/or coat and results in what is seen as white spots or patches on a solid base colour. These phenotypes include the white forehead stripe on the cattle's heads and colour-sidedness patterns as observed in the Texas Longhorn, Florida Cracker, English Longhorn, Scandinavian cattle and other African breeds (Olson 1999). Mutations of the KIT and KITL genes have been reported to have association with white spotting and roan phenotypes in Belgian Blue and Shorthorn cattle (Seitz et al., 1999). However, other studies such as that conducted by Hofstetter et al. (2019) reported BTA22, which harbour the MITF gene to be associated with the white phenotype in Brown Swiss cattle. According to Bennett and Lamoreux (2003), a large number of genes affects coat colour and colour patterns through different mechanisms (Bennett and Lamoreux 2003). According to Fontanesi et al. (2010) genetic heterogeneity is prevalent in coat colour phenotypes in cattle and other livestock species.

The Illumina HD bovine Beadchip (Illumina, Inc., San Diego, CA, United States) was developed by a multi-institutional consortium, Bovine HapMap, and contains 777,962 SNP markers. The taurine, zebu and composite breeds were used in the development of the array (Harris et al., 2011) and the chip has found utility in population genetics (Meseret et al., 2020), genomic evaluation and selection (Uemoto et al., 2015), copy number variation (Jiang et al., 2013) and case-control genome-wide association studies (GWAS) (Riffert 2015). Riffert (2015) observed the MC1R gene to be associated with the dark coat colours of the Brazilian Gir cattle. Colour-sidedness was reported to be a result of the copy number variant (CNV)

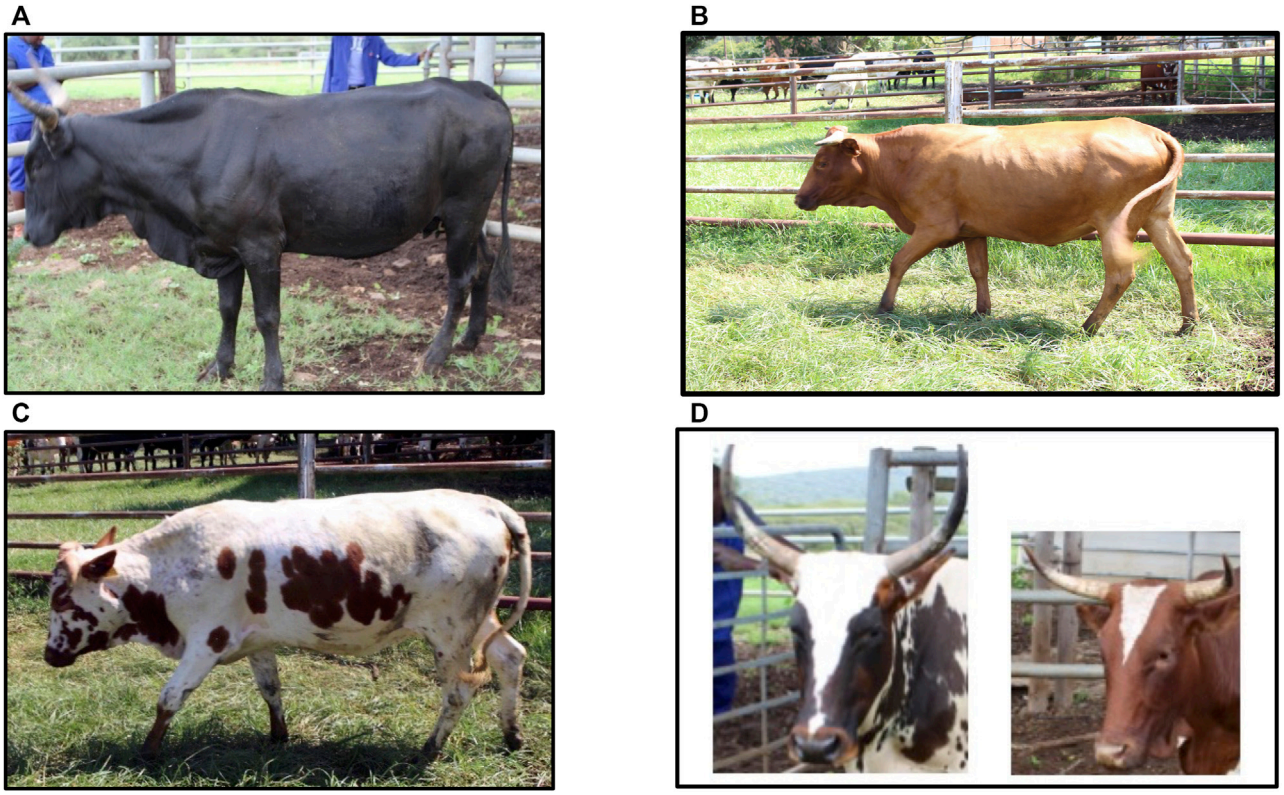


FIGURE 1 | Nguni cattle depicting base coat colours and the colour-sidedness and the white forehead stripe pattern. **(A)** Black (eumelanin) coat coloured Nguni cattle, **(B)** Red (pheomelanin) coat coloured Nguni cattle, **(C)** A colour-sided Nguni cattle, **(D)** white forehead stripe observed in Nguni cattle.

mechanism on a 480 kb region that harbours the KIT gene on BTA6 in cattle (Durkin et al., 2012). The Illumina BovineHD Beadchip was used to investigate and identify markers for slick hair coat on three tropical breeds, Senepol, Carora and Romosinuano, the study of which reported S-phase kinase-associated protein 2 (SKP2) and sperm flagellar 2 (SPEF2) as candidate genes for the slick hair phenotypes in these cattle breeds (Huson et al., 2014). The KIT gene was also associated with white base colour and spotting legs of Gir cattle (Riffert 2015). GWAS on facial markings have also been explored in the Fleckvieh cattle and the results showed that the MITF on BTA22 and KITLG on BTA5 were found to be associated with pigmentation patterns on the animal's head (Mészáros et al., 2015).

Despite its socio-cultural and economic importance, very little is known about the genetics of coat colour in Nguni cattle, absence of this information of which hinder efforts for genetic improvement for such traits. The aim of this study was therefore to investigate genomic regions (SNPs and candidate genes) and genetic mechanisms that are associated with the coat colours and patterns in South African Nguni cattle using the Illumina Bovine HD genotypes and Nguni cattle phenotyped for base coat colour, colour-sidedness and presence and absence of white forehead stripes. The Nguni cattle used in this study were from two research populations of Bartlow Combine and Kokstadt in KwaZulu-Natal province and consisted of unrelated animals representative of the Nguni cattle diversity.

MATERIALS AND METHODS

Animal Populations and Coat Colour Phenotypes

A total of 142 Nguni cattle were sampled from two research stations of the Bartlow Combine (n = 78) and Kokstadt (n = 66) in the KwaZulu-Natal region. The Bartlow Combine and Kokstadt research stations are located at the North and Southern parts of KwaZulu-Natal province of South Africa respectively. These two research stations are conservation farms for the Nguni cattle diversity and were chosen as a representative sample of the total Nguni cattle diversity. Using photo images, the coat colour patterns were phenotyped into 1) black, brown or red base coat colours and 2) colour-sidedness and the 3) white forehead stripe patterns as illustrated in **Figure 1**. Base coat colour was classified into two different categories of 1) eumelanin which constituted black (n = 39) and brown (n = 6) animals and 2) pheomelanin with the red (n = 19) animals. Of the 64 cattle used in the base colour analysis, 35 were from the Bartlow Combine station and 29 were from the Kokstadt station. Forty-six (46) colour-sided and 94 non-colour-sided Nguni cattle were phenotyped and used in the colour-sidedness trait analysis. A total of 82 Nguni cattle were phenotyped for the

TABLE 1 | Case-control GWAS tests used for the generation of Manhattan plots.

Trait	# Case	# Control
Base coat colour	Eumelanin (39 Black, 6 Brown)	Pheomelanin (19 Red)
Colour-sidedness	46 Colour-sided	94 Non-colour-sided
White forehead stripe	15 with the white forehead stripe	67 without the forehead stripe

white forehead stripe trait of which 15 animals had the white forehead stripe and 67 did not have the forehead stripe.

Genotyping and Quality Control

Genomic DNA was extracted from the 142 blood samples using the QIAGEN MagAttract HMW DNA Kit (Qiagen Hilden, Germany). The concentration of the extracted DNA was measured using the qubit fluorometer (Life Technologies, United Kingdom) and concentrations greater than 25 ng/μL were accepted for further analysis. DNA integrity was verified on a 1% agarose gel. High-quality DNA was genotyped using the Illumina Bovine HD BeadChip Array kit (Illumina, Inc., San Diego, CA, United States) and in accordance with the Infinium HD Assay Super Manual Protocol at the Agricultural Research Council Biotechnology Platform, South Africa.

SNP quality control was conducted using PLINK version 1.9 (Chang et al., 2015). A genotyping rate for individual animals' assessment was set at 0.95. SNPs were pruned based on a genotyping call rate of 0.95, Hardy-Weinberg < 0.00001, minor allele frequency <0.05. In addition, SNPs without chromosomal positions were removed from the data. The call rate of 0.95 removed 2 animals and 140 animals remained for downstream analysis. The assessment of SNP quality removed 8,463 SNPs for the genotyping call rate, 4,259 SNPs for Hardy-Weinberg and 142,290 SNPs for minor allele frequency. SNPs without chromosomal position (n = 847) were also removed and 622,103 SNPs remained for downstream analysis.

Within-Population Genetic Diversity and Population Structure Analysis

Genetic variation within populations was estimated using four measurements of minor allele frequency (MAF), observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficients (F_{IS}) all of which were analysed using PLINK version 1.9. Genetic relationships between the Nguni cattle from the Bartlow Combine and Kokstadt conservation stations were also analysed using Principal Component Analysis (PCA) and genetic relationship matrix (GRM) on Golden Helix SNP Variation Suite (SVS) version 8.5.0 (Golden Helix, Bozeman, MT, United States). PCA and GRM were also conducted for the animals within each coat colour phenotype (Base coat colour, Forehead stripe and colour sidedness) in order to investigate the genetic relationships between cases and controls for each GWAS.

Genome-wide Association Analysis

Genome-wide association tests for the base coat colours and patterns were done using Golden Helix SVS version 8.5.0 (Golden Helix, Bozeman, MT, United States) using the Efficient Mixed Model Association eXpedited analysis (EMMAX) (Kang et al., 2010). The eumelanin was considered the case compared to pheomelanin; colour-sided animals were considered the case against non-colour-sided animals; and the animals with a white forehead stripe were regarded as the case against those without a forehead stripe, as illustrated in **Table 1**. The mixed model used in this study was:

$$y = X\alpha + K\mu + e$$

where y was the coat colour type or pattern, X was the genotype (622,103 SNP genotypes), α was regarded as the vector of the fixed effect (genotype) while K was the relative kinship matrix, μ was regarded as the unknown random effect, and e was the random error. The Bonferroni correction threshold for multiple tests was used to detect the genome-wide significant SNPs, which were defined as α/K (α = 0.05 and K is the number of SNPs = 622,103). Manhattan plots were generated using the Bonferroni adjusted log of the p-values on Golden Helix SNP Variation Suite (SVS) version 8.5.0 (Golden Helix, Bozeman, MT, United States). Per marker pairwise F_{ST} analysis was conducted using Golden Helix SNP Variation Suite (SVS) version 8.5.0 (Golden Helix, Bozeman, MT, United States) to investigate loci highly differentiating cases from controls.

LD and Population Stratification

In order to guide gene annotation, pair-wise LD (r^2) values were estimated between adjacent SNPs for each autosome and the genome-wide LD over all autosomes in each Nguni herd using Plink v1.09 (Purcell et al., 2007). LD decay LD decay was then analyzed for three maximum distances between SNP pairs, ≥10 kb, ≥100 kb, and ≥1,000 and for each distance, SNP comparisons were binned applying bin sizes of 1 kb, 10 kb, and 100 kb, respectively. The average r^2 for each bin was estimated and plotted against the inter-marker distance.

Furthermore, a quantile-quantile (Q-Q) plot was generated using Golden Helix SNP Variation Suite (SVS) version 8.5.0 (Golden Helix, Bozeman, MT, United States) in order to assess the influence of population stratification on the GWAS. In addition the average genomic inflation factors (λ), defined as the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution drawn from the EMMAX analysis.

SNP Annotation

Candidate genes in close proximity (upto 1 MB downstream or upstream from the significant SNPs ($p < 0.00001$) were identified using the Golden Helix SVS GenomeBrowse using the BTAU5.0.1 (November 2015) reference genome. Pathways for the identified genes ±1 MB from the indicative SNP for each trait

TABLE 2 | Minor allele frequency (MAF), observed ($H_{O)}$ and expected ($H_{E)}$ heterozygosities and, inbreeding coefficient ($F_{IS)\ of\ Nguni\ cattle\ from}$ Bartlow and Kokstadt herds.

Herd	N*	MAF ±SD	H_O ± SD	H_E ± SD	F_{IS} ± SD
Bartlow	73	0.260 ± 0.137	0.361 ± 0.148	0.348 ± 0.135	-0.038 ± 0.037
Kokstadt	67	0.265 ± 0.135	0.365 ± 0.145	0.353 ± 0.131	-0.033 ± 0.060

N = number of animals.

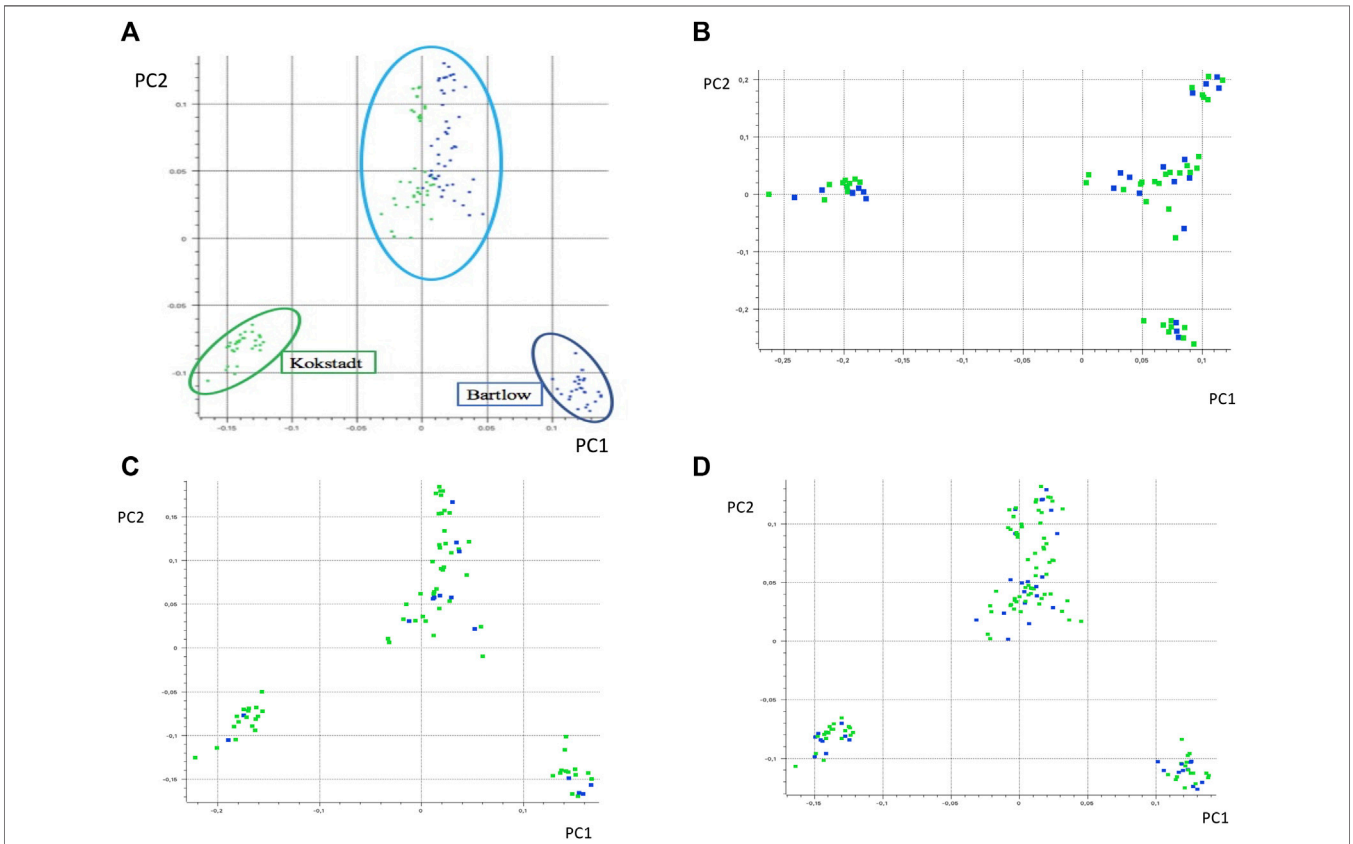


FIGURE 2 | Principal component based clustering of the Nguni cattle **(A)** from two different farms, Bartlow Combine and Kokstadt conservation stations; **(B)** phenotyped for base coat colourl **(C)** phenotyped for absence and presence of forehead strip and **(D)** phenotyped for colour sidedness or non-colour sidedness. Principal component analysis diagram showing principal component one (PC1) and principal component two (PC2) and the variance depicted alongside the axes of the diagram.

(base coat colour, colour-sidedness and white forehead stripe) were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and their association with coat colour patterns inferred.

RESULTS

Genetic Diversity

Minor allele frequencies of 0.260 ± 0.137 and 0.265 ± 0.135 were reported for Bartlow and Kokstadt herds respectively (**Table 2**). The observed (H_O) and expected (H_E) heterozygosity for the Bartlow cattle were 0.361 ± 0.148 and 0.348 ± 0.135 respectively, which were similar to those observed in Kokstadt cattle (H_O = 0.365 ± 0.145, H_E = 0.353 ± 0.131). The observed heterozygosity for Bartlow Combine and Kokstadt populations were greater than the expected heterozygosity ($H_O > H_E$), indicating the possibility of previously isolated populations coming together to form a single population. As a result, slightly negative inbreeding coefficients were observed for both Bartlow (−0.038 ± 0.037) and Kokstadt (−0.033 ± 0.060) cattle.

Population Structure

PC1 and PC2 together accounted for 40% of the total variance, and separated the Nguni cattle into three different clusters (**Figure 2A**). PC1 (21.1% of the variation) separated the Nguni cattle of Kokstadt station (green) from those of Bartlow Combine (blue). PC2 (18.8% of the variation) separated the two clusters of Cattle (Kokstadt and Bartlow combine) from a single cluster that was constituted by animals from the two research stations. Population structure, analysed using PCA analysis explained

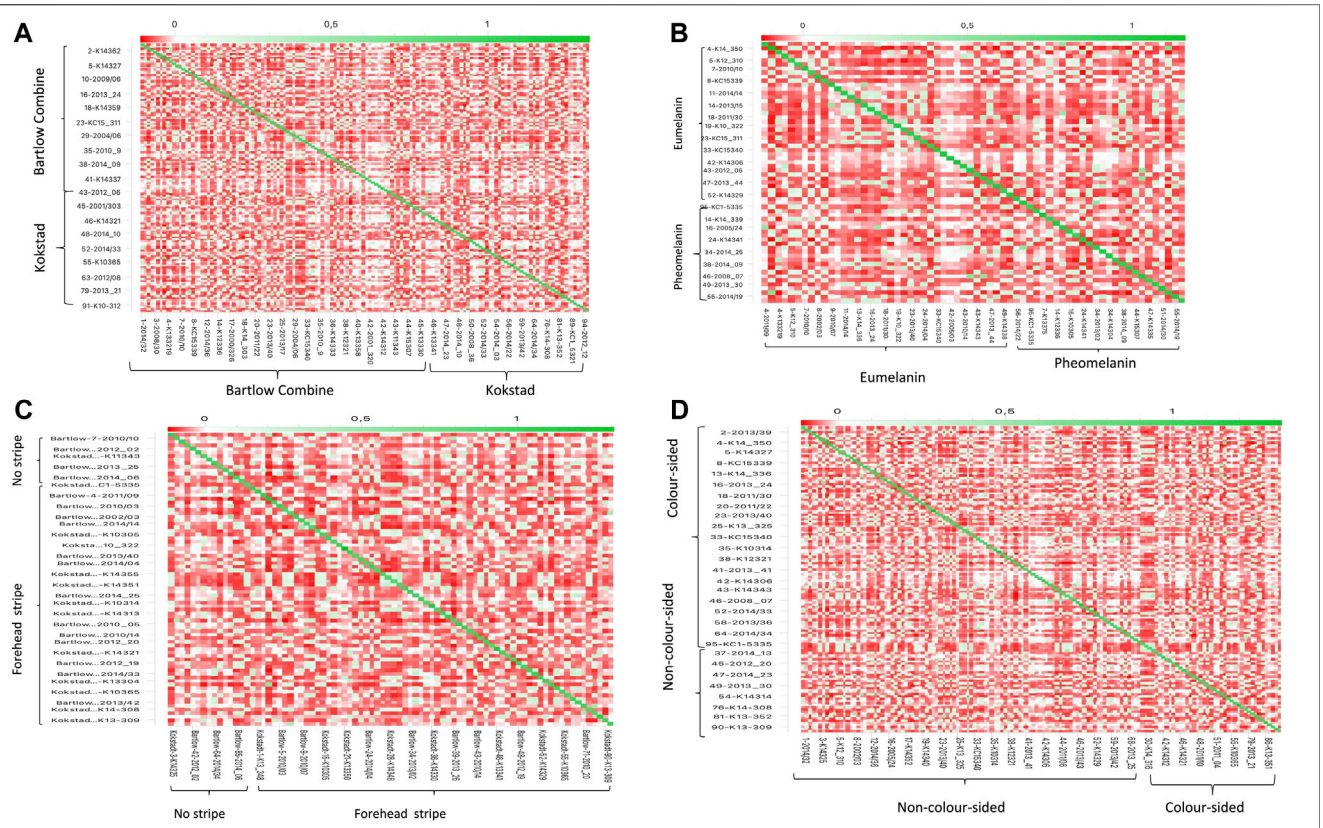


FIGURE 3 | Heat Map of the genomic relationship matrix for **(A)** Bartlow Combine and Kokstadt conservation stations; **(B)** Eumelanin and pheomelanin Nguni cattle **(C)** Nguni cattle with or without forehead stripes and **(D)** colour sidedness and non-colour sided Nguni cattle. The animal identities (IDs) are shown on the *x* and *y*-axis.

11, 13 and 18% of the genetic variation in the animals phenotyped for base coat colour, colour sidedness and Forehead stripe respectively (**Figures 2B–D**). In all three coat colour phenotypes, PC1 and PC2 explained less than 40% of the total genetic variation and animals randomly clustered into three groups with no clear distinction of cases and controls (**Figures 2B–D**). The genetic relationship matrix (GRM) heatmap (**Figure 3**) showed the kinship within and between herds with estimates ranging from 0 (low kinship) to 1 (high kinship). Low Kinship was observed in both herds of Bartlow (0.037 ± 0.013) and Kokstad (0.043 ± 0.013) and Relatively lower (-4.46×10^{-12} ± 0.013) kinship was observed between pairs of animals from the two herds (**Figure 3A**). Similarly low kinship was observed within and between cases and controls of base coat colour (1.47×10^{-13}), presence and absence of forehead stripe (4.05×10^{-15}) and colour sided (2.02×10^{-12}) phenotypes respectively (**Figures 3B–D**).

LD Analysis and Population Stratification

An average genomewide LD (r2) of 0.413 ± 0.219 for Bartlow and 0.402 ± 0.209 for Kokstad was observed. LD decay analysis demonstrated that LD stayed just above 0.2 upto 600 kb and decayed to just about 0.1 at 1000 kb window (**Supplementary Table S1**).

The quantile-quantile plots (Q-Q plots) show that the influence of population stratification was negligible (**Figures 4A–C**). Moreover, the average genomic inflation factors (λ) for the three phenotypes were close to 1 (1.03 for Base coat colour; 0.96 for colour sidedness and 1.09 for Forehead stripe). Both the QQ plots and λ suggested little or no residual population structure effects on the test statistic inflation. Despite the small sample size, the results of GWAS are therefore reasonable and worth further investigation.

Genome-wide Association: SNPs, Genes and Pathways

Base Coat Colour (Eumelanin vs. Pheomelanin)

The GWAS for base coat colour (eumelanin vs. pheomelanin) resulted in two and four significant ($p < 0.000001$) SNPs on BTA10 and BTA18 respectively. (**Figure 5A**). The two SNPs on BTA10 (BovineHD1000021741 and BovineHD1000021742) yielded twenty-three genes, five of which had roles in 7 biological pathways that included the MAPK signalling pathway (**Supplementary Table S1**). The SNPs on BTA18 yielded nine genes (**Supplementary Table S2**). Of these genes, MYLK3 gene was found to have a role in the calcium signalling pathway, cGMP-PKG signalling pathway, vascular smooth muscle contraction, apelin signalling pathway, focal adhesion, platelet activation, regulation of actin cytoskeleton, oxytocin signalling pathway and gastric acid secretion. MC1R gene is involved in the neuroactive ligand-receptor interaction and

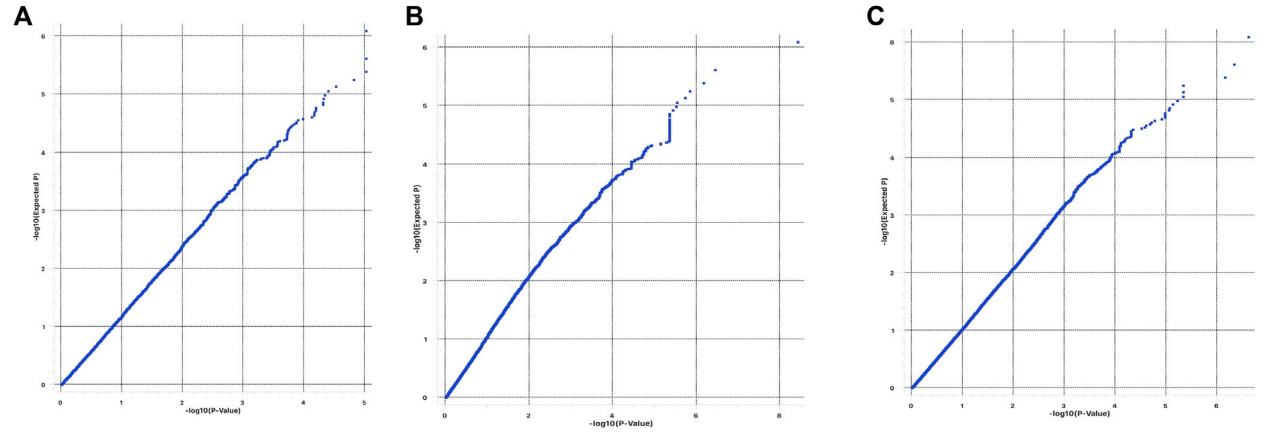


FIGURE 4 | QQplots from GWAS of **(A)** base coat colour; **(B)** Forehead stripe and **(C)** colour sided phenotypes.

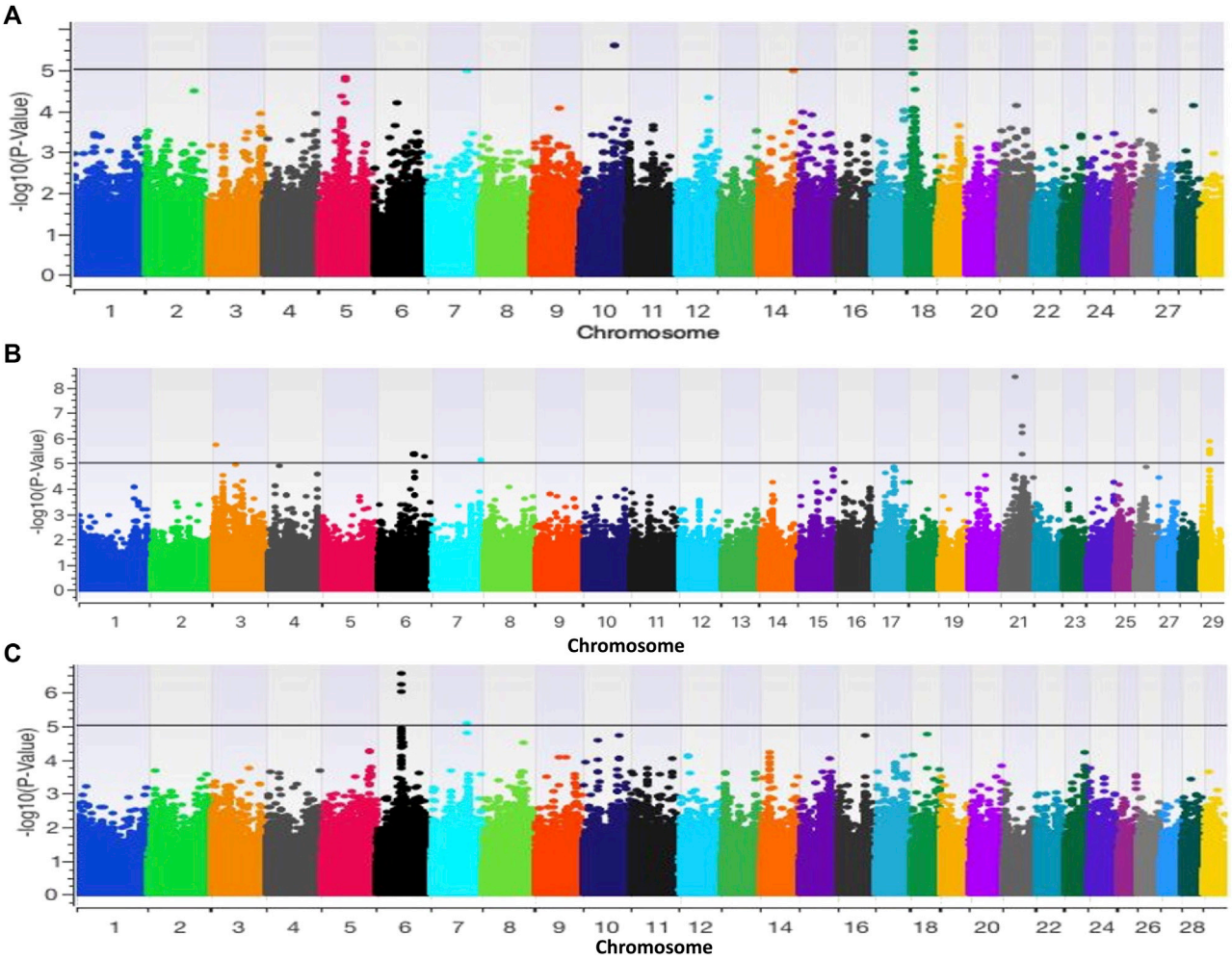


FIGURE 5 | Manhattan plots based on GWAS of **(A)** base coat colour; **(B)** Forehead stripe and **(C)** coloursided phenotypes.

melanogenesis pathways while the FANCA gene was observed in the Fanconi anemia pathway. Two of the identified biological pathways, the MAPK and melanogenesis pathways, are directly linked to the synthesis of coat colour (Fan et al., 2013; Xue et al., 2018).

Colour-Sidedness and White Forehead Stripe

The GWAS for colour-sided versus non-colour-sided animals resulted in four significant SNPs on BTA6 (3 SNPs) and on BTA7 (1 SNP) as illustrated in **Figure 5B**. Genes such as PCDH7, LOC784827, LOC104968962, LOC104968963, LOC785099 were within 1 Mb of significant SNPs and were not associated with any pathways based on KEGG (**Supplementary Table S3**). The GWAS for white forehead stripe resulted in 27 indicative SNPs located on BTA3, 6, 7, 21 and 29 (**Figure 5C**). BTA3 harboured one indicative SNP, BTA6 harboured 16 indicative SNPs, BTA7 harboured 2 indicative SNPs while BTA21 and BTA29 harboured 4 indicative SNPs each. The indicative SNPs and genes associated with white forehead stripe are illustrated in **Supplementary Table S4**. The EFNA5, MAPK10, PAK1 genes were associated with the MAPK signaling pathway, while the PPP2R3C gene was found to be associated with the adrenergic signaling pathway.

Per marker pairwise F_{ST} for the different phenotypes generated similar genomic regions as GWAS analysis for forehead stripe on chromosome 6 and 21 and for colour-sidedness phenotype on chromosome 6 as illustrated in **Supplementary Table S5**.

DISCUSSION

Coat colour and patterns have been widely used to differentiate cattle breeds (Yang et al., 2013), identify and select for other socio-cultural and adaptive traits. Coat colour traits are becoming increasingly important because of the roles they play in the animals' behavioural and adaptation attributes in different environments (Koseniuk et al., 2018). In this study, Nguni cattle from Bartlow Combine and Kokstadt research stations provided an opportunity to study the genetics of the different coat colour patterns observed on this breed. The study focused on base coat colours, colour-sidedness and the white forehead stripe in the hope to improve selection for coat colour in future breeding programmes.

The genetic variability of the Nguni cattle from Bartlow Combine and Kokstadt were similar to each other and higher that an average H_E of 0.28 reported in a genetic diversity study on six South African cattle breeds (Makina et al., 2014). Another study that was conducted using microsatellite markers reported a higher average value for the expected heterozygosity of 0.701 (Sanarana et al., 2016). The high heterozygosity values reported by Sanarana and others could possibly be due to longer periods of natural selection for adaptation and admixture in the studied populations (Ojango et al., 2011). Results for the genetic diversity analysis of the Bartlow Combine and Kokstadt Nguni cattle also showed that the two ecotypes had low inbreeding coefficients demonstrating that the individual animals from Bartlow Combine and Kokstadt were not related and the populations behaved like outbred populations. The Bartlow Combine conservation station was established in 1954 using cattle from the Mhlabisa and Nongoma districts. Today, collaboration between the Department of Agriculture, Forestry and Fisheries (DAFF) and the Industrial Development Corporation (IDC) allows Nguni cattle to be given to different farmers for breeding over a period of 5 years in an effort to increase the numbers of the Nguni cattle in rural communities (de Waal 2014). After the 5-year period, the same number of Nguni heifers and bulls must be returned to the herd. Similar dynamics characterise the Kokstadt herd and this continuous introduction of new stock from different sources to the research stations might explain why the two herds behaved like outbred populations with low inbreeding coefficients. The genetic relationship matrix confirmed the inbreeding coefficient results by reporting low kinship estimates within and between herds.

Population structure was analysed using principal component analysis. The two separate, but compact clusters observed within PC1 indicates that the two populations are genetically distinct from each other whilst a third genetic cluster consisting of a mixture of animals from Bartlow and Kokstadt farms demonstrates presence of admixed animals in line with the population dynamics of the sampled Nguni cattle whereby they were also purchased from various sources countrywide into the research stations where they were then disseminated into emerging farming communities (Kars 1993). The PCA results indicate that there was high genetic diversity within the Nguni cattle populations as well as low genetic diversity between the Bartlow and Kokstadt populations, which is expected of herds of such a broad founder population. Generally Nguni cattle exists as ecotypes reared in different ethnic and geographical communities and the herds at both Bartlow and Kokstadt were drawn from such diverse ecotypes. The 3 distinct clusters are therefore suggestive of the unique and overlapping ecotypes between the two herds. All three PCA and GRM analysis for the different phenotypes (base coat colour, forehead stripe and colour-sidedness) suggested weak population structure and low genetic relations between the animals belonging to the cases and those from the control groups. Based on the PCA and the GRM analysis, it was therefore assumed that the Nguni cattle used in this study were highly unrelated individuals, sampled from diverse gene pools and representative of the diversity prevalent in the breed. The GWAS analysis further assumed absence of population structure within and amongst the different phenotypes. These assumptions were further confirmed by the QQ plots and the average genomic inflation factors that both suggested little or no residual population structure effects on the test statistic.

Genome-wide association studies enable researchers to screen through the entire genome of the species of interest and identify genomic factors/regions associated with the trait of interest. GWAS has recently been used to map candidate genes that were associated with the coat colours of Iranian Markhoz goat (Nazari-Ghadikolaei et al., 2018) and the Brazilian Gir cattle (Riffert 2015), but has not yet been used to gain insight into the

genetics underlying the coat colour variations of the South African Nguni cattle. The GWAS approach first identifies SNPs that are associated with the different phenotypes followed by inference of the genomic regions within the associated SNPs, an analysis that requires defining the window period for which to screen for associated genes. LD analysis of the Nguni herds used in this study, reported an average LD between 0.40 and 0.41 in the two herds in agreement with previous studies that reported an LD of 0.45 in Nguni cattle breed (Makina et al., 2015). LD decay analysis showed a sustained LD that was still above 0.1 at 1000 kb window. The results did suggest as with that of Makina et al. (2015) and others utility of the genomewide SNP panels in GWAS and QTL mapping studies.

The determination of the SNP window within which to screen for genes has varied between studies with no consensus as to the distance cutoff (Brodie et al., 2016). According to Brodie et al. (2016), one way to approach gene identification in GWAS studies is to closely link the probable genes to pathways. This is based on the assumption that SNPs that are associated with the same phenotype are expected to affect the same biological processes and hence the same pathways. We therefore, in this study used that approach of first identifying genes in close proximity and then secondly inferring on the associated phenotypes and their link to coat colour traits.

The genome-wide association analysis for base coat colour (eumelanin vs. pheomelanin) yielded indicative SNPs on BTA10 and BTA18. The SNPs on BTA10 (BovineHD1000021741 and BovineHD1000021742) had no gene within 1 Mb region known to be associated with coat colour. The well-known MC1R gene was however found in the vicinity of the indicative SNPs on BTA18. The MC1R gene harboured by chromosome 18 was first reported to be responsible for the black and brown pigmentations of cattle coat colours in 1995 (Klungland et al., 1995). In this study, three alleles (E^D, E^+, e) of the bovine extension (E) locus, that make the MC1R, were characterized. The MC1R gene was further reported to be associated with the dark coloured coats of Gir cattle (Riffert 2015). The observation of MC1R association with base colour in Nguni cattle in this study further supports the role of this gene in coat colour in multiple cattle breeds. The MC1R gene has been identified and reported in other species such as pigeons' plumage (Ran et al., 2016) and the Chinese sheep coat colour variations (Yang et al., 2013). The study on pigeons plumage identified the MC1R gene to be associated with the trait of black plumage pigeons (Ran et al., 2016). The study on the Chinese sheep found haplotype H1 and H3 of the identified SNPs on the MC1R to be significantly associated with the Minxian Black-fur sheep population (Yang et al., 2013).

ASIP has been regarded as a regulatory protein for the melanin production because when it binds to the MC1R, it suppresses the formation of eumelanin and activates the production of pheomelanin (Videira et al., 2013). The ASIP gene was not detected in our base coat colour GWAS analysis probably because it is an antagonist to MC1R G-protein receptor. Another reason for the undetected ASIP gene in this study could be because there were mutations that occurred on the gene that led to production of eumelanin in the presence of functional and non-mutated genes such as the MC1R and its ligands, α-MSH and ACTH (Almathen et al., 2018). Studies that show mutations to the ASIP gene giving rise to the production of eumelanin have been reported in black Xalda sheep (Royo et al., 2008) and the silver fox (Våge et al., 1997). Another study reported that two missense and one deletion of in exon 4 identified on the ASIP gene were associated with dark coloured coats in alpaca (Feeley et al., 2011).

Two major pathways associated with base coat colour were observed 1) the MAPK and 2) melanogenesis pathways. The MAPK pathway is one of the core pathways that lead to the pigmentation of mammals (D'Mello et al., 2016). In melanosomes, MAPK activates the melanocyte-specific transcription factor Microphthalmia (MITF) that subsequently activates tyrosinase (TYR) and tyrosine-like protein 1 (TYRP1) (Wellbrock et al., 2002). Tyrosinase is important in the catalysis of the melanin precursor tyrosine to 3,4-dihydroxyphenylalanine (L-DOPA) that ultimately produces melanin after a few reactions. The PPP2R5E gene found on BTA10 in the study has role in the PI3K-Akt signalling pathway that regulates the MAPK in B16F10 mouse melanoma cells (Zhou et al., 2017). The AICAR adiponectin via the AMPK pathway has been used to regulate the production of tyrosinase that causes melisma in human and mouse melanocytes thus inactivating the MAPK pathway (Bang et al., 2017). A study conducted on humans showed that pigment cells that were treated with oestrogen had increased cAMP levels that ultimately activated the tyrosinase enzymes to produce eumelanin and cells treated with progesterone had an opposite effect (Natale et al., 2016). The HSPA2 gene found on BTA10 was also involved in the MAPK pathway. The MC1R gene identified on BTA18 in the GWAS results for base coat colours is associated with the core pathway that is responsible for melanogenesis. This gene is G-protein coupled receptor found on the surface of melanin-producing cells that binds α-MSH/ACTH/ASIP thus activating a series of signalling pathways in the cytoplasm of melanocytes (Millington 2006).

In our study, the KIT gene was surprisingly not observed in the vicinity of the indicative SNPs for colour-sided Nguni cattle. According to (Olson 1999), the gene for colour-sidedness (*Cs*), is inherited in its dominant state and the expression of the gene in the heterozygous form brings about variations in those animals. Colour-sidedness trait has been observed in Belgian Blue and Brown Swiss cattle (Durkin et al., 2012). Durkin et al. (2012) reported the colours-sidedness trait in cattle to represent a phenomenon that results from copy number variants (CNVs) particularly duplication and translocation of a segment that harbours the KIT gene from chromosome 6 to chromosome 29 (Durkin et al., 2012). Another study conducted on colour-sided African Nguni cattle reported similar results to Durkin et al. (2012) with two normal and ectopic signals on BTA6 and BTA29, respectively (Szczerbal et al., 2017). A recent study stated that the KIT gene showed significant association to white coat coloured Iranian Markhoz goat (Nazari-Ghadikolaei et al., 2018) and another study reported that mutations experienced by this gene results in white coat colours in many species including cattle (Hanna et al., 2014). SNPs on BTA6 and SNPs within the KIT locus in

this study were further than 1 Mb ruling out the possibility of the KIT gene having an influence on colour-sidedness in Nguni cattle. Our results were however similar to another study conducted on colour-sided cattle that could not find significant associations between KIT linked SNPs on BTA6 and colour-sided phenotypes. Mastrangelo et al. (2019), for example reported PLK2 on BTA20 as a potential candidate gene in colour sidedness phenotype in Cinisara cattle.

Besides the possibility of other genetic mechanisms other than those impacted by KIT gene, confounding known and unknown population structures implicate GWAS results and when not accounted for may lead to false positives and false negatives (Zhang and Pan 2015). Also, phenotypic complexities can interfere with determining the SNPs associated with the trait of interest on our GWAS analysis. The Nguni coat colour patterns are complex, which made phenotyping for colour sidedness with precision and objectivity challenging in this study. This included the existence of multiple phenotypes in a single animal i.e., colour-sidedness and forehead stripes and other phenotypes not reported in this study. It is therefore recommended that more comprehensive phenotyping, with a larger sample size and diverse colour sided patterns is conducted before concluding on the genetics of colour sidedness in this breed.

When melanogenesis is disrupted, leucism can occur. Leucism is a condition where there is lack of pigmentation due to the absence of melanocytes on the animals' skin and/or coat colours and results in what is seen as white spots or patches on a solid base colour. These phenotypes include the white forehead stripe on the cattle's heads and white patterns observed on the animals' body and head. White phenotypes can also be desirable traits to some farmers or breeders in hot humid environments because animals can tolerate heat and function at their optimal. Nguni cattle have been reported to be an optimal breed for the production of beef and have immunity against diseases (Mapholi et al., 2014). It is thus believed that the diversity of coat colours and myriad of coat colour patterns have played a role in some of its adaptive traits. The GWAS conducted for the white forehead stripe resulted in indicative SNPs on BTA6, 7, 21 and 29. The KIT gene was not found in the vicinity of the indicative SNPs. However, the four genes MAPK10, EFNA5, PPP2R3C and PAK1 harboured on BTA6, 7, 21 and 29, respectively were found to be closely associated with the melanogenesis. The MAPK10, EFNA5 and PAK1 are associated with the MAPK signaling pathway while MAPK10 is also found to be associated with the Wnt signaling pathway. The Wnt signaling pathway plays a critical role in the developmental and differentiation stages of melanocytes (D'Mello et al., 2016). Wnt1 and Wnt3a are responsible for the increment of melanocytes and differentiation of neural crest cells to melanocytes in the presence of beta-catenin, respectively (Dunn et al., 2005). When beta-catenin is bound to the Wnt proteins, it regulates MITF transcription via LEF/TCF transcription factors (Steingrímsson et al., 2004). The MITF gene is activated through the MAPK pathway and this gene is responsible for the regulation of tyrosinase (TYR) that activates melanogenesis in melanocytes. The PPP2R3C gene is associated with the adrenergic signaling pathway, which is also important in melanogenesis pathway. When the norepinephrine is bound to the adrenergic receptors, the inositol triphosphate/diacylgycerol that controls the release of calcium is activated (Park et al., 2009). Subsequently, the PKC-beta is activated thus activating TYR by phosphorylation in cells to commence with melanin synthesis. The TECRL gene has been previously found to have associations with pigmentation patterns of the Fleckvieh cattle's head (O'brien et al., 2014). Literature reports that the TECRL gene also plays a role in lipid chemical reactions and signaling pathways (O'brien et al., 2014).

Overall the GWAS on presence of white forehead stripe in Nguni cattle demonstrated role of different genes at play that impact pathways known to have an influence on melanogenesis and therefore pigmentation in vertebrates. This could imply either genetic heterogeneity of coat colour and patterns phenotype or complexity of coat colour phenotypes in Nguni cattle leading to confounding effects. A more comprehensive study with intensive phenotyping of white forehead stripe is therefore required.

CONCLUSION

Overall, the GWAS for coat colour patterns in the South African Nguni cattle shed light into the genetics of coat colour patterns in this breed. The GWAS for base coat colour reported SNPs, genes and pathways similar to those from other studies, with the observation of MC1R gene, which is well known to be associated with coat colour in cattle and other species. This gene was also observed to be linked to the well-known MAPK and melanogenesis pathways. Although, the GWAS for colour-sidedness and white forehead stripe did not report the well-known KIT gene, the significant genes such as MAPK10, EFNA5, PPP2R3C and PAK1 are synergistically involved in the synthesis of melanin. Another gene, TECRL, on BTA6 observed to be associated with white forehead stripe phenotype in this study has been reported to be associated with pigmentation patterns in Fleckvieh cattle. Because of the complexity of coat colour patterns in the Nguni cattle, this study recommends comprehensive sampling and phenotyping particularly of leucism traits for future GWAS studies.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets generated and/or analyzed during the current study are available from the OSF repository using the link: https://osf.io/t36bw/.

ETHICS STATEMENT

The animal study was reviewed and approved by University of KwaZulu-Natal Animal Ethics Committee.

AUTHOR CONTRIBUTIONS

LK was responsible for the DNA extraction, genotyping of samples, data analysis and wrote the draft manuscript. FM, ED, GM and JS conceptualised the study. KH assisted with the genomic analysis. FM provided genotyping equipment on the Illumina BovineHD 770K chip at the Agricultural Research Council-Biotechnology Platform. FM, ED and KH assisted with the writing of the manuscript and with revisions. GM and JS assisted with the genome-wide association studies (GWAS) methods and critically reviewed the manuscript.

FUNDING

Funding for the genomic research and data analysis was supported by the National Research Foundation through the South African/Austria Joint Scientific and Technological Cooperation Programme (STGR160512164890).

ACKNOWLEDGMENTS

We would like to thank Sphamandla Ngcamu (Bartlow Research Station) and Mpumelelo Magawana (Kokstadt Research Station) and their teams for assisting with biological sample collection. We also express gratitude to the Biotechnology Platform team at the Agricultural Research Council for assistance with genotyping and data analysis. The National Research Foundation and the University of KwaZulu-Natal funded this study for which we are grateful.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.832702/full#supplementary-material

REFERENCES

Almathen, f., elbir, h., bahbahani, h., mwacharo, j., and hanotte, o. (2018). Polymorphisms in Mc1r and Asip Genes Are Associated with Coat Color Variation in the Arabian Camel. *J. Hered.* 109, 700–706. doi:10.1093/jhered/esy024

Bang, s., won, k. H., moon, H.-R., yoo, h., hong, a., song, y., et al. (2017). Novel Regulation of Melanogenesis by Adiponectin via the Ampk/crtc Pathway. *Pigment Cel Melanoma Res.* 30, 553–557. doi:10.1111/pcmr.12596

Bennett, d. C., and lamoreux, m. L. (2003). The Color Loci of Mice - A Genetic Century. *Pigment Cel. Res.* 16, 333–344. doi:10.1034/j.1600-0749.2003.00067.x

Bester, j., matjuda, l., rust, j., and fourie, h. (2003). The Nguni: a Case Study. *Communitybased Manag. Anim. Genet. Resour.*, 45–69.

Brodie, a., azaria, J. R., and ofran, y. (2016). How Far from the Snp May the Causative Genes Be? *Nucleic Acids Res.* 44, 6046–6054. doi:10.1093/nar/gkw500

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. *GigaSci* 4. doi:10.1186/s13742-015-0047-8

Cieslak, m., reissmann, m., hofreiter, m., and ludwig, a. (2011). Colours of Domestication. *Biol. Rev.* 86, 885–899. doi:10.1111/j.1469-185x.2011.00177.x

De waal, h. (2014). *Overview of the Northern Cape Idc Nguni Cattle Development Project and the Free State Idc Nguni Cattle Development project*. Bloemfontein, South Africa: University of free state press.

D'mello, s., finlay, g., baguley, b., and askarian-amiri, m. (2016). Signaling Pathways in Melanogenesis. *Int. J. Mol. Sci.* 17, 1144.

Dunn, k. J., brady, m., Ochsenbauer-Jambor, c., snyder, s., incao, a., and pavan, w. J. (2005). Wnt1 and Wnt3a Promote Expansion of Melanocytes through Distinct Modes of Action. *Pigment Cel. Res.* 18, 167–180. doi:10.1111/j.1600-0749.2005.00226.x

Durkin, k., coppieters, w., Drögemüller, c., ahariz, n., cambisano, n., druet, t., et al. (2012). Serial Translocation by Means of Circular Intermediates Underlies Colour Sidedness in Cattle. *Nature,* 482, 81.doi:10.1038/nature10757

Feeley, n., bottomley, s., and munyard, k. (2011). Three Novel Mutations in Asip Associated with Black Fibre in Alpacas (vicugna Pacos). *J. Agric. Sci.* 4, 529–538. doi:10.1017/s0021859610001231

Fontanesi, l., scotti, e., and russo, v. (2010). Analysis of Snps in the Kit Gene of Cattle with Different Coat Colour Patterns and Perspectives to Use These Markers for Breed Traceability and Authentication of Beef and Dairy Products. *Ital. J. Anim. Sci.* 9, e42. doi:10.4081/ijas.2010.e42

Hanna, l. L. H., sanders, j. O., riley, d. G., abbey, c. A., and gill, c. A. (2014). Identification of a Major Locus Interacting with Mc1r and Modifying Black Coat Color in an F 2 Nellore-Angus Population. *Genet. selection Evol.* 46, 4. doi:10.1186/1297-9686-46-4

Harris, b., creagh, f., winkelman, a., and johnson, d. (2011). Experiences with the Illumina High Density Bovine Beadchip. *Interbull Bull.*

Horsburgh, k. A., prost, s., gosling, a., stanton, j.-a., rand, c., and matisoo-smith, e. A. (2013). The Genetic Diversity of the Nguni Breed of African Cattle (bos spp.): Complete Mitochondrial Genomes of Haplogroup T1. *Plos one* 8, e71956. doi:10.1371/journal.pone.0071956

Huson, h. J., kim, e.-s., godfrey, r. W., olson, t. A., mcclure, m. C., chase, c. C., et al. (2014). Genome-wide Association Study and Ancestral Origins of the Slick-Hair Coat in Tropically Adapted Cattle. *Front. Genet.* 5, 101. doi:10.3389/fgene.2014.00101

Jiang, l., jiang, j., yang, j., liu, x., wang, j., wang, h., et al. (2013). Genome-wide Detection of Copy Number Variations Using High-Density Snp Genotyping Platforms in Holsteins. *Bmc genomics* 14, 1–10. doi:10.1186/1471-2164-14-131

Kang, h. M., sul, j. H., service, s. K., zaitlen, n. A., kong, s.-y., freimer, n. B., et al. (2010). Variance Component Model to Account for Sample Structure in Genome-wide Association Studies. *Nat. Genet.* 42, 348. doi:10.1038/ng.548

Kars, a. D. (1993). *A Genetic Analysis of the Bartlow Combine Nguni Stud Philosophiae Doctor university of orange Free State*.

Klungland, h., vage, d., gomez-raya, l., adalsteinsson, s., and lien, s. (1995). The Role of Melanocyte-Stimulating Hormone (Msh) Receptor in Bovine Coat Color Determination. *Mamm. Genome* 6, 636–639. doi:10.1007/bf00352371

Koseniuk, a., ropka-molik, k., rubiś, d., and smołucha, g. (2018). Genetic Background of Coat Colour in Sheep. *Archiv fuer tierzucht* 61, 173. doi:10.5194/aab-61-173-2018

Makina, s. O., muchadeyi, f. C., van marle-köster, e., macneil, m. D., and maiwashe, a. (2014). Genetic Diversity and Population Structure Among Six Cattle Breeds in south africa Using a Whole Genome Snp Panel. *Front. Genet.* 5, 333. doi:10.3389/fgene.2014.00333

Makina, s. o., taylor, j. f., van marle-köster, e., muchadeyi, f. c., makgahlela, m. l., macneil, m. d., et al. (2015). Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. *Front. Genet.* 6, 337. doi:10.3389/fgene.2015.00337

Mapholi, n. O., marufu, m. C., maiwashe, a., banga, c. B., muchenje, v., macneil, m. D., et al. (2014). Towards a Genomics Approach to Tick (Acari: Ixodidae) Control in Cattle: a Review. *Ticks tick-borne Dis.* 5, 475–483. doi:10.1016/j.ttbdis.2014.04.006

Mapiye, c., chimonyo, m., muchenje, v., dzama, k., marufu, m. C., and raats, j. G. (2007). Potential for Value-Addition of Nguni Cattle Products in the Communal Areas of south africa: a Review. *Afr. J. Agric. Res.* 2, 488–495.

Marks, m. S., and seabra, m. C. (2001). The Melanosome: Membrane Dynamics in Black and white. *Nat. Rev. Mol. Cel. Biol.* 2, 738–748. doi:10.1038/35096009

Mastrangelo, s., sottile, g., sardina, m. T., sutera, a. M., tolone, m., di gerlando, r., et al. (2019). A Combined Genome-wide Approach Identifies a New Potential Candidate Marker Associated with the Coat Color Sidedness in Cattle. *Livestock Sci.* 225, 91–95. doi:10.1016/j.livsci.2019.05.009

Meseret, s., mekonnen, y. A., brenig, b., schütz, e., hanotte, o., gültas, m., et al. (2020). *Genetic Diversity and Population Structure of Six Ethiopian Cattle Breeds from Different Geographical Regions Using High Density Single Nucleotide Polymorphisms*. Livestock Science 234, 103979. doi:10.1016/j.livsci.2020.103979

Mészáros, g., petautschnig, e., schwarzenbacher, h., and sölkner, j. (2015). Genomic Regions Influencing Coat Color Saturation and Facial Markings in Fleckvieh Cattle. *Anim. Genet.* 46, 65–68.

Millington, g. (2006). Proopiomelanocortin (Pomc): the Cutaneous Roles of its Melanocortin Products and Receptors. *Clin. Exp. Dermatol. Clin. Dermatol.* 31, 407–412. doi:10.1111/j.1365-2230.2006.02128.x

Musemwa, l., mushunje, a., chimonyo, m., and mapiye, c. (2010). Low Cattle Market Off-Take Rates in Communal Production Systems of south africa: Causes and Mitigation Strategies. *J. Sustain. Dev. africa* 12, 209–226.

Natale, c. A., duperret, e. K., zhang, j., sadeghi, r., dahal, a., o'brien, k. T., et al. (2016). Sex Steroids Regulate Skin Pigmentation through Nonclassical Membrane-Bound Receptors. *Elife* 5, e15104. doi:10.7554/eLife.15104

Nazari-ghadikolaei, a., mehrabani-yeganeh, h., miarei-aashtiani, s. R., staiger, e. A., rashidi, a., and huson, h. J. (2018). Genome-wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. *Front. Genet.* 9, 105. doi:10.3389/fgene.2018.00105

O'brien, a. M. P., utsunomiya, y. T., mészáros, g., bickhart, d. M., liu, g. E., van tassell, c. P., et al. (2014). Assessing Signatures of Selection through Variation in Linkage Disequilibrium between Taurine and Indicine Cattle. *Genet. selection Evol.* 46, 19.

Ojango, j., mpofu, n., marshall, k., and andersson-eklund, l. (2011). Quantitative Methods to Improve the Understanding and Utilization of Animal Genetic Resources. *Anim. Genet. Train. resource* 3. version.

Olson, t. (1999). Genetics of Colour Variation. *The Genet. cattle*, 33–53.

Oosthuizen, m. P. (1996). *Uchibidolo: The Abundant Herds: a Descriptive Study of the Sanga-Nguni Cattle of the Zulu People, with Special Reference to Colour-Pattern Terminology and Naming-Practice*.

Park, h., kosmadaki, m., yaar, m., and gilchrest, b. (2009). Cellular Mechanisms Regulating Human Melanogenesis. *Cell Mol. Life Sci.* 66, 1493–1506. doi:10.1007/s00018-009-8703-8

Ran, j.-s., you, x.-y., jin, j., zhou, y.-g., wang, y., lan, d., et al. (2016). The Relationship between Mc1r Mutation and Plumage Color Variation in Pigeons. *Biomed. Research International*. doi:10.1155/2016/3059756

Riffert, v. E. (2015). *Mapping Genes Responsible for Coat Colour Patterns of the Gir Cattle Breed in brazil*. vienna: Master thesis, university of natural resources and life sciences.

Royo, l. J., alvarez, i., arranz, j., fernández, i., rodríguez, a., pérez-pardal, l., et al. (2008). Differences in the Expression of the Asip Gene Are Involved in the Recessive Black Coat Colour Pattern in Sheep: Evidence from the Rare Xalda Sheep Breed. *Anim. Genet.* 39, 290–293. doi:10.1111/j.1365-2052.2008.01712.x

Sanarana, y., visser, c., bosman, l., nephawe, k., maiwashe, a., and van marle-köster, e. (2016). Genetic Diversity in South African Nguni Cattle Ecotypes Based on Microsatellite Markers. *Trop. Anim. Health Prod.* 48, 379–385. doi:10.1007/s11250-015-0962-9

Seitz, j. J., schmutz, s. M., thue, t. D., and buchanan, f. C. (1999). A Missense Mutation in the Bovine Mgf Gene Is Associated with the Roan Phenotype in Belgian Blue and Shorthorn Cattle. *Mamm. Genome* 10, 710–712. doi:10.1007/s003359901076

Steingrímsson, e., copeland, n. G., and jenkins, n. A. (2004). Melanocytes and the Microphthalmia Transcription Factor Network. *Annu. Rev. Genet.* 38, 365–411.

Szczerbal, i., robinson, t. J., duran, a., davey, s. C., andersson, m., and switonski, m. (2017). Ectopic Position of Duplicated Kit Gene in African Nguni Cattle, Associated with Color Sidedness, Confirms its Shared Ancestry with the bos Taurus Lineage. *Anim. Genet.* 48, 122–123. doi:10.1111/age.12495

Tada, o., muchenje, v., and dzama, k. (2013). Preferential Traits for Breeding Nguni Cattle in Low-Input *In-Situ* Conservation Production Systems. *Springerplus* 2, 195. doi:10.1186/2193-1801-2-195

Uemoto, y., sasaki, s., kojima, t., sugimoto, y., and watanabe, t. (2015). Impact of Qtl Minor Allele Frequency on Genomic Evaluation Using Real Genotype Data and Simulated Phenotypes in Japanese Black Cattle. *Bmc Genet.* 16, 134. doi:10.1186/s12863-015-0287-8

Våge, d. I., lu, d., klungland, h., lien, s., adalsteinsson, s., and cone, r. D. (1997). A Non-epistatic Interaction of agouti and Extension in the Fox, vulpes vulpes. *Nat. Genet.* 15, 311–315. doi:10.1038/ng0397-311

Venter, y., and theron, h. (2016). *Nguni Cattle Breeder's Society. South African Stud Book and Animal Improvement Association*.

Videira, i. F. D. S., moura, d. F. L., and magina, s. (2013). Mechanisms Regulating Melanogenesis. *Anais brasileiros de dermatologia* 88, 76–83. doi:10.1590/s0365-05962013000100009

Wellbrock, c., weisser, c., geissinger, e., troppmair, j., and schartl, m. (2002). Activation of P59fyn Leads to Melanocyte Dedifferentiation by Influencing Mkp-1-Regulated Mitogen-Activated Protein Kinase Signaling. *J. Biol. Chem.* 277, 6443–6454. doi:10.1074/jbc.m110684200

Yang, g.-l., fu, d.-l., lang, x., wang, y.-t., cheng, s.-r., fang, s.-l., et al. (2013). Mutations in Mc1r Gene Determine Black Coat Color Phenotype in Chinese Sheep. *scientific World J.*

Zhang, y., and pan, w. (2015). Principal Component Regression and Linear Mixed Model in Association Analysis of Structured Samples: Competitors or Complements? *Genet. Epidemiol.* 39, 149–155. doi:10.1002/gepi.21879

Zhou, j., ren, t., li, y., cheng, a., xie, w., xu, l., et al. (2017). Oleoylethanolamide Inhibits α-melanocyte Stimulating Hormone-Stimulated Melanogenesis via Erk, Akt and Creb Signaling Pathways in B16 Melanoma Cells. *Oncotarget* 8, 56868. doi:10.18632/oncotarget.18097

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 | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 08 July 2022
doi: 10.3389/fvets.2022.938311

Genome-Wide Analysis of microRNAs Identifies the Lipid Metabolism Pathway to Be a Defining Factor in Adipose Tissue From Different Sheep

*Tian-Yi Liu, Hui Feng, Salsabeel Yousuf, Ling-Li Xie and Xiang-Yang Miao**

State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

microRNAs are a class of important non-coding RNAs, which can participate in the regulation of biological processes. In recent years, miRNA has been widely studied not only in humans and mice, but also in animal husbandry. However, compared with other livestock and poultry breeds, the study of miRNA in subcutaneous adipose tissue of sheep is not comprehensive. Transcriptome analysis of miRNAs in subcutaneous adipose tissue of Duolang sheep, and Small Tail Han sheep was performed using RNA-Seq technology. Differentially expressed miRNAs were screened between different breeds. Target genes were predicted, and then the joint analysis of candidate genes were conducted based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment. Finally, the RNA-Seq data were verified by real-time quantitative polymerase chain reaction (qRT-PCR). Herein, we identified 38 differentially expressed miRNAs (9 novel miRNAs and 29 known miRNAs). In addition, a total of 854 target genes were predicted by miRanda software. GO and KEGG pathway analysis demonstrated that regulation of lipolysis in adipocytes plays a key role in the deposition of subcutaneous adipose tissue in Duolang sheep and Small Tail Han sheep. The miRNAs might regulate fat deposits by regulating genes involved in regulation of lipolysis in adipocytes. Specifically, NC_ 040278.1_ 37602, oar-mir-493-3p, NC_ 040278.1_ 37521 and NC_ 040255.1_ 11627 might target PTGS2, AKT2, AKT3, and PIK3CA, respectively, and then play critical regulatory role. In conclusion, all the results provide a good idea for further revealing the mechanism of subcutaneous adipose tissue deposition and improving the meat production performance of sheep, and lay a foundation for promoting the development of animal husbandry.

Keywords: sheep, subcutaneous fat, microRNAs, high-throughput sequencing, fat deposition

OPEN ACCESS

Edited by:
Natalia A. Zinovieva, L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

Reviewed by:
Hui Li, Guangxi University, China
Xiaofei Guo, Tianjin Academy of Agricultural Sciences, China

****Correspondence:***
Xiang-Yang Miao
miaoxy32@163.com

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

Received: *07 May 2022*
Accepted: *06 June 2022*
Published: *08 July 2022*

Citation:
Liu T-Y, Feng H, Yousuf S, Xie L-L and Miao X-Y (2022) Genome-Wide Analysis of microRNAs Identifies the Lipid Metabolism Pathway to Be a Defining Factor in Adipose Tissue From Different Sheep. Front. Vet. Sci. 9:938311. doi: 10.3389/fvets.2022.938311

INTRODUCTION

miRNA is a kind of short-chain endogenous non-coding RNA with a length of 18–25 nt, which is highly conservative and widely distributed in animals, plants and viruses (1, 2). With the discovery of lin-4 and let-7 in Caenorhabditis elegans (3, 4), the research of miRNA has gradually entered people's vision. Although miRNAs cannot encode proteins, they still control various biological

processes and are key regulators of development and cellular homeostasis (5). In many biological processes, miRNAs regulate gene expression by recognizing homologous sequences and interfering with transcriptional, translational or epigenetic processes (6). In the past few years, the research of miRNA in cancer has exploded (7), including diseases such as hepatocellular carcinoma (8), breast cancer (9), and gastric cancer (10). Not only that, miRNA in type 2 diabetes (11), hypertension (12) and atherosclerosis (13) is also very common. Obesity has become a prevalent health problem as it is closely associated with many diseases.

Obesity is due to the proliferation, differentiation or enlargement of adipocytes, and miRNAs have both promoting and inhibiting effects on adipocyte differentiation (14). LPL is a direct target gene of miR-152 during preadipocyte differentiation. MiR-152 can inhibit the proliferation of 3T3-L1 preadipocytes and promote the differentiation of 3T3-L1 preadipocytes by negatively regulating LPL (15). The study found that miR-244 regulates the adipogenic differentiation of bovine preadipocytes by targeting LPL. When miR-224 was overexpressed, the mRNAs expression levels of the adipogenesis-related markers PPARγ, FASN, C/EBPα, C/EBPβ and PLIN1 decreased, whereas the opposite effect was produced when miR-224 was inhibited (16). The regulation of lipid metabolism by miRNA is a complex network regulation system, which mainly affects lipid metabolism by regulating genes related to lipid synthesis, oxidation, and transport (17, 18). For example, miR-33, as one of the widely studied miRNAs, is involved in the regulation of multiple lipid metabolism links. The expression of genes related to fatty acid β-oxidation can be inhibited, and the biosynthesis of high-density lipoprotein and the outflow of cholesterol can also be inhibited (19, 20). As the first miRNA found to be involved in the regulation of lipid metabolism, miR-122 has been deeply studied. MiR-122 is most abundant in the liver and plays an important role in mainly regulating lipid synthesis and oxidation processes, especially the accumulation of triglycerides. MiR-122 plays an important regulatory role in the translation of YY1 and FXR. Mir-122 upregulates FXR expression by targeting the 3'UTR of YY1 mRNA upstream, suppressing triglyceride levels in hepatocytes, and FXR is also involved in the regulation of lipid metabolism disorders and insulin resistance (21). Another study showed that the expression of miR-122 was inhibited by binding to the 3'UTR of Sirt1 in non-alcoholic fatty liver disease, and down-regulation of miR-122 inhibited adipogenesis genes, but activated the AMPK signaling pathway, further inhibiting hepatic lipogenesis and triglyceride secretion (22). Studies have found that miR-122 can also promote cholesterol synthesis (23, 24). The above studies show that miRNA plays a key regulatory role in the process of fat deposition.

As one of the main meat livestock and poultry resources in the world, sheep has warm meat, high protein content and low fat and cholesterol levels. Compared with pork, the meat quality is more delicate, and the content of essential amino acids is also higher than that of pigs, chickens and cattle (25, 26). Adipose tissue is an important factor affecting meat quality, mainly including the effects on sensory, flavor and tenderness (27, 28). Moreover, the back-fat of sheep is an important indicators of meat yield and hot carcass composition (29). And subcutaneous adipose tissue accounts for the highest proportion of total fat content in animals (30), which has the function of protecting animals and storing energy. Duolang sheep and Small Tail Han sheep belong to high-quality local breeds in China, and there are differences in fat deposition between the two breeds. Duolang sheep have large fat buttocks, strong meat production capacity, fresh and juicy meat, and belongs to both meat and fat sheep, and Duolang sheep also have the characteristics of rapid growth and development. Small Tail Han sheep are short and thin-tailed sheep, with strong environmental adaptability and reproductive ability, fast growth rate and stable genetic performance, but the meat body size is not obvious and the carcass meat production rate is low. At present, living standards have been greatly improved, and people's requirements for meat quality and quantity have also increased. Therefore, studying the molecular mechanism of sheep fat deposition can improve the quality of meat products to meet consumer demand. At the same time, adipose tissue is an important organ of energy metabolism, and excessive fat deposition can lead to obesity and the occurrence of a series of metabolic syndromes (31). Therefore, studying the mechanism of fat deposition can not only improve meat quality and promote the development of animal husbandry, but also prevent or treat a series of diseases affecting human health caused by excessive fat deposition. We selected the subcutaneous adipose tissue of Duolang sheep and Small Tail Han sheep as experimental materials, analyzed the gene expression profile of subcutaneous adipose tissue by transcriptome sequencing and bioinformatics methods, screened and identified the key candidate genes related to lipid metabolism and adipogenic differentiation and explored the molecular mechanism related to fat deposition.

MATERIALS AND METHODS

Sample Collection and RNA Isolation

All work in this study was approved by the Animal Welfare and Ethics Committee of Beijing Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (No. IAS2019-82). In order to detect the expression profile of miRNAs in sheep subcutaneous fat, we collected subcutaneous adipose tissue samples from 3 adult female Duolang sheep (D-PF-1, 2, 3) and 3 adult female Small Tail Han sheep (X-PF-1, 2, 3). All sheep were raised in the same conditions with free to drink and eat under natural light, and their diet meets the current nutritional needs. They were in good physical condition, aged 2 years old and the mean bodyweight of sheep were 50 ± 3 kg. We collected subcutaneous adipose tissue samples located at the back fat according to the agricultural industry standard of the people's Republic of China (NY/T 3469-2019). In order to reduce the pain of animals, we first stunned them with electricity and then slaughtered them. The whole sampling process was controlled within 30 min. The collected samples were immediately frozen in liquid nitrogen and stored in an environment of −80°C before the experiment.

Total RNA was isolated from each adipose tissue using TRIzol reagent (Invitrogen, Invitrogen Life Technologies, Carlsbad,

TABLE 1 | Quality control statistics of sequencing data for each sample.

Sample	Raw reads	Clean reads	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)	Useful reads (18–32 nt)
D_PF_1	19,305,542	19,237,461	0.0239	98.5	95.16	45.08	18,716,909
D_PF_2	19,293,493	19,262,447	0.0238	98.57	95.3	43.86	19,040,661
D_PF_3	18,321,730	18,244,173	0.0239	98.54	95.15	44.95	17,764,605
X_PF_1	19,036,709	18,972,125	0.0253	97.76	94.08	44.79	18,542,581
X_PF_2	23,072,665	22,989,021	0.0244	98.22	94.73	44.2	22,450,480
X_PF_3	19,383,731	19,336,376	0.0257	97.53	93.73	45.16	18,945,034

TABLE 2 | Reference genome comparison results for each sample.

Sample	Total reads	Total mapped	Mapped reads (+)	Mapped reads (–)
D_PF_1	18,716,909	17,572,239 (93.88%)	13,567,787	7,418,996
D_PF_2	19,040,661	18,495,698 (97.14%)	14,384,579	8,158,010
D_PF_3	17,764,605	16,903,298 (95.15%)	13,052,330	7,106,665
X_PF_1	18,542,581	17,787,015 (95.93%)	12,868,296	8,980,321
X_PF_2	22,450,480	21,147,342 (94.20%)	14,790,236	11,768,078
X_PF_3	18,945,034	17,696,401 (93.41%)	13,532,382	7,894,160

USA), and genomic DNA was removed using rDNase I RNase-free (TaKara). The RNA concentration and quality were then measured with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and the ND-2000 (NanoDrop Technologies), RIN > 8 was good quality.

Library Preparation and Sequencing

The RNA library was constructed using the TruSeqTM Small RNA sample prep Kit (Invitrogen) according to the instructions. The rRNA in the total RNA was first removed, the 3′ end adapter and the 5′ end adapter were, respectively, connected with the kit, and then the random primers were reversed to 1st cDNA. To enrich the library we performed 11-12 PCR cycles. The library was enriched and then purified (6% Novex TBE PAGE gel, 1.0 mm, 10 well) and quantified by TBS380 (Picogreen). Bridge PCR was performed on cBot to generate clusters. The library was sequenced with the Illumina NovaSeq 6000.

Quality Control and Sequence Alignment

In order to improve the sequencing quality and get clean reads, introduced adapter sequences, low-quality reads, sequences with high N rate (N stands for indeterminate bases), and sequences that are too short will be removed. The clean reads were aligned with the reference genome using Bowtie (32) software. The reference genome GCF_016772045.1 was obtained in the NCBI database. StringTie (32) software was used to splicing the mapped reads.

miRNA Identification and Structural Analysis

The reads aligned to the reference genome were aligned with the miRNA precursor and mature sequences in the miRBase V.22.1 (33) to obtain known miRNAs. The Rfam (34) was used to filter ncRNAs and repeats sequences such as ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA), and at that time, the types and numbers of these sequences were counted. The sRNAs that cannot be aligned with Rfam and miRBase were aligned to the reference genome, and the surrounding sequences were intercepted using miRDeep2 (35) software for secondary structure prediction to identify new miRNAs. In the process of developing from precursor to mature miRNA, the Dicer restriction site was specific, which makes the first base of miRNA mature sequence strongly biased to U. In addition, some non-canonical editing sites exist in miRNAs, thereby altering target genes. And the MiRME (36) method was used to detect various mutation and editing sites of miRNA. Finally, based on the seed

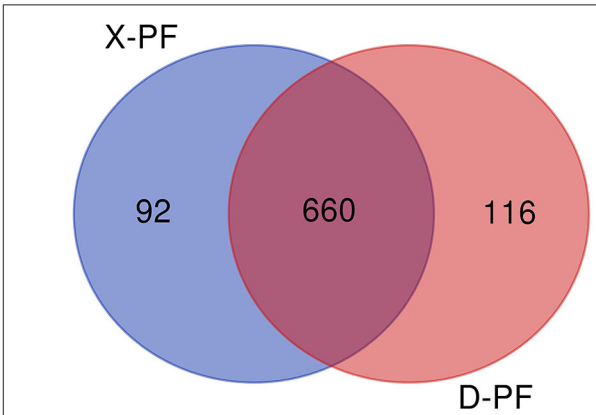


FIGURE 1 | miRNAs Venn diagram between two groups (D-PF vs. X-PF). D-PF represented the subcutaneous adipose tissue of Duolang sheep, X-PF represented the subcutaneous adipose tissue of Small Tail Han sheep.

TABLE 3 | Small RNA classification statistics.

Type	D_PF_1	D_PF_2	D_PF_3	X_PF_1	X_PF_2	X_PF_3
Known miRNA	13,353,021 (71.34%)	14,835,693 (77.92%)	13,060,553 (73.52%)	13,224,849 (71.32%)	16,825,434 (74.94%)	12,843,282 (67.79%)
Novel miRNA	798,022 (4.26%)	892,391 (4.69%)	735,426 (4.14%)	1,027,702 (5.54%)	1,330,213 (5.93%)	881,219 (4.65%)
rRNA	305,343 (1.63%)	91,711 (0.48%)	179,636 (1.01%)	149,916 (0.81%)	119,537 (0.53%)	184,698 (0.97%)
tRNA	532,977 (2.85%)	189,584 (1.0%)	405,022 (2.28%)	255,955 (1.38%)	951,902 (4.24%)	645,567 (3.41%)
snoRNA	3,185 (0.02%)	2,976 (0.02%)	4,895 (0.03%)	3,763 (0.02%)	3,455 (0.02%)	2,887 (0.02%)
snRNA	4,902 (0.03%)	4,423 (0.02%)	3,732 (0.02%)	2,679 (0.01%)	2,968 (0.01%)	4,771 (0.03%)
Unknown	863,087 (4.61%)	508,884 (2.67%)	697,672 (3.93%)	583,771 (3.15%)	504,830 (2.25%)	764,738 (4.04%)
Other	2,856,372 (15.27%)	2,514,999 (13.21%)	2,677,669 (15.07%)	3,293,946 (17.77%)	2,712,141 (12.08%)	3,617,872 (19.1%)
Sum	18,716,909	19,040,661	17,764,605	18,542,581	22,450,480	18,945,034

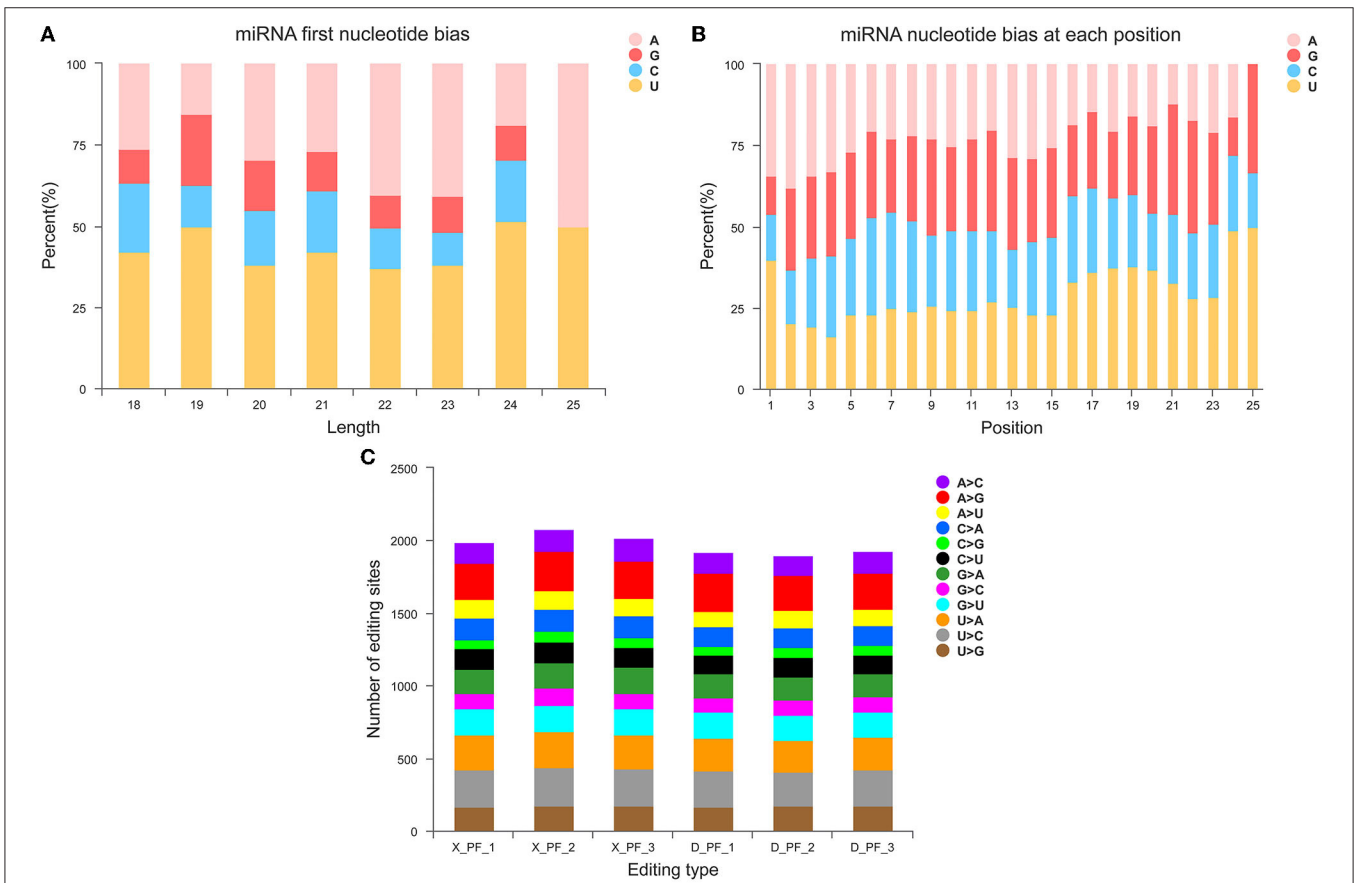


FIGURE 2 | Structural characteristics of miRNA. **(A)** First base preference distribution map of miRNAs with different lengths. **(B)** Distribution map of base preference at different miRNA sites. **(C)** miRNA base editing type distribution map.

sequence, the identified known miRNAs and new miRNAs were subjected to miRNA family analysis.

Differentially Expressed miRNAs

In order to facilitate the subsequent analysis of the differential expression among the samples, quantitative analysis was performed on the expression levels of the samples, respectively. RSEM (37) software was used for quantification, and the expression level of each miRNA was calculated according to the transcripts per million readsx (TPM) method (38). DESeq2 (39) is a statistical analysis based on negative binomial distribution, which can be used in experiments with biological replicates. Significant differently expressed miRNAs were extracted with p-value < 0.05 and $|\log_2 FC| \geq 1$.

miRNA Target Gene Prediction

Because miRNAs cannot encode proteins, they function through post-transcriptional regulation of their target genes. In animals, miRNA relies on the seed sequence (2–8 nt at the 5′ end) to closely bind to the 3′ non-coding region of the target gene,

thereby inhibiting the translation of the target mRNA. In this study, the miRanda (40) software was used to predict its target genes, and the parameters were set as: Score $\geq$ 160 and Energy ≤ -20. The predicted target genes were visualized by the application software Cytoscape (41). All of the mRNA data was obtained from fat collected from the same animals that were used for the miRNA analysis at the same relative time (age) and using a similar procedure.

GO and KEGG Enrichment Analysis of Differentially Expressed Genes

GO (42) can be used for functional enrichment analysis on the differentially expressed genes. The software Goatools was used to perform GO enrichment analysis on the differentially expressed genes, so as to obtain functional annotations of the genes, including three parts: biological process, molecular function and cellular components. The KEGG (43) pathway enrichment analysis was also performed on the genes, and the R software was used for the enrichment analysis of metabolic pathways and information processing pathways. Using Fisher's exact test and Benjamini and Hoceberg (BH) to correct *p*-value to obtain *p*-adjust, when *p*-adjust < 0.05, we considered significant enrichment.

Quantitative Real-Time Polymerase Chain Reaction

Real-time fluorescent quantitative PCR (qRT-PCR) was used to verify the reliability of the sequencing results. Five differentially expressed miRNAs and eight differentially expressed mRNAs were randomly selected for verification. In total, 0.5 μg RNA was taken to synthesize cDNA template through GeneAmp® PCR System 9700 (Applied Biosystems, USA). First, RNA, 4 × g DNA wiper Mix and nuclease free H_2O were reacted in GeneAmp® PCR System 9700 at 42°C for 2 min. Second, 5 × HiScript II Q RT SuperMix IIa were added and reacted at 50°C for 15 min, then for 5 s at 85°C. And then the reverse transcription reaction mix were dilute × 10 in nuclease free H_2O and maintained at −20°C. The qRT-PCR analysis was conducted using LightCycler® 480 II Real-time PCR Instrument (Roche, Swiss). U6 and ACTB were used as miRNA and mRNA internal references, respectively. Three biological replicates were employed for each gene. The relative expression levels of genes between samples were calculated using $2^{-\Delta\Delta Ct}$ method. Data obtained were analyzed using GraphPad Prism (V8.0.1). The student *t*-test ($p < 0.05$) was used for mean comparisons. All results were presented in bar charts with the means and their standard deviation (±SD). The sequences of the primers used are listed in **Supplementary Table S1**.

Statistical Analysis

All the data were presented as means ± SD. When comparisons were made, a student's *t*-test was performed and *p*-value < 0.05 was considered as statistically significant.

RESULTS

Evaluation of RNA-Sequencing Data

In order to study the mechanism of miRNAs in subcutaneous adipose tissue of Duolang sheep and Small Tail Han sheep, we constructed cDNA libraries of two breeds, and obtained the original data by high-throughput sequencing technology. We obtained a total of 118.41 M Raw Reads in 6 sheep subcutaneous adipose tissue samples, and the Raw Reads in each sample reached more than 18.32 M, the percentage of Q20 bases was more than 97%, and the percentage of Q30 bases was more than 93% (**Table 1**). After quality control, compared with the reference genome, the mapping rate of each sample ranges from 93.41 to 97.14%, indicating that the reference genome is fully annotated

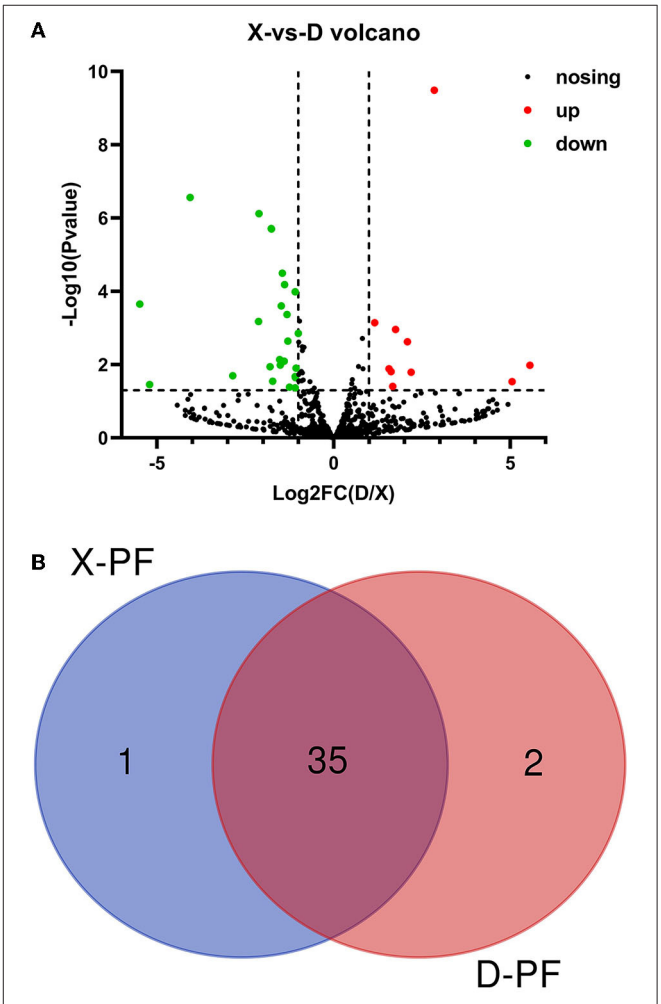


FIGURE 3 | Differential expression of miRNA. **(A)** Volcano map of differentially expressed miRNAs. The scatter in the figure represents the miRNAs, the black dots indicate the miRNAs with no significant differences, the red dots indicating a significant up-regulation of the miRNA, and the green dots represent a significant down-regulation of the miRNA. **(B)** The miRNA Venn diagram was differentially expressed in the two groups. D-PF represented the subcutaneous adipose tissue of Duolang sheep, X-PF represented the subcutaneous adipose tissue of Small Tail Han sheep.

and there is no contamination in the experiment, which is a good foundation for subsequent data analysis (**Table 2**).

miRNA Identification and Prediction

After comparing with miRBase database, we obtain the known miRNA information of each sample, and then compare and analyze the sequences without annotation information to obtain the newly predicted miRNA. A total of 868 miRNAs were obtained in this experiment, of which the number of known and new miRNAs were 149 and 719, respectively, 660 of the 868 miRNAs were coexpressed, and 116 and 92 miRNAs existed only in Duolang sheep and Small Tail Han sheep, respectively (**Figure 1**). The statistics of small RNA are shown in **Table 3**, including miRNA sequences and non-miRNA sequences. miRNA accounts for the largest proportion in each sample (**Supplementary Figure S1**).

miRNA Structural Analysis

Counting the known 149 miRNAs, we found them from 51 families and 271 species (**Supplementary Table S2**). Statistics on the first base of miRNAs of different lengths and base preferences at different sites (**Figure 2A**) and base preferences at different miRNA sites (**Figure 2B**), we found that four bases A, G, C, and U accounted for the proportion of first place varies, and the proportion of first place preference U is the largest (**Supplementary Table S3**), which is consistent with previous

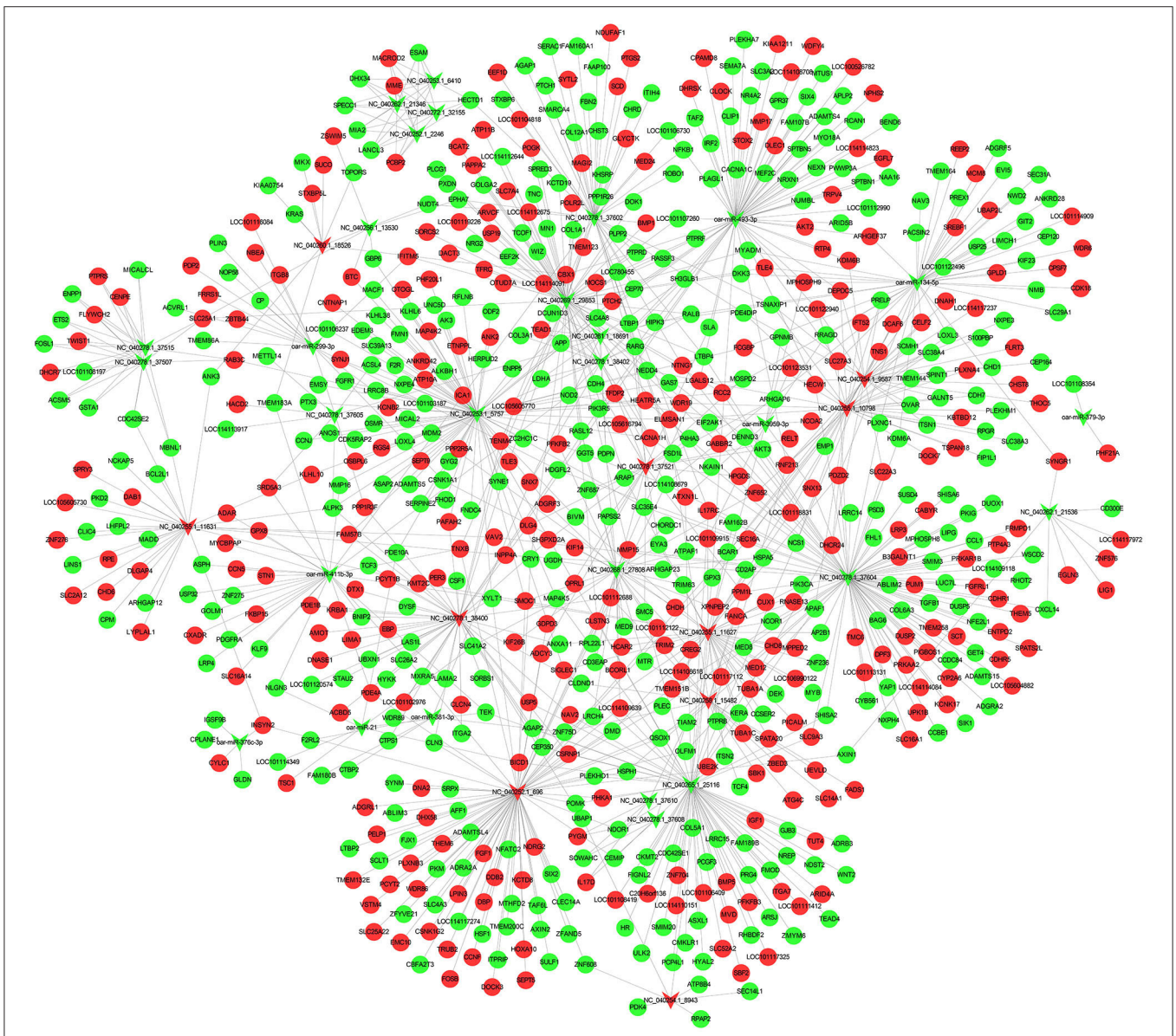


FIGURE 4 | Regulatory networks of genes. miRNA-mRNA regulatory network in sheep subcutaneous adipose tissue. The circle represents mRNA, and the arrow represents miRNA, the red represents up-regulated expression in Duolang sheep, and the green represents down-regulated expression.

studies in pigs. Analysis of miRNA base editing in each sample showed that the number of various base editing sites in the subcutaneous adipose tissue of the Small Tail Han sheep was slightly higher than that of the Duolang sheep (**Figure 2C**).

Differential Expression of miRNA

There are obvious differences in the deposition of adipose tissue between Duolang sheep and Small Tail Han sheep. By comparing the miRNAs expression of the two breeds of sheep, the differentially expressed miRNAs in the subcutaneous fat of the two breeds of sheep were screened. About 95% of the miRNAs were filtered out. There were 38 remaining differentially expressed miRNAs, of which 9 were known miRNAs and 29 were unknown miRNAs. Among the differentially expressed miRNAs, 10 were up-regulated and 28 were down-regulated in Duolang sheep (**Figure 3A**; **Supplementary Table S4**). To further analyze the interaction of the differentially expressed miRNAs in the two breeds, we constructed a Venn diagram (**Figure 3B**), of which one was only present in the Small Tail Han sheep and two were only in the Duolang sheep.

miRNA Target Gene Prediction and Functional Enrichment Analysis

miRNA functions by inhibiting the translation of target genes. In order to more directly observe the interaction between miRNAs and mRNAs, we useed Cytoscape to draw. When predicting the target genes of 38 differentially expressed miRNAs, we found that there are a large number of target genes, a total of 854 (**Figure 4**; **Supplementary Table S5**). The results show that one miRNA may regulate multiple target genes, and a target gene may also be regulated by multiple different kinds of miRNAs.

In order to explore the function of miRNAs, we followed the GO enrichment analysis of the target mRNAs of 38 different miRNAs. The results showed that in the biological process (**Figure 5A**), the target genes were significantly enriched in the regulation of protein phosphorylation, regulation of lipid catabolic process, regulation of MAPK cascade, cholesterol biosynthetic process, response to insulin. Differentially expressed target genes in molecular function (**Figure 5B**), mainly enriched in heparin binding, glycosaminoglycan binding, C-8 sterol isomerase activity, glycerate kinase activity and glycosylphosphatidylinositol phospholipase D activity. In terms of cellular components (**Figure 5C**), it was significantly enriched in cell surface, extracellular matrix, receptor complex and other related items. GO analysis showed that the differential miRNAs mainly regulated the lipid metabolism process and the regulation of enzyme activities in Duolang sheep and Small Tail Han sheep, resulting in differences in fat deposition between the two breeds.

KEGG enrichment was used to analyze the function of target genes (**Figure 6**). A total of 284 pathways were enriched, and the corrected p-adjust < 0.05 was set to screen the pathways. Among them, the significantly enriched pathways were PI3K Akt signaling pathway, AMPK, steroid biosynthesis, insulin signaling pathway, regulation of lipolysis in adipocytes, phospholipase D signaling pathway, MAPK and other pathways related to lipid metabolism. Therefore, it is speculated that the differential expression of miRNAs may regulate fat deposition by changing the signal transduction of two breeds of sheep.

The Relationship Between miRNA and Target Genes in Regulation of Lipolysis in Adipocytes

Regulation of lipolysis in adipocytes plays an important role in the process of fat deposition. We found that miRNA target genes are partially enriched in this signaling pathway, including PTGS2, ADRB3, LRCH4, PIK3CA, ADCY3, AKT3 and AKT2. The query results based on string database show that there is an interaction relationship between PTGS2, AKT2, AKT3 and PIK3CA (**Figure 7**). The miRNAs that target and regulate them are NC_040278.1_37602, oar-miR-493-3p, NC_040255.1_11627 and NC_040278.1_37521 (**Figure 8**). These differentially expressed miRNAs may cause the difference in fat deposition between Duolang sheep and Small Tail Han sheep.

QRT-PCR

To verify the accuracy of the RNA-Seq data, we detected the relative expression levels of 13 transcripts by qRT-PCR (**Figure 9**). The results showed that four miRNAs (NC_040278.1_37602, NC_040253.1_5757, NC_040262.1_21536 and NC_040278.1_37507) were down-regulated and NC_040255.1_11631 was up-regulated, which was consistent with the RNA-Seq results. In addition, we verified the expression of target genes (AKT3, PTGS2, COL1A1, MGST3, PCK1, PPP2R5A, FADS1, and LOC101113583) and proved that the sequencing results were reliable.

DISCUSSION

In recent years, related research on ncRNA in cancer has been widely reported. miRNA is a short-chain non-coding RNA that can specifically bind to the 3′ UTR of mRNA to inhibit mRNA translation or promote mRNA degradation. The deposition of adipose tissue not only plays a role in storing energy, but also can improve slaughter weight, slaughter rate, carcass quality and economic benefits for livestock and poultry in the fattening period. Previous studies on miRNA have mostly appeared in model animals such as mice, or in pigs that are genetically similar to humans, and there is relatively little research on miRNA in sheep. In this study, Duolang sheep and Small Tail Han sheep with different fat deposition ability were used to explore the regulation mechanism of miRNAs on fat deposition. We sequenced the miRNAs in sheep subcutaneous adipose tissue and analyzed the obtained data. The results showed that the structural characteristics of miRNAs were consistent with previous studies in other animals. And a total of 38 differentially expressed miRNAs were obtained, including 29 new miRNAs. MiRanda software was used to predict the target genes of miRNAs. The target genes were analyzed by GO and KEGG function enrichment, so as to speculate the function of miRNA. The target genes were mainly involved in regulation of lipolysis in adipocytes. The above results showed that miRNAs played a regulatory role in the deposition of sheep adipose tissue. Finally,

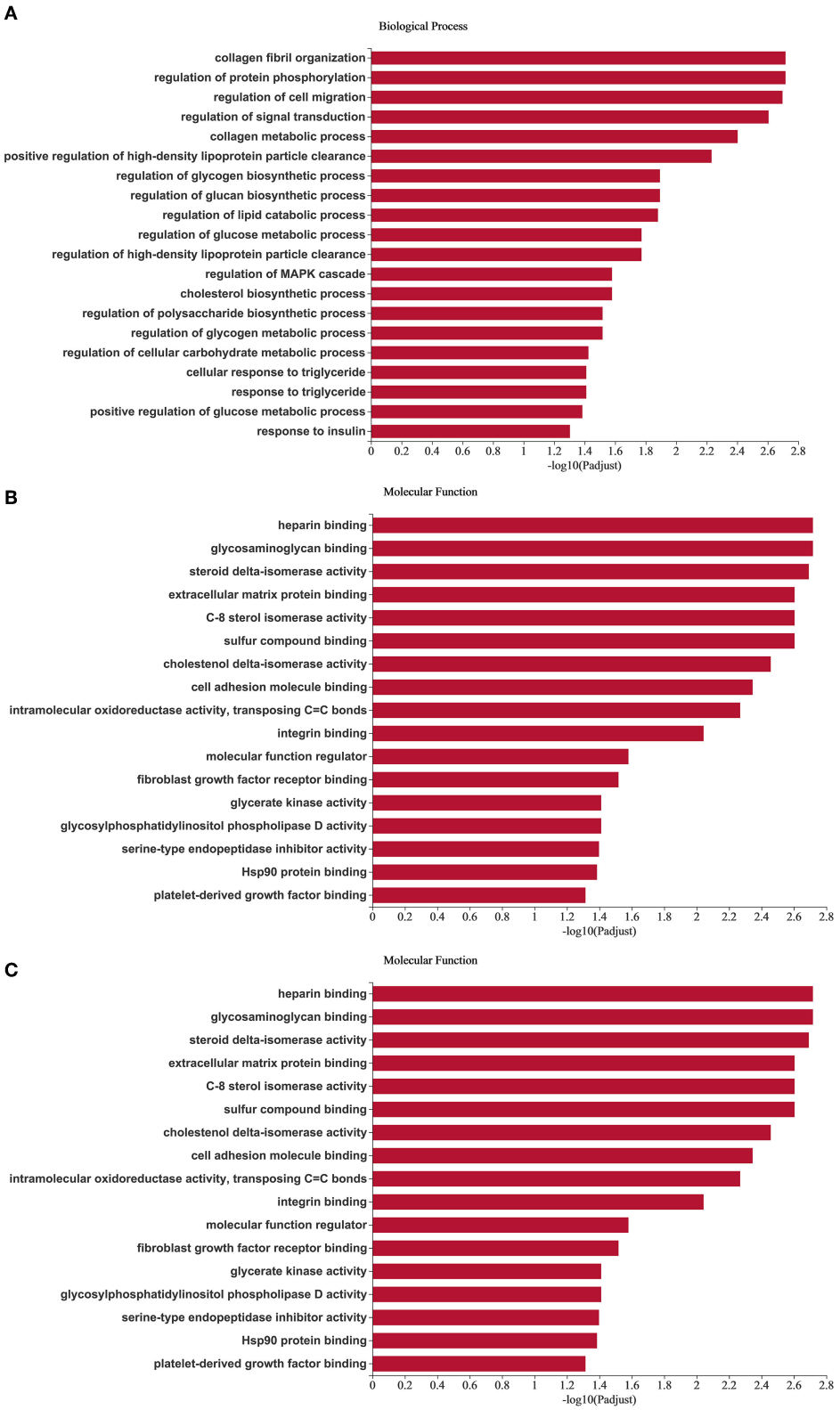


FIGURE 5 | GO enrichment analysis of target genes. **(A–C)** Are biological processes, molecular function, and cellular component, respectively. The X-axis is the enrichment significance, and the Y-axis is GO term.

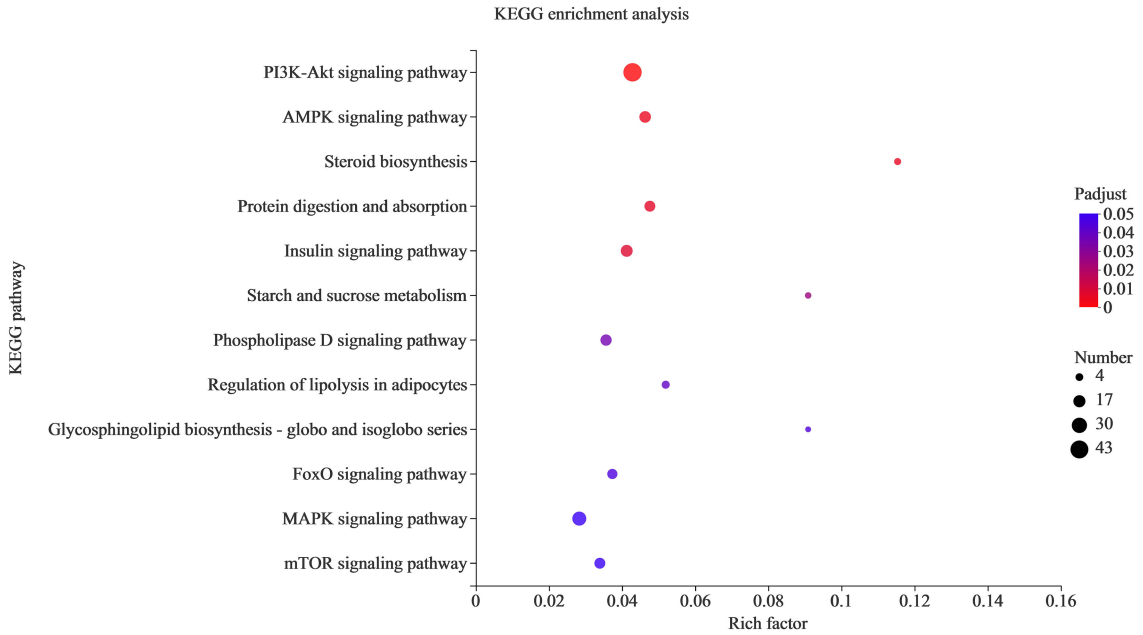


FIGURE 6 | KEGG pathway analysis of the target genes. The X-axis is the rich factor, and the Y-axis is KEGG pathway.

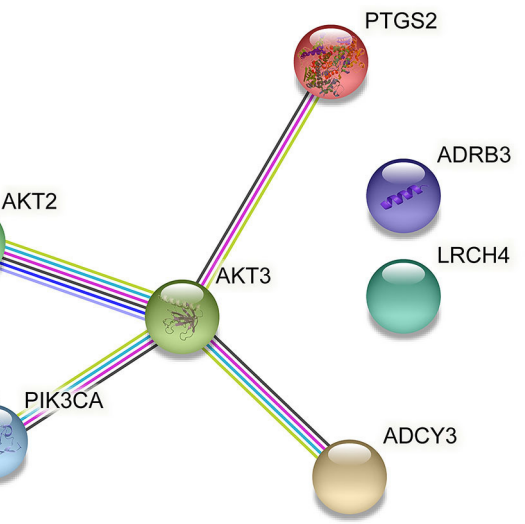


FIGURE 7 | Gene interactions enriched in Regulation of lipolysis in adipocytes.

the sequencing reliability of 13 RNAs were further verified by qRT-PCR reaction. In sum, this study explored the role of miRNAs in regulating subcutaneous fat deposition in sheep and might provide a basis for studying the molecular mechanism of miRNAs regulating lipid metabolism.

Alterations in lipolysis are often associated with obesity, including increased rates of basal lipolysis, which may contribute to the development of insulin resistance, and impaired responses to stimulate lipolysis. Obesity is characterized by excess white adipose tissue and enlarged adipocyte size due to increased triacylglycerol storage. This study found that PTGS2, AKT2, AKT3, and PIK3CA were significantly enriched in the regulation of lipolysis in adipocytes. These 4 differentially expressed genes were regulated by 4 miRNAs (NC_ 040278.1_ 37602, oar-mir-493-3p, NC_ 040278.1_ 37521 and NC_ 040255.1_ 11627). The unknown miRNA NC_040278.1_37602 was down-regulated in subcutaneous adipose tissue of Duolang sheep. It might target PTGS2 and involve in regulation of lipolysis in adipocytes pathway, thus regulating fat deposition. Fu et al. (44) found that the differentially expressed gene PTGS2 was significantly enriched in the arachidonic acid metabolism signaling pathway in high-fat diet-induced obesity mice, and these genes were closely related to glycolipid metabolism in adipose tissue. Fluoxetine has been shown to increase *in vivo* and *in vitro* hepatic lipid accumulation. Fluoxetine treatment increased mRNA expression of prostaglandin biosynthetic enzymes PTGS2 (45). The differentially expressed mRNA was combined with miRNA using the mode of up-down or down-up to construct the regulatory network. In this study, PTGS2 was up-regulated in subcutaneous adipose tissue of Duolang sheep, and it was speculated that NC_040278.1_37602 could affect the deposition of sheep subcutaneous adipose tissue by regulating PTGS2.

The oar-mir-493-3p was down-regulated in subcutaneous adipose tissue of Duolang sheep. It has been found that ssc-mir-493 is highly expressed in the skeletal muscle of Duroc pigs and interacts with PDK4 (46). PDK4 (47) is located in the mitochondrial matrix and inhibits the pyruvate dehydrogenase complex, thereby catalyzing the conversion of pyruvate to acetylcoa. Therefore, it is responsible for reducing

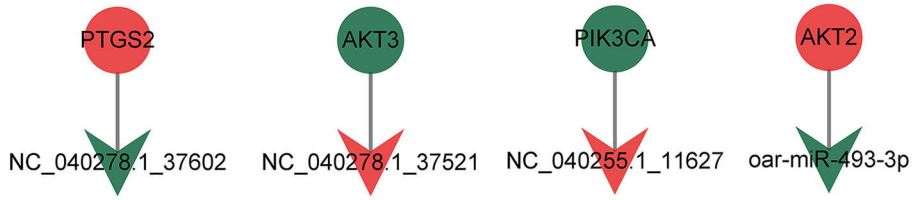


FIGURE 8 | The targeting relationship of genes in Regulation of lipolysis in adipocytes.

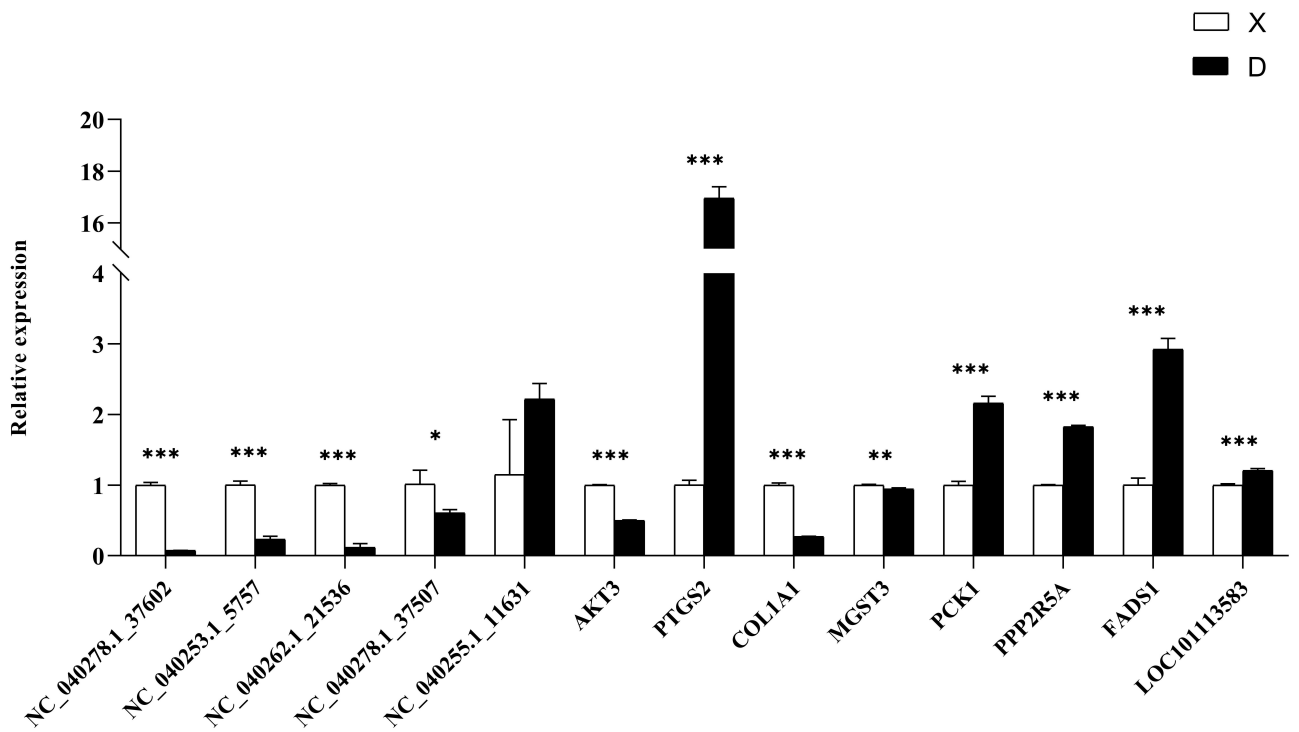


FIGURE 9 | qRT-PCR verification of differentially expressed genes. The differential expression of genes in subcutaneous adipose tissues between Duolang and Small Tail Han sheep was verified qRT-PCR. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

the utilization of glucose and the up regulation of fatty acid oxidation. It is speculated that mir-493 may be involved in the regulation of lipid metabolism by targeting PDK4. Some studies have analyzed the miRNAs in human islets, liver and skeletal muscle, and found that mir-493 is also enriched in islets, and the expression level is higher than that in liver and skeletal muscle (48). Other studies have found that mir-493 can regulate IGF1R and its downstream effector molecule mapk1. These may lay a foundation for revealing the potential regulatory mechanism of mir-493 in affecting lipid metabolism and related diseases by affecting insulin synthesis and secretion. In this study, oar-mir-493-3p might target AKT2 and involve in regulation of lipolysis in adipocytes pathway, thus regulating fat deposition. AKT, also known as protein kinase B, includes three closely related isoforms, AKT1, AKT2 and AKT3. AKT2 is abundant in brown fat and is up-regulated in 3T3-L1 adipocytes (49) and is critical for adipose tissue growth, not because it controls differentiation per se, but because it promotes lipid accumulation (50). The study found that BCAAs are associated with obesity-related metabolic disorders, BCAAs aggravate obesity-related hepatic glucolipid metabolism disorders by weakening AKT2 signaling pathway and supplementing BCAAs significantly increased hepatic gluconeogenesis in high-fat diet-induced obese mice and inhibited hepatic lipogenesis. This resulted from severe attenuation of AKT2 signaling through mTORC1 and mTORC2-dependent pathways. BCKAs suppressed AKT2 activation through mTORC1 and mTORC2 signaling and promote AKT2 ubiquitin-proteasome-dependent degradation through the mTORC2 pathway (51). In this experiment, AKT2

was up-regulated in the subcutaneous adipose tissue of Duolang sheep, and oar-mir-493-3p regulated the subcutaneous adipose tissue deposition of sheep by targeting AKT2.

The unknown miRNA NC_040278.1_37521 was up-regulated in subcutaneous adipose tissue of Duolang sheep, and it might target AKT3 and involve in regulation of lipolysis in adipocytes pathway. Studies have found that AKT3 specifically phosphorylates WNK1 at T58. Lack of AKT3 in adipocytes increases WNK1 protein levels, leading to the activation of SGK1 (52). In turn, SGK1 promotes adipogenesis by phosphorylating and inhibiting the transcription factor FOXO1, which in turn activates the transcription of PPARγ in adipocytes, thereby promoting adipogenesis, and mice lacking AKT3 have increased adipocyte numbers, white adipose tissue expansion when fed a high-fat diet, and glucose homeostasis is impaired. Thus, the interaction between AKT3, WNK1 and SGK1 regulates adipogenesis *in vivo*, and dysregulation of this pathway can lead to increased adipogenesis and obesity as well as insulin resistance (53). AKT3 was down-regulated in subcutaneous adipose tissue of Duolang sheep, and NC_040278.1_37521 affects the deposition of adipose tissue by regulating it.

The unknown miRNA NC_040255.1_11627 was up-regulated in subcutaneous adipose tissue of Duolang sheep, and it might target PIK3CA and involve in regulation of lipolysis in adipocytes pathway. The protein encoded by the PIK3CA gene is the catalytic subunit of PI3Ks, a family of lipid kinases that can specifically phosphorylate the 3-hydroxyl group of phosphatidylinositol to produce second messenger inositols (54). NC_040255.1_11627 functions by regulating PIK3CA. PIK3CA was also significantly enriched in signaling pathways such as PI3K-Akt and mTOR in this study. More and more studies have shown that the mTOR signaling pathway plays an important role in maintaining energy homeostasis and glucose and lipid metabolism. Its activation can lead to obesity. As a downstream target of the PI3K/Akt pathway, mTOR can regulate lipid synthesis through PPARγ and SREBP1 (55). Activation of AMPK inhibits the activity of mTOCR1 (56). AMPK can phosphorylate mTOCR1 to promote glucose catabolism, fatty acid oxidation and autophagy, and inhibit fatty acid and protein synthesis (57). We therefore selected miRNAs (NC_040278.1_37602, oar-miR-493-3p, NC_040278.1_37521 and NC_040255.1_11627) and target genes (PTGS2, AKT2, AKT3, and PIK3CA) as important candidate genes by functional analysis. The analysis identified pathways related to fat metabolism, including regulation of lipolysis in adipocytes, that may indirectly affect subcutaneous adipose tissue deposition in sheep.

CONCLUSIONS

In summary, miRNAs could regulate fat deposition in Duolang and Small Tail Han sheep. According to the results of RNA-Seq, miRNAs can target mRNAs and then play an important role in regulation of lipolysis in adipocytes. By this way, NC_040278.1_37602, oar-miR-493-3p, NC_040278.1_37521, and NC_040255.1_11627 might target PTGS2, AKT2, AKT3, and PIK3CA, respectively, play key regulatory roles. This study can provides valuable information to supplement the sheep miRNA database and further study the biology of sheep subcutaneous fat miRNA. Moreover, it lays a foundation for exploring the mechanism of fat deposition in sheep, which was beneficial to the development of animal husbandry.

DATA AVAILABILITY STATEMENT

The RNA-Seq data have been deposited in the NCBI Sequence Read Archive (SRA) database with accession number PRJNA801884.

ETHICS STATEMENT

All the procedures involving animals were approved by the animal care and use committee at the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (NO. IAS2019-82), where the study was conducted. All the experiments were performed in accordance with the relevant guidelines and regulations set by the Ministry of Agriculture of the People's Republic of China.

AUTHOR CONTRIBUTIONS

T-YL, HF, SY, and L-LX performed the experiment. T-YL analyzed the data. HF, SY, and L-LX interpreted the data. X-YM conceived and designed the study. T-YL and X-YM wrote the paper. All authors read and approved the final manuscript.

FUNDING

This work was supported by a grant from the Major Science and Technology Project of New Variety Breeding of Genetically Modified Organisms (Nos. 2009ZX08008-004 and 2008ZX08008-003), the Agricultural Science and Technology Innovation Programme (ASTIPIAS05) and the Basic Research Fund for Central Public Research Institutes of CAAS (Y2016JC22, Y2018PT68).

ACKNOWLEDGMENTS

We thank all the researchers who contributed to this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.938311/full#supplementary-material

REFERENCES

1. Miao, X, Luo, Q, Qin, X. Genome-wide analysis reveals the differential regulations of mRNAs and miRNAs in Dorset and small tail Han sheep muscles. *Gene.* (2015) 562:188–96. doi: 10.1016/j.gene.2015.02.070
2. Lu TX, Rothenberg ME. MicroRNA. *J Allergy Clin Immunol.* (2018) 141:1202–7. doi: 10.1016/j.jaci.2017.08.034
3. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* (1993) 75:843–54. doi: 10.1016/0092-8674(93)90529-Y
4. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of Let-7 heterochronic regulatory RNA. *Nature.* (2000) 408:86–9. doi: 10.1038/35040556
5. Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions. *Brief Bioinform.* (2014) 15:1–19. doi: 10.1093/bib/bbs075
6. Miao X, Luo Q, Qin X, Guo Y. Genome-wide analysis of microRNAs identifies the lipid metabolism pathway to be a defining factor in adipose tissue from different sheep. *Sci Rep.* (2015) 5:18470. doi: 10.1038/srep18470
7. Lee YS, Dutta A. microRNAs in cancer. *Annu Rev Pathol.* (2009) 4:199–227. doi: 10.1146/annurev.pathol.4.110807.092222
8. Komoll RM, Hu Q, Olarewaju O, von Döhlen L, Yuan Q, Xie Y, et al. MicroRNA-342-3p is a potent tumour suppressor in hepatocellular carcinoma. *J Hepatol.* (2021) 74:122–34. doi: 10.1016/j.jhep.2020.07.039
9. Sengupta D, Deb M, Kar S, Pradhan N, Parbin S, Kirtana R, et al. Dissecting miRNA facilitated physiology and function in human breast cancer for therapeutic intervention. *Semin Cancer Biol.* (2021) 72:46–64. doi: 10.1016/j.semcancer.2020.05.017
10. Shimura T, Toden S, Kandimalla R, Toiyama Y, Okugawa Y, Kanda M, et al. Genomewide expression profiling identifies a novel miRNA-based signature for the detection of peritoneal metastasis in patients with gastric cancer. *Ann Surg.* (2021) 274:e425–34. doi: 10.1097/SLA.0000000000003647
11. Ying W, Gao H, Dos Reis FCG, Bandyopadhyay G, Ofrecio JM, Luo Z, et al. Mir-690, an exosomal-derived miRNA from M2-polarized macrophages, improves insulin sensitivity in obese mice. *Cell Metab.* (2021) 33:781–90.e5. doi: 10.1016/j.cmet.2020.12.019
12. Ji C, Guo X. The clinical potential of circulating microRNAs in obesity. *Nat Rev Endocrinol.* (2019) 15:731–43. doi: 10.1038/s41574-019-0260-0
13. Zhu J, Liu B, Wang Z, Wang D, Ni H, Zhang L, et al. Exosomes from nicotine-stimulated macrophages accelerate atherosclerosis through miR-21-3p/PTEN-mediated VSMC migration and proliferation. *Theranostics.* (2019) 9:6901–19. doi: 10.7150/thno.37357
14. Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV, et al. MicroRNA-143 regulates adipocyte differentiation. *J Biol Chem.* (2004) 279:52361–5. doi: 10.1074/jbc.C400438200
15. Fan Y, Gan M, Tan Y, Chen L, Shen L, Niu L, et al. Mir-152 regulates 3T3-L1 preadipocyte proliferation and differentiation. *Molecules.* (2019) 24:3379. doi: 10.3390/molecules24183379
16. Zhang Y, Wang Y, Wang H, Ma X, Zan L. MicroRNA-224 impairs adipogenic differentiation of bovine preadipocytes by targeting LPL. *Mol Cell Probes.* (2019) 44:29–36. doi: 10.1016/j.mcp.2019.01.005
17. Aryal B, Singh AK, Rotllan N, Price N, Fernández-Hernando C. microRNAs and lipid metabolism. *Curr Opin Lipidol.* (2017) 28:273–80. doi: 10.1097/MOL.0000000000000420
18. Thum T, Condorelli G. Long noncoding RNAs and microRNAs in cardiovascular pathophysiology. *Circ Res.* (2015) 116:751–62. doi: 10.1161/CIRCRESAHA.116.303549
19. Yang Z, Cappello T, Wang L. Emerging role of microRNAs in lipid metabolism. *Acta pharmaceutica Sinica B.* (2015) 5:145–50. doi: 10.1016/j.apsb.2015.01.002
20. Shao F, Wang X, Yu J, Shen K, Qi C, Gu Z. Expression of Mir-33 from an SREBP2 intron inhibits the expression of the fatty acid oxidation-regulatory genes CROT and HADHB in chicken liver. *Br Poult Sci.* (2019) 60:115–24. doi: 10.1080/00071668.2018.1564242
21. Wu GY, Rui C, Chen JQ, Sho E, Zhan SS, Yuan XW, et al. MicroRNA-122 inhibits lipid droplet formation and hepatic triglyceride accumulation via Yin Yang 1. *Cell Physiol Biochem.* (2017) 44:1651–64. doi: 10.1159/000485765
22. Long JK, Dai W, Zheng YW, Zhao SP. Mir-122 promotes hepatic lipogenesis via inhibiting the LKB1/AMPK pathway by targeting SIRT1 in non-alcoholic fatty liver disease. *Mol Med.* (2019) 25:26. doi: 10.1186/s10020-019-0085-2
23. Girard M, Jacquemin E, Munnich A, Lyonnet S, Henrion-Caude A. Mir-122, a paradigm for the role of microRNAs in the liver. *J Hepatol.* (2008) 48:648–56. doi: 10.1016/j.jhep.2008.01.019
24. Qin N, Chen Y, Jin MN, Zhang C, Qiao W, Yue XL, et al. Anti-obesity and anti-diabetic effects of flavonoid derivative (Fla-Cn) via microRNA in high fat diet induced obesity mice. *Eur J Pharm Sci.* (2016) 82:52–63. doi: 10.1016/j.ejps.2015.11.013
25. Kumar YP, Mishra DS, Prakash P, Sethi MS. Morphological differences in longissimus dorsi muscle of pig, buffalo, sheep, goat and poultry. *Anat Anz.* (1976) 140:136–42.
26. Peng JB. *Study on the Process Optimization and Shelf Life Prediction of Fermented Mutton Jerky*. Hangzhou: Zhejiang University (2021).
27. Ge K, Ye P, Yang L, Kuang J, Chen X, Geng Z. Comparison of slaughter performance, meat traits, serum lipid parameters and fat tissue between chaohu ducks with high- and low-intramuscular fat content. *Anim Biotechnol.* (2020) 31:245–55. doi: 10.1080/10495398.2019.1664565
28. Guo Y, Zhang X, Huang W, Miao X. Identification and characterization of differentially expressed miRNAs in subcutaneous adipose between Wagyu and Holstein cattle. *Sci Rep.* (2017) 7:44026. doi: 10.1038/srep44026
29. Yilmaz O, Kizilaslan M, Arzik Y, Behrem S, Ata N, Karaca O, et al. Genome-wide association studies of preweaning growth and in vivo carcass composition traits in ESME sheep. *J Anim Breed Genet.* (2022) 139:26–39. doi: 10.1111/jbg.12640
30. Kouba M, Bonneau M, Noblet J. Relative development of subcutaneous, intermuscular, and kidney fat in growing pigs with different body compositions. *J Anim Sci.* (1999) 77:622–9. doi: 10.2527/1999.773622x
31. Kang D, Zhou G, Zhou S, Zeng J, Wang X, Jiang Y, et al. Comparative transcriptome analysis reveals potentially novel roles of homeobox genes in adipose deposition in fat-tailed sheep. *Sci Rep.* (2017) 7:14491. doi: 10.1038/s41598-017-14967-9
32. Xiao C, Wei T, Liu LX, Liu JQ, Wang CX, Yuan ZY, et al. Whole-transcriptome analysis of preadipocyte and adipocyte and construction of regulatory networks to investigate lipid metabolism in sheep. *Front Genet.* (2021) 12:662143. doi: 10.3389/fgene.2021.662143
33. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* (2008) 36:D154–8. doi: 10.1093/nar/gkm952
34. Han F, Zhou L, Zhao L, Wang L, Liu L, Li H, et al. Identification of miRNA in sheep intramuscular fat and the role of miR-193a-5p in proliferation and differentiation of 3T3-L1. *Front Genet.* (2021) 12:633295. doi: 10.3389/fgene.2021.633295
35. Mackowiak SD. Identification of novel and known miRNAs in deep-sequencing data with miRDeep2. *Curr Protoc Bioinformatics.* (2011) 36:12. doi: 10.1002/0471250953.bi1210s36
36. Zheng Y, Ji B, Song R, Wang S, Li T, Zhang X, et al. Accurate detection for a wide range of mutation and editing sites of microRNAs from small RNA high-throughput sequencing profiles. *Nucleic Acids Res.* (2016) 44:e123. doi: 10.1093/nar/gkw471
37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* (2011) 12:323. doi: 10.1186/1471-2105-12-323
38. Zhou G, Wang X, Yuan C, Kang D, Xu X, Zhou J, et al. Integrating miRNA and mRNA expression profiling uncovers miRNAs underlying fat deposition in sheep. *Biomed Res Int.* (2017) 2017:1857580. doi: 10.1155/2017/1857580
39. Fei X, Jin M, Wang Y, Li T, Lu Z, Yuan Z, et al. Transcriptome reveals key microRNAs involved in fat deposition between different tail sheep breeds. *PLoS ONE.* (2022) 17:e0264804. doi: 10.1371/journal.pone.0264804
40. Riffo-Campos ÁL, Riquelme I, Brebi-Mieville P. Tools for sequence-based miRNA target prediction: what to choose? *Int J Mol Sci.* (2016) 17:1987. doi: 10.3390/ijms17121987
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

42. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* (2000) 25:25–9. doi: 10.1038/75556
43. Kanehisa M, Goto S, Kawashima S, Okuno Y & Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* (2004) 32:D277–80. doi: 10.1093/nar/gkh063
44. Fu P, Zhu R, Jia J, Hu Y, Wu C, Cieszczyk P, et al. Aerobic exercise promotes the functions of brown adipose tissue in obese mice via a mechanism involving COX2 in the Vegf signaling pathway. *Nutr Metab.* (2021) 18:56. doi: 10.1186/s12986-021-00581-0
45. Ayyash A, Holloway AC. Fluoxetine-induced hepatic lipid accumulation is mediated by prostaglandin endoperoxide synthase 1 and is linked to elevated 15-Deoxy-Δ(12,14) PGJ. *J Appl Toxicol.* (2021) 42:1004–15. doi: 10.1002/jat.4272
46. Mármol-Sánchez E, Ramayo-Caldas Y, Quintanilla R, Cardoso TF, González-Prendes R, Tibau J, et al. Co-expression network analysis predicts a key role of microRNAs in the adaptation of the porcine skeletal muscle to nutrient supply. *J Anim Sci Biotechnol.* (2020) 11:10. doi: 10.1186/s40104-019-0412-z
47. Newhardt MF, Batushansky A, Matsuzaki S, Young ZT, West M, Chin NC, et al. Enhancing cardiac glycolysis causes an increase in PDK4 content in response to short-term high-fat diet. *J Biol Chem.* (2019) 294:16831–45. doi: 10.1074/jbc.RA119.010371
48. Bolmeson C, Esguerra JL, Salehi A, Speidel D, Eliasson L, Cilio CM. Differences in islet-enriched miRNAs in healthy and glucose intolerant human subjects. *Biochem Biophys Res Commun.* (2011) 404:16–22. doi: 10.1016/j.bbrc.2010.11.024
49. Altomare DA, Lyons GE, Mitsuuchi Y, Cheng JQ, Testa JR. Akt2 mRNA is highly expressed in embryonic brown fat and the AKT2 kinase is activated by insulin. *Oncogene.* (1998) 16:2407–11. doi: 10.1038/sj.onc.1201750
50. Sanchez-Gurmaches J, Martinez Calejman C, Jung SM, Li H, Guertin DA. Brown fat organogenesis and maintenance requires Akt1 and Akt2. *Mol Metab.* (2019) 23:60–74. doi: 10.1016/j.molmet.2019.02.004
51. Zhao H, Zhang F, Sun D, Wang X, Zhang X, Zhang J, et al. Branched-chain amino acids exacerbate obesity-related hepatic glucose and lipid metabolic disorders via attenuating Akt2 signaling. *Diabetes.* (2020) 69:1164–77. doi: 10.2337/db19-0920
52. Ding L, Zhang L, Kim M, Byzova T, Podrez E. Akt3 kinase suppresses pinocytosis of low-density lipoprotein by macrophages via a novel WNK/SGK1/Cdc42 protein pathway. *J Biol Chem.* (2017) 292:9283–93. doi: 10.1074/jbc.M116.773739
53. Ding L, Zhang L, Biswas S, Schugar RC, Brown JM, Byzova T, et al. Akt3 inhibits adipogenesis and protects from diet-induced obesity via Wnk1/Sgk1 signaling. *JCI Insight.* (2017) 2:e95687. doi: 10.1172/jci.insight.95687
54. Rathinaswamy MK, Burke JE. Class I phosphoinositide 3-kinase (PI3K) regulatory subunits and their roles in signaling and disease. *Adv Biol Regul.* (2020) 75:100657. doi: 10.1016/j.jbior.2019.100657
55. Blanchard PG, Festuccia WT, Houde VP, St-Pierre P, Brûlé S, Turcotte V, et al. Major involvement of mtor in the Pparγ-induced stimulation of adipose tissue lipid uptake and fat accretion. *J Lipid Res.* (2012) 53:1117–25. doi: 10.1194/jlr.M021485
56. González A, Hall MN, Lin SC, Hardie DG. Ampk and Tor: the Yin and Yang of cellular nutrient sensing and growth control. *Cell Metab.* (2020) 31:472–92. doi: 10.1016/j.cmet.2020.01.015
57. Herzig S, Shaw RJ. AMPK: guardian of metabolism and mitochondrial homeostasis. *Nat Rev Mol Cell Biol.* (2018) 19:121–35. doi: 10.1038/nrm.2017.95

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 | Frontiers in Genetics

REVIEW
published: 14 July 2022
doi: 10.3389/fgene.2022.907411

Opportunities of Genomics for the Use of Semen Cryo-Conserved in Gene Banks

J. Kor Oldenbroek and Jack J. Windig*

Centre for Genetic Resources, The Netherlands and Animal Breeding and Genomics, Wageningen University & Research, Netherlands, Netherlands

Shortly after the introduction of cryo-conserved semen in the main farm animal species, gene banks were founded. Safeguarding farm animal genetic diversity for future use was and is the main objective. A sampling of sires was based on their pedigree and phenotypic information. Nowadays, DNA information from cryo-conserved sires and from animals in the living populations has become available. The combination of their DNA information can be used to realize three opportunities: 1) to make the gene bank a more complete archive of genetic diversity, 2) to determine the history of the genetic diversity from the living populations, and 3) to improve the performance and genetic diversity of living populations. These three opportunities for the use of gene bank sires in the genomic era are outlined in this study, and relevant recent literature is summarized to illustrate the great value of a gene bank as an archive of genetic diversity.

Keywords: genomics, genetic diversity, priorization, optimization, introgression, selection, drift

OPEN ACCESS

Edited by:
Johann Sölkner, University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
João Gouveia, Universidade Federal do Vale do São Francisco, Brazil
N Srivastava, Indian Veterinary Research Institute (IVRI), India

****Correspondence:***
J. Kor Oldenbroek
kor.oldenbroek@wur.nl

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

Received: *29 March 2022*
Accepted: *07 June 2022*
Published: *14 July 2022*

Citation:
Oldenbroek JK and Windig JJ (2022) Opportunities of Genomics for the Use of Semen Cryo-Conserved in Gene Banks. Front. Genet. 13:907411. doi: 10.3389/fgene.2022.907411

INTRODUCTION

Conservation of Genetic Diversity in Gene Banks

Genetic diversity is an important characteristic of a population of animals. It creates the opportunity for artificial selection to improve desired traits of the animals. Genetic diversity is important in natural and captive populations because it facilitates their adaptation to a wide variety of environments. Genetic diversity is influenced by genetic drift, selection, migration, and mutation. The loss of genetic diversity within breeds, resulting in inbreeding, is mainly at stake in populations under intense selection for a few traits and in small populations with a high genetic relationship between the animals. The conservation of the genetic diversity between breeds is also relevant because recently many breeds were set aside from the mainstream production chains. These rare breeds may be a source of unique genetic diversity if they still have a sufficient effective population size (Leroy et al., 2015).

Concern about the loss of genetic diversity in farm animals has become widespread, for example, by the activities of the FAO (Oldenbroek, 2007). This loss can effectively be prevented, among other measures, by the storage of frozen semen or embryos (Smith, 1984). In the 1940's; of the past century, artificial insemination techniques were developed. The main driver in cattle was the existence of venereal diseases transferred by natural mating, which caused infertility. In the 1950's, cryo-conservation of semen was developed to facilitate the logistics and a wider use of sires not hindered by the short longevity of fresh semen. Already at the start of artificial insemination in cattle with frozen semen, Swedish AI-studs started with the long-term storage of cryo-conserved semen from each bull used for breeding (Oldenbroek, 1999).

Smith (1984) stated that "the possible returns from retaining genetic diversity may be large, while the costs for cryo-conservation in a gene bank, by comparison, are trivially small on a national basis." In his view, any breed at risk should be cryo-conserved. However, he predicted that "the continuous genetic improvement in current stocks may make it increasingly difficult for unimproved conserved stocks to compete, unless there are reversals in breeding goals or drastic changes in husbandry conditions." Despite the fact that breeds, not used by mainstream breeding programs, are presently kept alive in small numbers by motivated farmers, their existence and their genetic diversity are not safeguarded in the future by live-conservation only. These breeds are often less productive and generate less income and are kept by hobby-farmers or older professional farmers, often without a successor for their farm. Therefore, the integration of cryo-conservation of genetic diversity in a gene bank with the conservation of genetic diversity in live populations is the most powerful conservation strategy (Oldenbroek, 2007).

An example of a gene bank in progress is described by Blackburn et al. (2019). The US gene bank already contains more than one million samples from over 55,000 animals from 165 livestock and poultry breeds. The collection was developed to safeguard the genetic diversity of species and breeds important for livestock production. The oldest samples are from animals that lived 60 years ago. About 50% of the collection comprises rare breeds, with less than 1,000 animals. Their collection completeness is 45%. The completeness is calculated as a percentage of the target goal, which combines the germplasm quantity and the minimum number of animals to reconstitute a breed with an effective population size of 50. The larger populations are more complete, up to 98%. Gene bank collections are used indeed. Samples from over 6,000 animals in the collection have been used for adding diversity to breeds, genomic evaluation, reconstituting populations, or various research projects. Especially for the rare breeds, confronted with an array of obstacles not at stake in the larger populations, the gene bank is considered in the US to be the best security for the U.S. livestock sector.

Opportunities for Genomics to Use Conserved Genetics

In the era of genomics, the management of gene banks can take their decisions for sampling and for use of the donors of semen based on genomic criteria. In the past, the genetic criteria could only be based on the history of breeds and on pedigree analysis (Passemard et al., 2018). "Measures of genetic relationships and inbreeding based on pedigrees are expectations, while molecular genetic estimates of inbreeding are the particular realizations of such expectations" (Fernández and Bennewitz, 2017). This is the reason that genomic-based measures of genetic relationships between breeds measure the genetic distances among breeds more accurately. The cost of generating genomic data on a large scale has decreased sharply in recent years. For these reasons, genomic information has become the standard for choices to be made in the conservation of farm animal genetic diversity (Oldenbroek, 2017).

Nowadays, when gene bank collections become more or less complete, other questions become relevant (Passemard et al., 2018): what are the useful additions from the live populations; how to optimize the collection; and last but not least, how can these collections serve the breeding programs for the live, small, and even mainstream populations? What are the opportunities of genomic techniques in this respect? The DNA information of cryo-conserved sires and from animals in the living populations can be combined. This combination can be used 1) to make the gene bank a more complete archive of genetic diversity, 2) to determine the genetic background of the living populations, and 3) to improve the performance and genetic diversity of living populations. These three opportunities for the use of gene bank sires in the genomic era will be outlined in this study. Relevant recent literature (from 2016 onwards) will be summarized to illustrate the great value of a gene bank as an archive of genetic diversity.

PRIORITIZATION AND OPTIMIZATION PROCESSES IN GENE BANK COLLECTIONS BASED ON BETWEEN AND WITHIN BREED GENETIC DIVERSITY

Methods to Prioritize Breeds or Lines for Conservation in Gene Banks

Conservation of breeds in gene banks requires sampling and freezing of semen and other genetic material. It is labor-intensive when the sires are not used in regular breeding programs. The sampling of breeds and male animals within breeds is based on choices for which genomic similarities between breeds and animals might be decisive. Unique haplotypes and alleles may also influence such decisions. Fernández and Bennewitz (2017) described three methods useful to prioritize breeds with SNPs or WGS data: 1) the Weitzman diversity based on genetic distances and 2) the core set diversity based on genetic similarities, within and across breeds. 3) the so-called cluster analysis. The cluster approach attempts to detect an unknown number of groups (clusters) in the whole population. In this method, the genotypic data determine the structure of the clusters based on similarities and differences in genetic markers. The distance between the centers of the clusters indicates the genetic distances between breeds. The circumference of the clusters indicates the within-breed genetic diversity. In all these methods, population size plays a role. Rare breeds often have a small effective size and little within breed diversity, but they can have a large distance from other breeds. In the core set method, the little within-breed diversity is weighted against the large between-breed diversity. The outcome determines the genetic contribution of a breed to the genetic diversity of the species. In the Weitzman method, only the genetic distances to the other breeds are decisive in this respect.

The core set diversity method can be extended to the "safe set +1" approach (Eding et al., 2002). The start is the detection of the group of breeds not at risk of extinction or the breeds that are expected not to become extinct, and then the diversity stored in

that set of breeds is calculated. The breeds outside this set are added one by one, and the increase in diversity for each breed is calculated. The breeds giving the highest increase in diversity get the highest priority for conservation.

Hulsegge et al. (2019a) used a combination of core set and cluster analysis to study the relationship between the different Landrace pig lines conserved in the Dutch gene bank and their genetic relationship with the present population of Landrace pigs from Topigs Norsvin. Two clusters were identified in the conserved lines: the Norwegian/Finnish Landrace lines and the Dutch Landrace lines. These lines were bred in the past by different companies. With the gene bank samples, it was possible to assess the effect of a series of mergers of breeding companies in which lines were set aside or were used to breed the present Dutch Landrace line. Structure analysis revealed that all Landrace lines in the gene bank had a unique diversity and contributed almost equally to the present Dutch Landrace line. The core set method revealed that the genetic diversity level of the current Dutch Landrace breed was 0.89, while from the whole set, it was 0.99. Thus, a large quantity of the genetic diversity of the conserved Dutch Landrace lines in the gene bank is still present in the Dutch Landrace line of Topigs Norsvin. But, the gene bank lines harbor 10 per cent of the total genetic diversity not present in this current Dutch Landrace line.

Huson et al. (2020) compared the genetic variation of 49 heavily used Jersey bulls from the Island of Jersey to the genetic variation of 47 non-Island Jersey bulls and cows, mainly from the U.S, using a 777 K SNP chip. The Island of Jersey's population has been isolated for a long time from Jersey cattle elsewhere. Principal component analysis demonstrated that the two populations clearly differed but clustered together when Guernsey and Holstein cattle were incorporated in the analysis. The two Jersey populations demonstrated similar inbreeding levels despite large differences in population size and gene flow and slightly higher estimates of inbreeding parameters compared to the Holstein and Guernsey populations. This study provided an overview of how genetic variation in the Jersey breed was shaped, which can serve as a reference for future management of its genetic diversity.

Signer-Hasler et al. (2017) compared original local Swiss cattle breeds (Original Braunvieh (OB), Simmental (SI), Eringer (ER), and Evolèner (EV)) to more widely used breeds (Brown Swiss, Braunvieh Holstein, Red Holstein, and Swiss Fleckvieh). This was carried out with the genotypes of 9,214 cryo-conserved sires. They demonstrated low levels of genomic inbreeding and high levels of genetic diversity in the original Swiss cattle populations ER, OB, and SI compared to the other breeds and explained it by a greater use of natural service in the original Swiss breeds. The EV population had a high level of genomic inbreeding because it is regionally restricted with a low number of herd-book cows.

In several other studies, the analyzed DNA data originated partly from cryo-conserved males and partly from living animals. For example, Upadhyay et al. (2019) analyzed a mixture of gene bank bulls and live animals of nine native Swedish cattle breeds and described the genetic history and the population structure of the Swedish cattle breeds. They could detect clusters of breeds based on a common history and lines within breeds and detect differences in genetic diversity within breeds. Gautason et al. (2020) studied, partly with cryo-conserved bulls, the genetic position of Icelandic cattle, bred in isolation over 1,000 years, in the pool of Nordic and West-European breeds. They concluded that the Icelandic cattle have the highest relationship with the Finncattle breeds (Eastern, Northern, and Western Finncattle), and the Swedish Mountain cattle. Due to the long period of isolation, Islandic cattle is highly genetically distinct from the other Nordic and West-European breeds. Schmidtmann et al. (2021) analyzed ten cattle breeds belonging to the group of red dairy breeds in Northern Europe and originating from Germany, the Netherlands, and Denmark. The genomic composition was analyzed with 36195 SNP's. Genetic relationships and shared ancestries differed between the breeds. Gene flow from the Red Holstein breed to two German Angler breeds was clearly established.

Methods Useful for the Optimization of Gene Bank Collections

The characterization of samples by marker genotypes (SNP) or by Whole Genome Sequencing (WGS) of germplasm can give important additional information to the existing pedigree and phenotype-based information (Berg and Windig, 2017). The selection of animals in living populations for sampling of their semen, oocytes, or embryos for cryo-conservation can be performed more accurately with genomic information. More accurately, in the sense that the selected animals increase the genetic diversity of the samples of the species already conserved. Also, it guarantees that specific alleles or haplotypes are included in the cryo-collection. Software to estimate optimal contribution, including constraints for breeding value or genetic diversity, is available, for example,

Gencont, http://www.genebankdata.cgn.wur.nl/gencont/gencont.html.
EVA, http://www.nordgen.org, and
MateSel, http://matesel.une.edu.au.

Doekes (2020) stated that "a common strategy of national gene banks is to conserve all national breeds." But funds are usually restricted, and choices between breeds often have to be made. This may depend on which other breeds are already conserved. Due to historical events such as the time of separation or migration, closely related breeds share more diversity than less related breeds. The three main criteria for deciding which breeds should be conserved in a gene bank are genetic diversity, utility, and extinction probability (Bennewitz et al., 2007). When breeds are genotyped, we obtain more accurate estimates of the within-breed diversity. Bennewitz et al. (2007) outlined a strategy that maximizes utility in the selection of breeds for conservation. This utility combines the within breed diversity, the specific traits and the value of sustainable utilization of the breed, for example, for economic, cultural, and nature conservancy applications.

The most important aim of a breed's conservation is to establish a core collection per breed that is sufficiently large enough to reconstitute that breed when it becomes extinct. Within breeds, the main goal is to maximize genetic diversity in the core collection. For a core collection, genomics can be used

to select the animals from the living population. The actual number of samples needed in such a core collection, the strategic size consisting of the number of sires per breed and the number of samples per sire, is determined by the objectives of the collection. For the restoration of a population, the number of samples partly depends on the insemination success of the cryo-conserved semen (see FAO (2012) for more details). An optimization process for the core collection may result in moving samples from individual animals from the core collection to the working collection. Semen from the working collection can be distributed upon request for research or support of the live population.

For the selection of animals for the core collection of breeds, SNP or WGS data can be used in optimal contribution methods. SNP data have the disadvantage of ascertainment bias because they are often developed for the use in mainstream breeds. WGS also detects minor alleles present in rare breeds. Eynard et al. (2016) concluded that the amount of diversity conserved is approximately the same using SNP or WGS data in an optimal contribution method. But using WGS data, the loss in small minor alleles is much less than based on data from SNP chips.

Hulsegge et al. (2016) used the maximum diversity strategy to determine the contribution of the Dutch Red and White Friesian (DFR) cattle (gene bank samples and samples from living females) and the contribution of different lines within this breed to the national cattle gene pool. With the use of the principle component analysis, genomic relationship measurements, and the core set plus one approach, it could be concluded that: 1) the DFR breed has a small but unique contribution to the Dutch cattle gene pool; 2) it is closely related to the Black and White Dutch Friesian breed; 3) of the seven lines that can be distinguished within the DFR breed, only two lines are clearly separate. 4) one of the separate lines comprises unique diversity not present within the DFR breed nor in the rest of the Dutch cattle gene pool and 5) the second separate line comprises unique diversity for the DFR breed but its genetic diversity overlaps with that of the Holstein Friesian breed, due to the use of a few Holstein Friesian bulls in the past. Thus both the population structure of a breed and its relationship with other breeds should be taken into account in the conservation decisions for a breed.

Van Breukelen et al. (2019) characterized the genomic diversity in the gene bank for Dutch native cattle breeds with SNP data. The data set consisted of 715 bulls from seven native breeds and 165 Holstein Friesian bulls and was used to calculate genetic similarities. With optimal contribution selection, core sets of bulls were established with a minimized similarity. The gene bank was composed in the past based on pedigree information. This led to a partial optimization of semen collection. The mean similarity within breeds based on the number of straws conserved per bull was 0.32%–1.49% lower than in the case where each bull would have contributed equally. Mean similarity could be further reduced within core sets by 0.34%–2.79% using OCS.

Engelsma et al. (2014) determined the impact of conserving a specific allele in a cryo-collection. That might be a unique allele with a positive effect on the performance or an allele responsible for a genetic defect. The more selection pressure is on a specific allele, the less diversity will be conserved across the genome. This method is not attractive because genetic diversity around the targeted locus will be lost.

In conclusion, for the prioritization of breeds and lines for conservation in gene banks and for the optimization of gene bank collections, several genomic methods were developed and approved to be effective, for example, the core set diversity approach, cluster analysis, and optimum contribution selection. To obtain a complete archive of the genetic diversity of a species within a breed, the analyzed genomic data should be as accurate as possible. Whole-genome sequences are to be preferred over SNPs because WGS data give information on rare variants of alleles and are free of ascertainment bias.

THE DETERMINATION OF THE HISTORY OF THE GENETIC DIVERSITY OF THE LIVING POPULATIONS

In the history of populations, selection plays an important role in several ways. Natural selection results in changes in allele frequencies that enable the adaptation of the animals to current environments. Artificial selection within breeding programs influences the allele frequencies at QTL, leading to desired performance traits. Gene banks are reservoirs to detect the selection of breeds over recent generations by calculating the changes in allele frequencies (Boitard et al., 2021). In an ideal breeding program in full control, all sires (and females) contribute equally to the next generation. But in less controlled breeding programs (e.g., cattle and companion animals), genetic drift is a real risk. In these programs, an intense selection for performance often causes the popular sire effect (Wellmann and Bennewitz, 2019). This effect may lead to the loss of part of the genomes of founder animals over generations that may be missed in the present live populations. Then, gene bank sires might be of value for the live populations when they contain parts of the genome of the lost founders. Their use can re-introduce these parts of the genome and increase the genetic diversity in the live population. A comparison of the live and cryo-bank populations can indicate which part of the genome in the live population is missing and can be re-introduced through the use of gene bank sires with a focus on SNPs, WGS, and minor alleles. In the history of populations, often selection and genetic drift are simultaneously at stake, both resulting in inbreeding and the loss of genetic diversity. The effects of selection and genetic drift are often confounded and cannot easily be untangled.

Genetic Diversity That Might Be Lost Without Conservation Efforts

Doekes et al. (2018a) assessed the effects on genetic diversity and genetic merit by using cryo-conserved bulls born before 2015 in addition to bulls born between 2010 and 2015. Optimal contribution selection was performed to minimizing mean relatedness (thus maximize diversity) or to maximize genetic

merit but with a restriction on relatedness. They concluded that the additional merit from cryo-bank bulls could be considerable when 1) the relative emphasis on diversity was higher, or 2) the index under selection changed. The additional merit of using the cryo-bank bulls was relatively low for the total merit index currently in use but higher (in ascending order) for the sub-indices production, udder health, and fertility. They concluded that: "anticipating changes in breeding goals in the future, the germplasm collection is a valuable resource for commercial breeding populations."

WGS data detect all genetic variants, including those with a low Minor Allele Frequency (MAF), which are largely absent from Single Nucleotide Polymorphism (SNP) chips. Therefore, WGS data are expected to measure more accurately the genetic relationships in populations. Eynard et al. (2015) compared the effects of using pedigree, SNP or WGS data of 118 Holstein bulls for the prioritization of animals for conservation. The benefit of using WGS was small for common variants, but considerable for variants with a MAF below 5%. Eynard et al. (2016) also investigated the effect of optimal contribution selection based on either data from a 50 k SNP chip or based on WGS data in a population of 277 bulls from the 1,000 bull genome project. Selection with a lot of emphasis on genetic improvement gave a high risk of loss, especially for the rare variants in the population. This risk was lower when the selection was based on genetic relationships determined with WGS data.

Hulsegge et al. (2022) investigated changes in genetic diversity in Dutch Friesian (DF) cattle. This breed is one of the founding breeds of Holstein cattle, but the two breeds have been bred separately for over 100 years. In the 1970's and 1980's, Holstein cattle largely replaced DF cattle, and the latter decreased sharply in numbers. Genetic diversity was compared with WSG data between a group of cryo-conserved historic DF sires (hDF) from about 40 years ago, a group of recently used DF bulls (rDF) and a group of recently used HF bulls, respectively. A large overlap of genetic diversity exists between the three groups due to their common history. However, each of the three groups has a number of group-specific SNPs, and the two DF groups are genetically clearly different from the rHF group. The genetic difference between the rDF and rHF is slightly larger than that between the hDF and rHF. In the past 40 years, the genetic diversity was reduced in the DF breed and it became more homogeneous. However, the breeders managed to maintain a low level of inbreeding. Especially, inbreeding due to recent ancestors was largely absent in rDF.

Stronen et al. (2019) studied the effects of managing small unique lineages within breeds separately or managed together at breed level. What is the best strategy to avoid the risks of inbreeding and genetic drift due to low effective population sizes? They examined the genetic diversity of native and commercial cattle (*Bos taurus*) breeds, including the very small population of Danish Jutland cattle. They established the population structure and genetic diversity within breeds and lineages by genotyping 770 K SNP loci. They included older cryopreserved samples to determine whether the use of cryo-conserved semen is a real opportunity for the re-introduction of lost genetic diversity. They proved the genetic uniqueness of native domestic breeds and emphasized the need for diligent conservation plans, taking into account the unique lineages, in which inbreeding is balanced with carefully designed outcrossing. The use of cryo-conserved semen of founders can indeed support the preservation of traditional genetic characteristics of native domestic breeds.

Dechow et al. (2020) found that two Holstein Friesian bulls born in the 1950s determine the male lineage of more than 99% of all known Holstein artificial insemination (AI) cryo-conserved bulls in the United States. All Holstein bulls can be traced back to only two bulls born in the late 1800s. This means that the genetic variation for the Y chromosome in US Holstein bulls is very limited because the Y chromosome is only transmitted from sire to son. From two additional male lineages, semen was available in the USDA gene bank. Semen from bulls of those lineages was used to produce seven bulls and eight heifers by *in vitro* embryo production with oocytes from elite modern genetic females. The genomic breeding values of these seven bulls indicate that the lost lineages can be re-introduced in one generation using elite dams, resulting in a breed average genetic value for economic merit for the seven bulls. This genetic value was reached through a high genetic merit for fertility, a near-average genetic merit for fat and protein yield, and a below-average genetic merit for udder and physical conformation.

Runs of Homozygosity as Indicators for Genetic Diversity

Before the era of genomics, there was general consensus on how to manage genetic diversity in livestock populations. Rates of pedigree-based inbreeding and kinship (ΔF and Δf) had to be limited to <1% per generation, and pedigree-based optimal contribution selection (OCS) was the method of choice to do so. Woolliams (2007) recommended that the rates of genomic inbreeding in small populations of livestock should remain below 0.5–1% per generation. Higher rates of inbreeding should be avoided. They lead to inbreeding depression through the presence of homozygous recessive alleles with negative effects and to a deterioration of traits due to the absence of favorable dominance effects at QTL, which are expressed in heterozygotes.

Nowadays, rates of genomic-based inbreeding and kinship are the characteristics used to manage diversity. They differ from pedigree based rates of inbreeding used in the past in selection schemes (Meuwissen and Oldenbroek, 2017). When data from dense marker genotyping are used to calculate rates of inbreeding, these rates include loci directly affected by selection driving allele frequency changes. The assumption for the pedigree based rate of inbreeding is that inbreeding is determined by neutral loci not linked to loci affected by selection. In reality, these unlinked loci are unlikely to exist. Thus, the realized molecular inbreeding is expected to be higher in breeding programs without an optimal contribution strategy. This is caused by within family selection of animals that get the same advantageous chromosome regions in the process of Mendelian sampling. Thus, when a pedigree based inbreeding should not be higher than 1 per cent per generation, the genomic rate of inbreeding should not be higher than 0.5 per cent (Meuwissen and Oldenbroek, 2017).

Runs of homozygosity (ROHs) are frequently used to measure inbreeding. ROHs are long stretches of two homologous chromosomes within the same individual that are identical. They are homozygous for all the loci within these stretches (Fernández and Bennewitz, 2017). ROHs reflect Identical By Descent (IBD) because it is highly unlikely that two identical long haplotypes are not copies of the same ancestral one. It is expected that a long ROH comes from a recent ancestor, and therefore it mirrors recent inbreeding. The shorter ROHs come from more distant ancestors. The proportion of the genome that is included in such an ROH is a measure of inbreeding (F_{ROH}).

When a lot of stakeholders take part in the selection of parents and the number of their offspring, as it is the case in nearly all species except commercial pig and poultry breeding, mean inbreeding coefficients may fluctuate from generation to generation. Populations can suffer from inbreeding effects due to bottle necks in the past, a high selection intensity for a limited number of traits or a popular sire effect. Effects of selection on inbreeding and popular sire effects can be at stake in each generation. This implies that inbreeding rates from generation to generation should be carefully controlled in populations under selection, and in particular in rare breeds with a small effective size.

Doekes et al. (2021) used pedigree data and found that recent inbreeding caused more inbreeding depression than inbreeding from more distant ancestors. This pleads for the use of Genomic Optimal Contribution Selection (GOCS) with a relationship matrix based on long ROH segments. Rates of inbreeding and kinship require comparisons of average inbreeding coefficients between several different generations. Gene banks often contain bulls that played an important role in breeding previous generations. Their DNA can be used to determine genomic inbreeding and kinship rates.

Doublet et al. (2019) studied the effect of genomic selection on the genetic diversity of three French dairy breeds: Montbéliarde, Normande, and Holstein. Their data originated from (partly) cryo-conserved bulls born between 2005 and 2015. They calculated annual genetic gains and inbreeding rates based on runs of homozygosity (ROH) and pedigree data. They paid special attention to the mean ROH length within breeds before and after the implementation of genomic selection. No significant change in inbreeding rates was found in the two national breeds, Montbéliarde and Normande. A significant increase in inbreeding rate was calculated for the Holstein breed at 0.55% per year based on ROHs and 0.49% per year based on pedigree data. This is equal to a rate of 1.36 and 1.39% per generation, respectively. The mean ROH length was longer for the Holstein breed than for the other two breeds, due to higher levels of inbreeding in recent generations. They concluded that the annual genetic gain increased for bulls from the three major French dairy cattle breeds after the start of genomic selection. However, the massive use of a popular sire in the Holstein breed caused the increase in the mean ROH length.

Gautason et al. (2021) used 50 k genotypes of more than 8,000 Icelandic cattle, including 636 cryo-conserved bulls, to estimate the genomic and pedigree-based inbreeding using long ROHs. They also used 47 Icelandic bulls genotyped with a 770 k SNP to compare them with other Nordic dairy cattle breeds. Average inbreeding coefficients according to pedigree and ROHs were 0.0621 and 0.101, respectively. They also computed ROH-based autosomal inbreeding coefficients. No severe historical inbreeding was found. The effective population sizes for the years 2009–2017 according to pedigree, ROHs, genomic relationship matrix, excess of homozygosity and individual increase in inbreeding were 81, 65, 60, 58, and 92, respectively. They concluded that inbreeding rates in Icelandic cattle are in line with FAO guidelines.

Doekes et al. (2018b) evaluated genome-wide and region-specific genetic diversity and inbreeding in the Dutch-Flemish Holstein Friesian (HF) selection scheme. In recent decades this scheme changed drastically. This implies the introduction of optimal contribution selection (OCS; around 2000), a major change in the composition of the breeding goal (around 2000), and the implementation of genomic selection (GS; around 2010). Pedigree and genotype data (~75.5 k SNPs) of 6,280 cryo-conserved AI-bulls were used to estimate rates of genome-wide inbreeding and kinship used to calculate the effective population sizes. Region-specific inbreeding trends were evaluated using ROHs. The effective population size between 1986 and 2015 varied between 69 and 102. Two major divisions were established in the genome-wide inbreeding and kinship trends. Around 2000, the inbreeding and kinship levels temporarily decreased. After the introduction of genomic selection from 2010 onwards, they sharply increased, with pedigree-based, ROH-based and marker-based inbreeding rates reaching levels of 1.8, 2.1 and 2.8% per generation, respectively. Across the genome, a substantial variation in the accumulation of inbreeding was found.

Meyermans et al. (2021) studied the genetic diversity in two populations of Belgian dual-purpose red cattle breeds. They are threatened because Belgian farmers nowadays prefer more specialized cattle breeds. In total, 270 animals, including 58 cryo-conserved bulls, of the Belgian Red and Belgian White Red cattle were genotyped with medium density SNP arrays. Genetic diversity parameters were: runs of homozygosity, effective population size, and genetic distances (Fst). ROH-based genomic inbreeding coefficients were estimated at 7.0% for Belgian Red and 6.1% for Belgian White Red cattle. The two populations had a low effective population size of 68 and 86, respectively. This illustrates the threat to their existence.

In conclusion, estimations of genetic relationships in small populations carrying rare alleles or carrying alleles rare in the larger breeds should be carried out based on WGS data. When the economic value of traits is changed in a breeding program, gene bank sires may be of value to realize such breeding goals. They can also be used to re-introduce parts of the genomes of lost founders in less controlled breeding programs. ROH's measure clearly the effects of selection and genetic drift (although often confounded) on inbreeding at the level of the whole genome as well on parts of it. These effects are often transferred into effective population sizes to illustrate better the course of the genetic diversity in the population.

THE IMPROVEMENT OF THE PERFORMANCE AND THE GENETIC DIVERSITY OF LIVING POPULATIONS

Construction of Reference Populations for Breeds

Knowledge of pedigrees is the first prerequisite to starting a breeding program. But not all farmers are participants in official pedigree recording programs. Small populations of rare breeds might benefit a lot when the population size can be extended because this can prevent inbreeding. Anecdotal and phenotypic information can indicate that animals belong to a certain breed, but breeding organizations have to be sure of that before these animals can fulfill a role in their breeding program. Genomics gives the opportunity to construct a reference population of individuals whose breed of origin is recorded over several generations. Gene bank sires always have an official registered pedigree and their DNA-composition may offer an important contribution to such a reference population for their breed. It is important that the reference population comprises the total genetic diversity of the breed at stake. The SNP markers can be selected from that "complete" reference population to be able to discriminate accurately amongst the breeds.

Hulsegge et al. (2019b) constructed reference populations for the Dutch cattle breeds partly based on SNP markers of gene bank sires. For the purity test, they used a threshold value equal or higher than 0.775 for which an non-registered animal is assigned to a breed. Out of tens of thousands of SNP markers, only 133 SNPs were needed to assign animals correctly to Dutch cattle breeds. Crossbred animals and animals from foreign breeds were identified as well.

Gebrehiwot et al. (2021) developed small SNP panels that accurately estimate the total proportion of dairy breeds and determine the parents of individuals in West and East African crossbred dairy cows. In the African smallholder system, pedigrees are not officially recorded, and often crossbreeding with dairy breeds is at stake. The identified low-cost SNP assays could complete the pedigree records in African smallholder systems. They facilitate effective breeding decisions to breed animals with the desired composition of the breeds available.

Construction of Reference Populations for Genomic Selection

In the genomic era, the genomic prediction of breeding values is an important, but challenging opportunity for the programs of small local breeds (Meuwissen and Oldenbroek, 2017). It offers the opportunity to increase the performance of breeds. Genomic prediction is based on large haplotype blocks created by family structure and a small effective population size. These haplotype blocks may contain several QTL. The effects of the alleles of the individual QTL are confounded. But the combination of many small local breeds with SNP and phenotypic data in the genomic selection scheme offers a large variety of haplotype blocks that can be used for genomic prediction. Then, the effects of individual QTL, present in several breeds, may be untangled. The knowledge of haplotypes and QTLs facilitates choosing the animals for conservation.

In small local populations, it is often impossible to create a reference population of sufficient size. Then data from animals of other breeds has to be added before the genomic prediction can start (Hozé et al., 2014). But, animals from the current population have a much higher reference population value than animals from other populations. Hence, it is important to create as many as possible reference animals from the current population. Males and females with phenotypic and SNP data are relevant. Cryo-conserved sires with phenotypic data can enlarge the reference populations and make genetic relationships in the reference population and between the reference and the "test" population of young animals and embryos more accurate. The latter is important because small populations need to use all opportunities to achieve accurate genomic predictions. When variable selection genomic prediction methods are used, the across and within breed genomic predictions can be carried out effectively (Kemper et al., 2015). The incorporation of reference populations from breeds that are related to the current breed is to be preferred (Lund et al., 2014). However, the across breed reference populations need to be significantly larger than the reference populations for a single breed (Wientjes, 2016).

Marjanovic et al. (2020) studied the possibilities for genomic selection in red dairy breeds, based on genomic and phenotypic data of cryo-conserved sires. The different breed-specific reference populations were all too small for accurate genomic prediction. Therefore, they studied the effect of adding individuals from another breed. The effective number of chromosome segments (*Me*) was used to estimate the relatedness between individuals from different breeds. It can also be used to prioritize breeds for conservation. The *Me* is also used an important parameter for the accuracy of genomic prediction. The *Me* can be estimated both within a population and between two populations or breeds. It is expressed as the reciprocal of the variance of genomic relationships. The six red Dutch rare breeds indicated a high variability of *Me*. Between breeds, the *Me* ranges from ~3,500 to ~17,400. It indicates the levels of relatedness between the breeds were different. Three clusters of breeds were found: 1) the MRY, Deep Red, and Improved Red; 2) the Dutch Friesian and Dutch Belted; and 3) the Groningen White Headed. The relatedness between breeds within the first two clusters is high. However, across-breed genomic prediction is still hampered due to the low number of genotyped individuals. An increase in this number is very effective. It appeared that for each of the six breeds, 11–133 reference animals from other breeds are needed to achieve the same accuracy of genomic prediction as an additional individual from the same breed.

Methods for Introgression of Traits

Sometimes a gene bank contains animals with a trait that is not present in a living population, but that trait can have a high value for that population. Then, introgression of the desired allele(s) responsible for that trait in the donor animals into the recipient population can be considered (Meuwissen and Oldenbroek,

2017). Here, two methods for introgression from a donor to a recipient breed will be discussed: 1) the transfer of an allele or 2) the transfer of set of alleles. The two can be realized by crossing parents from the donor breed to the recipient breed, followed by systematic backcrossing with the recipient breed. In each generation, parents for the next generation are chosen who carry the desired allele(s). The donor breed may be a small local breed that contains a desirable trait. The recipient maybe a mainstream breed which is lacking the desired trait. Further it is assumed that the allele(s) coding for the desired trait is (are) known, or narrow QTL regions with the allele(s) are known. This is known as Marker Assisted Introgression (MAI). It is a useful tool to introduce traits from a (conserved) donor population into a (mainstream) recipient population. A short generation interval facilitates this process. It is only worthwhile for a trait that has indeed a high value for the recipient breed. Hence, this value should compensate for the loss of several generations of selection in the mainstream breed because the donor breed often has lower performance.

Genomic introgression (GI) is a method to introgress a trait from a cryo-conserved donor population into a mainstream recipient population in the case of a complex trait. The genetic architecture of the trait is unidentified, for example, the architecture is complex and many genes determine the trait (Meuwissen and Oldenbroek, 2017). Applying GI, the first step is to produce crossbred offspring from donor and recipient animals. This will increase the genetic variance for the trait of interest in the offspring. The second step is a genomic selection program to improve the total performance. The weight of the trait of interest in the selection program is sufficient to obtain a positive selection response. The animals from the donor breed usually have lower performance than the recipient breed. Therefore, the higher variance in the crossbred generation is used to create higher genetic progress in subsequent generations.

Historically, introgression has been used to upgrade and to improve an important trait in breeds. The FecB fecundity QTL is first found in Australian Booroola sheep. This QTL, improving the litter size significantly, has been introgressed into a large number of other breeds (Fogarty, 2009).

Polledness in cattle has received increasing interest in recent years for welfare reasons. It is based on two different dominant alleles that are situated very close together in the cattle genome (Allais-Bonnet et al., 2013). In the dominating mainstream dairy breeds, polledness was absent. They are found in smaller breeds, especially in northern Europe (Cozzi et al., 2015), where they are fixed or segregating. Cryo-collections of breeds may contain the relevant alleles. Genomics can identify the carriers, which can be used for introgression. But, in many breeds where polledness was never observed, it is unlikely to detect carriers in cryo-conserved semen (Windig et al., 2015).

Methods for Removing Unwanted Introgression of Foreign Breeds

Many breeds have experienced migration or introgression in the past. Herdbooks often want or have to register only purebred animals. By selecting animals that minimize genomic co-ancestry between current animals and animals carrying the introgressed parts of the genome, the original genetic background of a breed can be recovered. With genomic information, this expires more efficiently than with pedigree information (Amador et al., 2013; Amador et al., 2014).

Kohl et al. (2020) found that applying standard Optimum Contribution Selection methodology in small local breeds with historical introgression could lead to a more intense selection of introgressed genetic material. The reason is that the introgressed alleles improve the rate of genetic gain and reduce the average kinship as an outcome of OCS. Consequently, small local breeds become extinct. In a simulation study, they used the advanced OCS (aOCS) approach that takes into account the introgressed genetic material. They created populations from the historical gene pool by using aOCS and took care that the simulated populations were comparable with real data. Historical breeding decisions that favored introgressed material could have been avoided by using aOCS. The genetic gain would have been at least 12.2% lower. However, the presence of introgressed genetic material, the genetic diversity, and native genetic diversity would have been more satisfactory for a small local breed whose breed purity should be enhanced.

Wang et al. (2017) found In the German Angler and Vorderwald cattle, a significant positive correlation between Migrant Contributions (MC) and estimated breeding values of the selection candidates. This means that traditional OCS would increase MC. They included MC in OCS and modified the kinships that account for the breed origin of alleles. Three OCS alternatives were simulated, taking into account minimizing kinship, minimizing MC, and maximizing genetic gain in the offspring generation. In the simulations, the inbreeding rate should not become higher than 1%, and at least 30% of the maximum progress should be achieved for all other criteria. In the traditional OCS (reference scenario), the highest breeding values were found with the restriction of classical kinship. In this case, the magnitude of MC in the progeny generation was not in control. When constraining or minimizing MC, the kinship of native alleles increased compared to the reference scenario. They concluded that including MC and kinship at native alleles is necessary when you want to maintain the genetic originality and the diversity of native alleles in breeding program aiming at genetic gain and control of inbreeding.

Schaler et al. (2018) studied the possibilities of reverse introgression in two local red cattle breeds from Northern Germany. They had pedigree data for 90,783 individuals for the German Angler breed and 187,255 individuals for the Red Dual-Purpose cattle breed. Information on sex, date of birth, breed percentage, and conventional breeding values was available. The native genetic contribution of individuals could be included as an additional trait in the total merit index as an attempt to recover a part of the native genetic background. Marker information that accounts for Mendelian sampling improved the native contributions. The maintenance of a sufficient genetic diversity of native alleles needs an advanced OCS with proper constraints.

In conclusion, gene bank sires have well recorded pedigrees and are used to construct reference populations to test breed purity. In combination with their breeding values for performance traits, they can strengthen the relationships in populations needed in genomic selection schemes. Methods for the introgression of traits from gene bank sires into living populations are well developed and effective. Optimum contribution methodology is available and is used to recover the native genetic background of breeds in which crossbreeding took place in the past.

OVERALL CONCLUSION

Gene banks can be considered as a sustainable and reliable archive of genetic diversity. Genomics gives three relevant stakeholders important tools to improve their efforts to conserve and use genetic diversity: 1) Gene bank management has the opportunity to use genomics to prioritize breeds and lines and animals within breeds for conservation and to optimize the collections. They may facilitate the creation of reference populations to test animals for breed purity and for genomic selection. At request, they can provide straws for genomic research and for the activities of rare breed associations and commercial companies. 2) rare breed associations have the opportunity to use genomics to add non-recorded animals to their populations, to re-use sires whose genomes are no longer present in the populations, to monitor the relationship and inbreeding over generations and to remove parts of the genome of other introgressed breeds, and 3) commercial breeding companies have the opportunity to use the genotypic and phenotypic data of gene bank sires into their reference populations for genomic selection, may re-use sires when afterwards it reveals that parts of their genome are interesting and no longer present in their current breeding sires, can monitor kinship and inbreeding over generations and can consider the introgression of interesting genes in gene bank sires not present in their current breeding animals.

AUTHOR CONTRIBUTIONS

JKO and JJW wrote the manuscript together.

REFERENCES

Allais-Bonnet, A., Grohs, C., Medugorac, I., Krebs, S., Djari, A., Graf, A., et al. (2013). Novel Insights into the Bovine Polled Phenotype and Horn Ontogenesis in Bovidae. *PLoS ONE* 8 (5), e63512. doi:10.1371/journal.pone.0063512

Amador, C., Fernández, J., and Meuwissen, T. H. (2013). Advantages of Using Molecular Coancestry in the Removal of Introgressed Genetic Material. *Genet. Sel. Evol.* 45, 13. doi:10.1186/1297-9686-45-13

Amador, C., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic Selection for Recovery of Original Genetic Background from Hybrids of Endangered and Common Breeds. *Evol. Appl.* 7, 227–237. doi:10.1111/eva.12113

Bennewitz, J., Eding, H., Ruane, J., and Simianer, H. (2007). "Chapter 6: Selecting Breeds for Conservation," in *Utilisation and Conservation of Farm Animal Genetic Resources*. Editor K. Oldenbroek (Wageningen: Academic Publishers). ISBN 978-90-8686-032-6.

Berg, P., and Windig, J. J. (2017). "Chapter 6: Management of Cryo-Collections with Genomic Tools," in *Genomic Management of Animal Genetic Diversity* (Netherlands: Wageningen Academic Publishers). ISBN: 978-90-8686-297-9.

Blackburn, H. D., Wilson, C. S., and Krehbiel, B. (2019). Conservation and Utilization of Livestock Genetic Diversity in the United States of America through Gene Banking. *Diversity* 11 (12), 244. doi:10.3390/d11120244

Boitard, S., Paris, C., Sevane, N., Servin, B., Bazi-Kabbaj, K., and Dunner, S. (2021). Gene banks as reservoirs to detect recent selection: the example of the Asturiana de los Valles Bovine breed. *Front. Genet.* 12, 575405. doi:10.3389/fgene.2021.575405

Cozzi, G., Gottardo, F., Brscic, M., Contiero, B., Irrgang, N., Knierim, U., et al. (2015). Dehorning of Cattle in the EU Member States: A Quantitative Survey of the Current Practices. *Livest. Sci.* 179, 4–11. doi:10.1016/j.livsci.2015.05.011

Dechow, C. D., Liu, W. S., Specht, L. W., and Blackburn, H. (2020). Reconstitution and Modernization of Lost Holstein Male Lineages Using Samples from a Gene Bank. *J. Dairy Sci.* 103, 4510–4516. doi:10.3168/jds.2019-17753

Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J., and Windig, J. J. (2018a). Value of the Dutch Holstein Friesian Germplasm Collection to Increase Genetic Diversity and Improve Genetic Merit. *J. Dairy Sci.* 101, 10022. doi:10.3168/jds.2018-15217

Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J., and Windig, J. J. (2018b). Trends in Genome-wide and Region-specific Genetic Diversity in the Dutch-Flemish Holstein-Friesian Breeding Program from 1986 to 2015. *Genet. Sel. Evol.* 50, 15. doi:10.1186/s12711-018-0385-y

Doekes, H. P., Bijma, P., and Windig, J. J. (2021). How Depressing Is Inbreeding? A Meta-Analysis of 30 Years of Research on the Effects of Inbreeding in Livestock. *Genes* 12, 926. doi:10.3390/genes12060926

Doekes, H. P. (2020). *Genomic Characterization and Conservation of Genetic Diversity in Cattle*. PhD thesis. Wageningen, Netherlands: Wageningen University, 216. ISBN: 978-94-6395-423-5. doi:10.18174/523321

Doublet, A.-C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., et al. (2019). The Impact of Genomic Selection on Genetic Diversity and Genetic Gain in Three French Dairy Cattle Breeds. *Genet. Sel. Evol.* 51, 52. doi:10.1186/s12711-019-0495-1

Eding, H., Crooijmans, R. P., Groenen, M. A., and Meuwissen, T. H. (2002). Assessing the Contribution of Breeds to Genetic Diversity in Conservation Schemes. *Genet. Sel. Evol.* 34, 613–633. doi:10.1186/1297-9686-34-5-613

Engelsma, K. A., Veerkamp, R. F., Calus, M. P. L., and Windig, J. J. (2014). Consequences for Diversity when Animals are Prioritized for Conservation of the Whole Genome or of One Specific Allele. *J. Anim. Breed. Genet.* 131, 61–70. doi:10.1111/jbg.12052

Eynard, S. E., Windig, J. J., Leroy, G., van Binsbergen, R., and Calus, M. (2015). The Effect of Rare Alleles on Estimated Genomic Relationships from Whole Genome Sequence Data. *BMC Genet.* 16, 24. doi:10.1186/s12863-015-0185-0

Eynard, S. E., Windig, J. J., Hiemstra, S. J., and Calus, M. P. L. (2016). Whole-genome Sequence Data Uncover Loss of Genetic Diversity Due to Selection. *Genet. Sel. Evol.* 48, 33. doi:10.1186/s12711-016-0210-4

FAO (2012). *Cryoconservation of Animal Genetic Resources*. Rome: FAO Animal Production and Health. Guidelines No. 12.

Fernández, J., and Bennewitz, J. (2017). "Chapter 2: Defining Genetic Diversity Based on Genomic Tools," in *Genomic Management of Animal Genetic Diversity* (Netherlands: Wageningen Academic Publishers). ISBN: 978-90-8686-297-9.

Fogarty, N. M. (2009). A Review of the Effects of the Booroola Gene (FecB) on Sheep Production. *Small Ruminant Res.* 85, 75–84. doi:10.1016/j.smallrumres.2009.08.003

Gautason, E., Schönherz, A. A., Sahana, G., and Guldbrandtsen, B. (2020). Relationship of Icelandic Cattle with Northern and Western European Cattle Breeds, Admixture and Population Structure. *Acta Agric. Scand. Sect. A Animal Sci.* 69 (1-2), 25–38. doi:10.1080/09064702.2019.1699951

Gautason, E., Schönherz, A. A., Sahana, G., and Guldbrandtsen, B. (2021). Genomic Inbreeding and Selection Signatures in the Local Dairy Breed Icelandic Cattle. *Anim. Genet.* 52, 251–262. doi:10.1111/age.13058

Gebrehiwot, N. Z., Strucken, E. M., Marshall, K., Aliloo, H., and Gibson, J. P. (2021). SNP Panels for the Estimation of Dairy Breed Proportion and Parentage Assignment in African Crossbred Dairy Cattle. *Genet. Sel. Evol.* 53, 21. doi:10.1186/s12711-021-00615-4

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., and Croiseau, P. (2014). Efficiency of Multi-Breed Genomic Selection for Dairy Cattle Breeds with Different Sizes of Reference Population. *J. Dairy. Sci.* 97, 3918–3929. doi:10.3168/jds.2013-7761

Hulsegge, B., Calus, M. P. L., Oldenbroek, J. K., and Windig, J. J. (2016). Conservation Priorities for the Different Lines of Dutch Red and White Friesian Cattle Change when Relationships with Other Breeds are Taken into Account. *J. Anim. Breed. Genet.* 134, 69–77. doi:10.1111/jbg.12233

Hulsegge, I., Calus, M., Hoving-Bolink, R., Lopes, M., Megens, H.-J., and Oldenbroek, K. (2019a). Impact of Merging Commercial Breeding Lines on the Genetic Diversity of Landrace Pigs. *Genet. Sel. Evol.* 51, 60. doi:10.1186/s12711-019-0502-6

Hulsegge, I., Schoon, M., Windig, J., Neuteboom, M., Hiemstra, S. J., and Schurink, A. (2019b). Development of a Genetic Tool for Determining Breed Purity of Cattle. *Livest. Sci.* 223, 60–67. doi:10.1016/j.livsci.2019.03.002

Hulsegge, I., Oldenbroek, K., Bouwman, A., Veerkamp, R., and Windig, J. (2022). Selection and Drift: A Comparison between Historic and Recent Dutch Friesian Cattle and Recent Holstein Friesian Using WGS Data. *Animals* 12, 329. doi:10.3390/ani12030329

Huson, H. J., Sonstegard, T. S., Godfrey, J., Hambrook, D., Wolfe, C., Wiggans, G., et al. (2020). A Genetic Investigation of Island Jersey Cattle, the Foundation of the Jersey Breed: Comparing Population Structure and Selection to Guernsey, Holstein, and United States Jersey Cattle. *Front. Genet.* 11, 366. doi:10.3389/fgene.2020.00366

Kemper, K.E., Reich, C. M., Bowman, P. J., van der Jagt, C. J., Chamberlain, A. J., Mason, B. A., et al. (2015). Improved Precision of QTL Mapping Using a Nonlinear Bayesian Method in a Multi-Breed Populations Leads to a Greater Accuracy of Across-Breed Genomic Predictions. *Genetics Selection Evolution* 47, 29.

Kohl, S., Wellmann, R., and Herold, P. (2020). Implementation of Advanced Optimum Contribution Selection in Small-Scale Breeding Schemes: Prospects and Challenges in Vorderwald Cattle. *Animal* 14 (3), 452–463. doi:10.1017/S1751731119002295

Leroy, G., Besbes, B., Boettcher, P., Hoffmann, I., Capitan, A., and Baumung, R. (2015). Rare Phenotypes in Domestic Animals: Unique Resources for Multiple Applications. *Anim. Genet.* 47, 141–153. doi:10.1111/age.12393

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., and Brøndum, R. F. (2014). Genomic Evaluation of Cattle in a Multi-Breed Context. *Livest. Sci.* 166, 101–110. doi:10.1016/j.livsci.2014.05.008

Marjanovic, J., Hulsegge, B., and Calus, M. (2020). Relatedness between Numerically Small Dutch Red Dairy Cattle Populations and Possibilities for Multibreed Genomic Prediction. *J. Dairy Sci.* 104, 4498–4506. doi:10.3168/jds.2020-19573

Meuwissen, T. H. E., and Oldenbroek, J. K. (2017). "Chapter 5: Management of Genetic Diversity Including Genomic Selection in Small *In Vivo* Populations," in *Genomic Management of Animal Genetic Diversity* (Netherlands: Wageningen Academic Publishers). ISBN: 978-90-8686-297-9.

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2021). Genomics Confirm an Alarming Status of the Genetic Diversity of Belgian Red and Belgian White Red Cattle. *Animals* 11, 3574. doi:10.3390/ani11123574

Oldenbroek, J. K. (1999). "Chapter 1: Introduction," in *Genebanks and the Conservation of Farm Animal Genetic Resources* (Lelystad, Netherlands: DLO-Institute for Animal Science and Health). ISBN: 90-75124-06-6.

Oldenbroek, J. K. (2007). "Chapter 1: Introduction," in *Utilisation and Conservation of Farm Animal Genetic Resources* (Netherlands: Wageningen Academic Publishers). ISBN 978-90-8686-032-6.

Oldenbroek, J. K. (2017). *Genomic Management of Animal Genetic Diversity*. Netherlands: Wageningen Academic Publishers. ISBN: 978-90-8686-297-9.

Passemard, A. S., Hiemstra, S. J., Tixier-Boichard, M., and Danchin-Burge, C. (2018). "Mapping the Diversity and Characteristics of European Farm Animal Genetic Collections: Banks or Museums?," in Proceedings of the World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand, February 7–11, 2018 (Auckland, New Zealand).

Schaler, J., Wellmann, R., Bennewitz, J., Thaller, G., and Hinrichs, D. (2018). Genetic Diversity and Historic Introgression in German Angler and Red Dual Purpose Cattle and Possibilities to Reverse Introgression. *Acta Agric. Scand. Sect. A Animal Sci.* 68 (2), 63–72. doi:10.1080/09064702.2019.1600011

Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M., HinrichsThaller, D. G., et al. (2021). Assessing the Genetic Background and Genomic Relatedness of Red Cattle Populations Originating from Northern Europe. *Genet. Sel. Evol.* 53, 23. doi:10.1186/s12711-021-00613-6

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., et al. (2017). Population Structure and Genomic Inbreeding in Nine Swiss Dairy Cattle Populations. *Genet. Sel. Evol.* 49, 83. doi:10.1186/s12711-017-0358-6

Smith, C. (1984). Genetic Aspects of Conservation in Farm Livestock. *Livest. Prod. Sci.* 11, 37–48. doi:10.1016/0301-6226(84)90005-8

Stronen, A. V., Pertoldi, C., Iacolina, L., Kadarmideen, H. N., and Kristensen, T. N. (2019). Genomic Analyses Suggest Adaptive Differentiation of Northern European Native Cattle Breeds. *Evol. Appl.* 12, 1096–1113. doi:10.1111/eva.12783

Upadhyay, M., Eriksson, S., Mikko, S., Strandberg, E., Stålhammar, H., Groenen, M. A. M., et al. (2019). Genomic Relatedness and Diversity of Swedish Native Cattle Breeds. *Genet. Sel. Evol.* 51, 56. doi:10.1186/s12711-019-0496-0

Van Breukelen, A. E., Doekes, H. P., Windig, J. J., and Oldenbroek, K. (2019). Characterization of Genetic Diversity Conserved in the Gene Bank for Dutch Cattle Breeds. *Diversity* 11, 229–313. doi:10.3390/d11120229

Wang, Y., Bennewitz, J., and Wellmann, R. (2017). Novel Optimum Contribution Selection Methods Accounting for Conflicting Objectives in Breeding Programs for Livestock Breeds with Historical Migration. *Genet. Sel. Evol.* 49, 45. doi:10.1186/s12711-017-0320-7

Wellmann, R., and Bennewitz, J. (2019). Key Genetic Parameters for Population Management. *Front. Genet.* 10, 667. doi:10.3389/fgene.2019.00667

Wientjes, Y. C. J. (2016). *Multi-population Genomic Prediction*. PhD Thesis. Netherlands: Wageningen University, 268.

Windig, J. J., Hoving-Bolink, R. A., and Veerkamp, R. F. (2015). Breeding for Polledness in Holstein Cattle. *Livest. Sci.* 179, 96–101. doi:10.1016/j.livsci.2015.05.021

Woolliams, J. (2007). "Chapter 7: Genetic Contributions and Inbreeding," in *Utilisation and Conservation of Farm Animal Genetic Resources*. Editor K. Oldenbroek (Netherlands: Wageningen Academic Publishers). ISBN 978-90-8686-032-6.

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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/fgene.2022.940736

High-Density Genomic Characterization of Native Croatian Sheep Breeds

Ivana Drzaic[1], Ino Curik[1], Boris Lukic[2], Mario Shihabi[1], Meng-Hua Li[3], Juha Kantanen[4], Salvatore Mastrangelo[5], Elena Ciani[6], Johannes A. Lenstra[7] and Vlatka Cubric-Curik[1]**

[1]Department of Animal Science, University of Zagreb Faculty of Agriculture, Zagreb, Croatia, [2]Department of Animal Production and Biotechnology, Faculty of Agrobiotechnical Sciences Osijek, Chair for Domestic Animal Breeding and Genetics, J. J. Strossmayer University of Osijek, Osijek, Croatia, [3]College of Animal Science and Technology, China Agricultural University, Beijing, China, [4]Production Systems, Natural Resources Institute Finland (Luke), Jokioinen, Finland, [5]Dipartimento di Scienze Agrarie, Alimentari e Forestali, University of Palermo, Palermo, Italy, [6]Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Universita Degli Studi di Bari "Aldo Moro", Bari, Italy, [7]Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands

A recent comprehensive genomic analysis based on 50K SNP profiles has shown that the regional Balkan sheep populations have considerable genetic overlap but are distinctly different from surrounding breeds. All eight Croatian sheep breeds were represented by a small number of individuals per breed. Here, we genotyped 220 individuals representing the native Croatian sheep breeds (Istrian Sheep, Krk Island Sheep, Cres Island Sheep, Rab Island Sheep, Lika Pramenka, Pag Island Sheep, Dalmatian Pramenka, Dubrovnik Sheep) and mouflon using the Ovine Infinium® HD SNP BeadChip (606,006 SNPs). In addition, we included publicly available Balkan Pramenka and other Mediterranean sheep breeds. Our analyses revealed the complex population structure of Croatian sheep breeds and their origin and geographic barriers (island versus mainland). Migration patterns confirmed the historical establishment of breeds and the pathways of gene flow. Inbreeding coefficients ($F_{ROH>2\ Mb}$) between sheep populations ranged from 0.025 to 0.070, with lower inbreeding coefficients observed in Dalmatian Pramenka and Pag Island Sheep and higher inbreeding in Dubrovnik sheep. The estimated effective population size ranged from 61 to 1039 for Krk Island Sheep and Dalmatian Pramenka, respectively. Higher inbreeding levels and lower effective population size indicate the need for improved conservation management to maintain genetic diversity in some breeds. Our results will contribute to breeding and conservation strategies of native Croatian sheep breeds.

Keywords: Croatian sheep breeds, effective population size, genomic characterization, inbreeding, population structure

OPEN ACCESS

Edited by:
Ottmar Distl,
University of Veterinary Medicine Hannover, Germany

Reviewed by:
Guillermo Giovambattista, CONICET Institute of Veterinary Genetics (IGEVET), Argentina
Tereza Manousaki, Hellenic Centre for Marine Research (HCMR), Greece

****Correspondence:***
Ivana Drzaic
ikovac@agr.hr
Vlatka Cubric-Curik
vcubric@agr.hr

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

Received: *10 May 2022*
Accepted: *16 June 2022*
Published: *15 July 2022*

Citation:
Drzaic I, Curik I, Lukic B, Shihabi M, Li M-H Kantanen J, Mastrangelo S, Ciani E, Lenstra JA and Cubric-Curik V (2022) High-Density Genomic Characterization of Native Croatian Sheep Breeds. Front. Genet. 13:940736. doi: 10.3389/fgene.2022.940736

INTRODUCTION

Sheep were domesticated about 10,000 years ago in the region of Anatolia and, along with goats, were among the first domesticated farm animals. Sheep were first hunted by humans and over time became managed wild populations, then kept in controlled herds, and finally, humans began to breed sheep (Zeder, 2008; Larson and Burger, 2013). Trought the history breeding for diseriable traits was always present but only about 200 years ago, the organized formation of sheep breeds began, resulting in fragmented populations, and reduced genetic diversity (Taberlet et al., 2008). Genetic diversity is defined as the variety of alleles and genotypes present in a population (Frankham et al.,

2002). Each breed has a unique genetic characteristic due to mutations and drift caused by geographic isolation and bottlenecks, artificial selection and adaptation to climate, nutrition and diseases and parasites (Barker, 2001). Many sheep breeds are local breeds that have adapted to specific locations for thousands of years and are closely associated with culture and history (Gandini and Villa, 2003). These small, unique breeds contribute to the overall diversity of the species, but diversity within breeds is low. Highly productive breeds have displaced local breeds and unique gene combinations of some local breeds are being lost (Groeneveld et al., 2010). Erosion of within-breed diversity could be a problem even for breeds whose overall populations remain very large. Monitoring population trends is a prerequisite for rapid and effective action to protect breeds from extinction. Measures to prevent the loss of livestock diversity will be more effective if the factors that drive genetic erosion and extinction risk are well understood. Maintaining genetic diversity is important for rapid adaptation to challenges (Andersson, 2012). Thus, characterizing genetic diversity is an important aspect of developing sustainable strategies for breed improvement (Groeneveld et al., 2010).

Livestock breeding, especially sheep species, was an important part of agriculture in Croatia. The first written records of sheep breeding in Croatia date back to 1781, and two decades later, the number of sheep per citizen in Europe was highest in Dalmatia (Defilipis, 1966), a region of Croatia whose name comes from Illyrian words *dalma,* meaning sheep. Since then, the number of sheep on Croatian territory has steadily decreased. The erosion of agricultural production combined with wars led to a drastic decrease in the number of some native Croatian breeds, which were threatened by extinction. Maintaining genetic diversity in livestock production is critical for meeting future challenges such as climate change, emerging diseases, and food security for a growing human population (Barker, 2001; Groeneveld et al., 2010). Croatia has a number of unique sheep breeds. The native Croatian sheep breeds belong to the Pramenka type, which is characterized by coarse wool, low production, and high resistance to environmental conditions. Within the Pramenka type, there are many breeds/populations that differ in size, wool quality and colour, as well as in their adaptation to the specific microclimatic conditions in their breeding area. Throughout history, Croatian farmers have been eager to improve production characteristics and therefore imported high-performance breeds. Merino rams (Pag Island Sheep, Rab Island Sheep, Krk Island Sheep, Cres Island Sheep and Dubrovnik sheep) and meat breeds such as Merinolandschaf (Dalmatian Pramenka and Lika Pramenka) or Bergamasca Sheep (Istrian Sheep) were mainly used to improve primitive Pramenka type. With the progress of selection, different breeds were created. Genetic diversity in sheep needs to be estimated to identify unique breeds that may be in danger of extinction.

The genetic study of native Croatian breeds started with β-lactoglobulin polymorphisms in Pag Island Sheep (Cubric-Curik et al., 2002). Since then, scientists have studied genetic diversity (Feligini et al., 2005; Ivanković et al., 2005; Cubric-Curik et al., 2009; Kasap et al., 2021; Špehar et al., 2022) and population structure (Lawson Handley et al., 2007; Ćinkulov et al., 2007; Ferencakovic et al., 2012; Salamon et al., 2014) of native Croatian sheep breeds using different approaches. Recently, 50 K SNP markers (Ciani et al., 2020) and whole-genome sequences (Deng et al., 2020; Lv et al., 2021) have been used in population genetic studies presenting native Croatian breeds, but with a very small number of animals per breed. Ciani et al. (2020) showed the separation of Balkan Pramenka from the rest of the European breeds and the belonging of native Croatian sheep breeds to the Pramenka cluster. The authors suggested that Balkan breeds are evolutionary connected with the domestication process and are one of the hub regions from which the migration of sheep spread to the rest of Europe. Recently, Shihabi et al. (2022) identified specific adaptive selection signals on the X chromosome of the individuals used in this analysis.

The main objective of this study was to evaluate the conservation status (diversity, inbreeding, and effective population size), population structure, and admixture of eight native Croatian sheep breeds. Our analysis was based on high-density autosomal SNP chip genotype information sampled from 20 to 45 individuals per breed. Our study is an extension of the study by Ciani et al. (2020), in which only a few individuals representing several native Croatian breeds were analyzed using 50 K SNP chip information. More specifically, the data used refer to a sample of 201 individuals representing eight native Croatian breeds (Cres Island Sheep, Dalmatian Pramenka, Dubrovnik Ruda, Istrian Sheep, Krk Island Sheep, Lika Pramenka, Pag Island Sheep, and Rab Island Sheep) and 10 mouflon sampled in Croatia. They provide an accurate estimate of genetic diversity, ROH-based inbreeding, current and historical effective population size, and allow in-depth comparison with other Mediterranean breeds and some other Pramenka sheep breeds.

MATERIALS AND METHODS

Sample Collection

The Croatian sheep represented in this study were: Istrian Sheep (ISS), Pag Island Sheep (PIS), Krk Island Sheep (KIS), Cres Island Sheep (CIS), Rab Island Sheep (RIS), Lika Pramenka (LPS), Dalmatian Pramenka (DPS) and Dubrovnik sheep (DRS) breeds, and Mouflon (EMC). The description of the native Croatian breeds is in **Supplementary Data S1**. These eight sheep breeds represent different geographical regions in Croatia and can be divided into island breeds (ISS, CIS, KIS, RIS and PIS) and mainland breeds (LPS, DPS, and DRS). Sampling locations are the places of origin of the breeds and are presented in **Figure 1**. For breeds statistics, see **Supplementary Table S1**. For this study, blood or tissue samples were randomly collected from 107 female and 105 male sheep located on 105 family farms representing 20 Cres Island Sheep, 20 Krk Island Sheep, 20 Rab Island Sheep, 20 Lika Pramenka, 25 Istrian Sheep, 26 Dubrovnik sheep, 26 Dalmatian Pramenka, 45 Pag Island Sheep, and 10 mouflons. Blood was collected by superficial venipuncture using sterile 10 ml EDTA Vacutainers (BD, Becton, Dickinson and Company, Oxford, United Kingdom) and stored at −86°C until further use. Tissue samples were collected using the Allflex TSU applicator (Allflex,

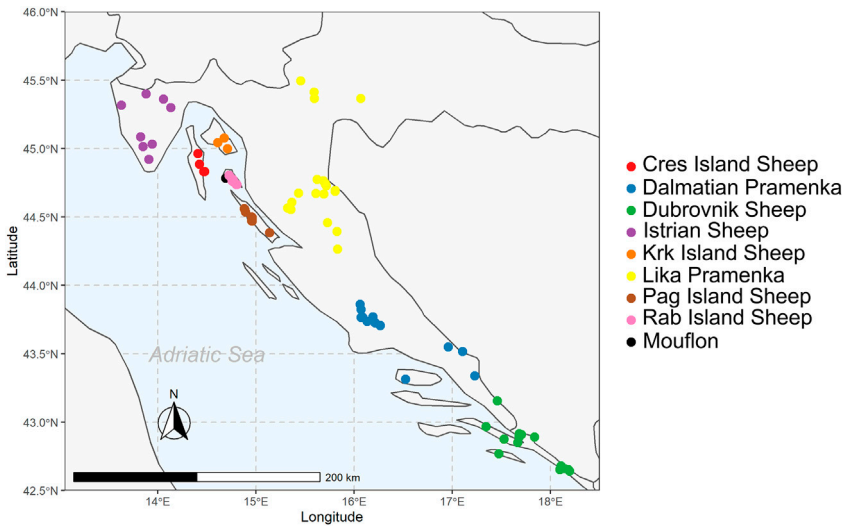


FIGURE 1 | Map of sampling locations for native Croatian sheep breeds. The map illustrates the geographic locations where samples were collected. Each breed is presented with a different colour.

France) and stored at +4°C until further use. All samples were collected according to national and European ethical protocols and directives.

DNA Extraction and SNP Genotyping

DNA extraction was performed using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Germany) according to the sample preparation and extraction protocol. DNA quality was checked by 1% agarose gel electrophoresis, while DNA quantity was determined using a NanoPhotometer P330 spectrophotometer (IMPLEN, Germany).

Croatian sheep and mouflon samples were genotyped using the Ovine Infinium HD BeadChip (Illumina, San Diego, CA, United States) by Gene Seek, Neogen Genomics (Neogene Europe Ltd., Scotland, United Kingdom). Genotypes were mapped to the Oar4.0 map. Quality control and filtering of SNPs were obtained using SNP & Variation Suite v8.7.0 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The accuracy and efficiency of SNP genotyping were assessed by applying a cut-off value of 0.7 for the Illumina GenCall score and 0.4 for the Illumina GenTrain score. All SNPs with more than 10% missing genotypes and individuals with more than 5% missing genotypes were removed. SNPs without position or SNPs that had been assigned to sex chromosomes or mitochondrial genome were also excluded. After quality control, a total of 201 sheep and 10 mouflons remained with 470,962 SNPs.

To compare the genetic relationships between Croatian sheep breeds and other Mediterranean sheep and to investigate population structure and admixture patterns, a reference panel of four Balkan Pramenka type sheep (Valachian—VAL, Sumavaska—SUM, Serbian Pramenka—PRA, Carpathian Mountain sheep—MKS) obtained from Cao et al. (2020), one Pramenka from this study (North—Macedonia Pramenka—NMS) and 13 other publicly available Mediterranean sheep breeds (Merino—MER, Noire du Velay—NVE, Causse du Lot—CDL, Rava—RAV, Blanche du Massif central—BMC, Meat Lacune—LAM, Dairy Lacune—LAC, Limousine—LIM, Tarasconnaise—TAR, Corse—COR, Manech Tete Rouge—MTR, Prealpes du Sud—PAS, Mourerous—MOU, Altamurana—ALT, Leccese - LEC) was used. Two datasets were created: I. Croatian dataset with 201 animals and 470962 SNPs and II. Mediterranean dataset with 574 animals (Croatian sheep included) and 348,968 SNPs (after merging and quality control). Map of the breed origin for the Mediterranean dataset is presented in **Figure 2A**.

Population Divergence and Relationship

To estimate genetic diversity within the population, expected heterozygosity, observed heterozygosity and the inbreeding coefficient (F_{IS}) were calculated using the software PLINK 1.9 (Purcell et al., 2007). In addition, Wright's F_{ST} fixation indices were calculated to determine the degree of genetic differentiation among selected sheep breeds using SNP & Variation Suite v8.7.0 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) and mean F_{ST} was calculated. Principal Component Analysis (PCA) was performed to investigate the genetic relationship among Croatian sheep breeds as well as at the Mediterranean level. PCA was performed for both datasets: Mediterranean and Croatian, using the R package "SNPRelate" (Zheng et al., 2012). Phylogenetic relationships based on Nei's genetic distances (Nei, 1978) were presented with a Neighbor network using SplitsTree4 software (Huson and Bryant, 2006). Nei genetic distances were calculated in the program R using the package "stAMPP" (Pembleton et al., 2013).

Population Structure

Genetic relationships and population structures were assessed with the software STRUCTURE v.2.3.4. (Pritchard et al., 2000)

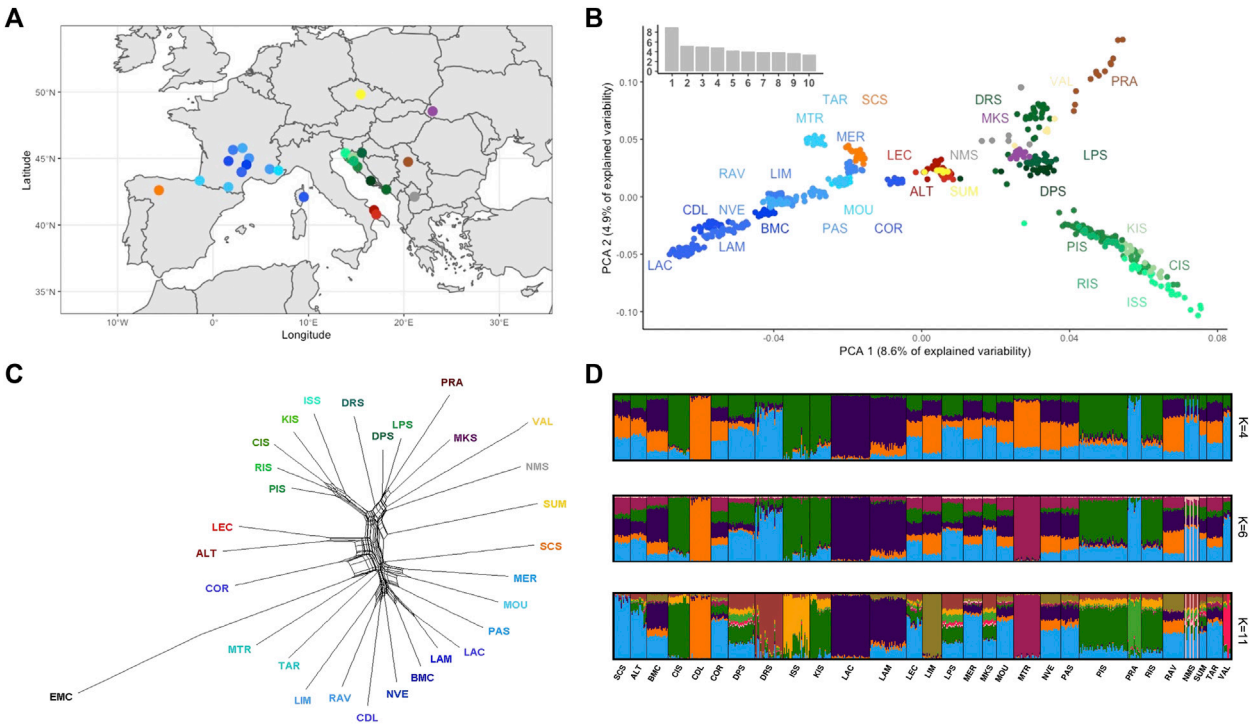


FIGURE 2 | (A) Map of the origin of the breeds used in the Mediterranean dataset; **(B)** PCA plot for Mediterranean sheep breeds with principal component 1 and principal component 2.; **(C)** Neighbour network based on the Nei genetic distances constructed for Mediterranean dataset; **(D)** Graphical presentations of the population structure analyses for a sample of 574 Mediterranean sheep. Each sheep is represented by a single vertical line broken into K colour segments, with lengths proportional to the estimated membership of the inferred cluster. Each country is presented with different colour: Croatia—green; France—blue; Spain—orange; Italy—red; North-Macedonia—grey; Ukraine—purple; Czech Republic—yellow; Serbia—brown. Within a country, each breed is represented with a different main colour and three-letter coding: DRS, Dubrovnik Sheep; LPS, Lika Pramenka; DPS, Dalmatian Pramenka; PIS, Pag Island Sheep; RIS, Rab Island Sheep; CIS, Cres Island Sheep; KIS, Krk Island Sheep; ISS, Istrian Sheep; PRA, Serbian Pramenka; VAL, Valachian; MKS, Mount Carpatian Sheep; NMS, North Macedonian Pramenka; SUM, Sumavaska; ALT, Altamunrana; LEC, Leccese; SCS, Churra; COR, Corse; MER, Merino; NVE, Noire du Velay; CDL, Causse du Lot; RAV, Rava; BMC, Blanche du Massif Central; LAM, Meat Lacaune; LAC, Milk Lacaune; LIM, Limousine; TAR, Tarasconnaise; MTR, Manech Tete Rouge; PAS, Preaples du Sud; MOU, Mourerous.

using the Bayesian model based on 83,795 SNPs in eight Croatian sheep breeds and 104,623 SNPs in 28 Mediterranean sheep breeds after pruning based on LD. Pruning of the data was performed in a sliding window of 50 SNPs, moving in increments of five SNPs along each chromosome and removing one SNP pair at a time with a pairwise $r^2 > 0.1$ using the software PLINK 1.9 (Purcell et al., 2007). After pruning the SNP missing rate was 0.000–0.003 and 0.000–0.031 per individual for Croatian dataset and Mediterannean dataset, respectively, and 0.003–0.011 for Mourerous and Altamurana population, respectively. For the Croatian and Mediterranean datasets, an ancestry model was constructed for a putative number of ancestral populations of 1–10 and 1–30, respectively. For each K value, 10 runs of 100,000 Markov chain Monte Carlo iterations were performed after a burn-in period of 10,000 iterations. The most likely number of clusters was determined using the ΔK method (Evanno et al., 2005) implemented in software Structure Selector (Li and Liu, 2018). Results from STRUCTURE were analyzed and visualized using the software Structure Selector (Li and Liu, 2018).

Gene Flow

Gene flow between Croatian and Mediterranean sheep breeds was examined using TreeMix v. 1.13 (Pickrell and Pritchard, 2012). TreeMix was applied to infer population trees and migration events between divergent populations. To determine the position of the root in the maximum likelihood tree, the European mouflon (n = 10) was defined as the outgroup population. Between 0 and 15 migration events were inferred.

Genomic Inbreeding Based on Runs of Homozygosity

The proportion of the genome that is autozygose is estimated by identification of Runs Of Homozygosity (ROH) with SNP & Variation Suite v8.7.0. ROH was detected for each individual separately using the following criteria: (I) the minimum number of SNPs in ROH was set to 50, which was calculated using the method proposed by Lencz et al. (2007) to minimize the number of false-positive ROH:

TABLE 1 | Genetic diversity indices for native Croatian sheep breeds. H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient; $F_{ROH>2\ Mb}$, genomic inbreeding calculated for ROH length over 2 Mb; N_{e0}, current effective population size. Each breed is represented with a three-letter coding: DRS, Dubrovnik Sheep; LPS, Lika Pramenka; DPS, Dalmatian Pramenka; PIS, Pag Island Sheep; RIS, Rab Island Sheep; CIS, Cres Island Sheep; KIS, Krk Island Sheep; ISS, Istrian Sheep.

Breed	H_o ± SE	H_e ± SE	F_{IS} ± SE	$F_{ROH>2\ Mb}$ ± SE	N_{e0}
CIS	0.340 ± 0.004	0.341 ± 0.000	0.002 ± 0.011	0.049 ± 0.011	148
DPS	0.346 ± 0.003	0.348 ± 0.000	0.004 ± 0.008	0.025 ± 0.008	1039
DRS	0.339 ± 0.005	0.341 ± 0.000	0.005 ± 0.015	0.070 ± 0.015	157
ISS	0.342 ± 0.004	0.336 ± 0.000	-0.017 ± 0.011	0.053 ± 0.010	161
KIS	0.340 ± 0.004	0.340 ± 0.000	0.001 ± 0.012	0.058 ± 0.011	62
LPS	0.345 ± 0.004	0.347 ± 0.000	0.005 ± 0.012	0.033 ± 0.012	598
PIS	0.339 ± 0.003	0.345 ± 0.000	0.016 ± 0.008	0.035 ± 0.008	1005
RIS	0.337 ± 0.006	0.345 ± 0.000	0.022 ± 0.018	0.055 ± 0.017	559

$$l = \frac{\ln \frac{\alpha}{n_s * n_i}}{\ln(1 - het)}$$

where α is the percentage of false-positive ROH (set to 0.05 in this study), n_s is the number of SNPs per individual, n_i is the number of individuals, *het* is heterozygosity across all SNPs; (II) the maximal gap between adjacent SNPs was set to 250 Kb; (III) the minimum SNP density per ROH was set to 1 SNP every 50 Kb; (IV) the minimum length constituting ROH was set to 2 Mb. To account for genotyping error, ROHs were calculated separately for each of the five categories specified according to the ROH length: >2 Mb, 2–4 Mb, 4–8 Mb, 8–16 Mb, and >16 Mb. Based on ROH length, the number of allowed heterozygotes and missing genotypes were defined according to Ferenčaković et al. (2013).

The individual genomic inbreeding coefficients based on ROH (F_{ROH}) were estimated as the sum of the length of all ROH per individual divided by the total length of the autosomal genome covered by SNPs, as described by McQuillan et al. (2008). The F_{ROH} was calculated for different minimum ROH lengths because the short segments are associated with individual autozygosity derived from a distant common ancestor, whereas long ROH segments are correlated with recent inbreeding (Ferenčaković et al., 2013). For each animal, we calculated the $F_{ROH2–4\ Mb}$, $F_{ROH4–8\ Mb}$, $F_{ROH8–16\ Mb}$, and $F_{ROH>16\ Mb}$ based on ROH with different minimum lengths. Based on the estimated individual $F_{ROH>2\ Mb}$, the breed inbreeding coefficient was calculated by averaging the F_{ROH} estimates per breed.

Effective Population Size

To better understand the demographic history of native Croatian sheep breeds, we estimated current and historical effective population size (N_e) form linkage disequilibrium (LD). Trends in effective population size were estimated using an optimization method that implements a genetic algorithm (Mitchell, 1998) to infer the recent demographic history of a population from SNP data of a small sample of contemporary individuals, as implemented in the software GONE (Santiago et al., 2020). Default parameters were applied. The most recent estimate of N_e was taken in the current generation. Furthermore, N_e estimates from 50 generations ago were used for comparison with results from other authors (Kijas et al., 2012; Beynon et al., 2015; Deniskova et al., 2019).

RESULTS AND DISCUSSION

Population Divergence and Relationships

Genetic diversity parameters are presented in **Table 1**. Observed heterozygosity ranged from 0.327 ± 0.006 in Rab Island Sheep to 0.346 ± 0.003 in Dalmatian Pramenka. The average expected heterozygosity within the breeds ranged from 0.336 in Istrian Sheep to 0.348 in Dalmatian Pramenka. The inbreeding coefficient (F_{IS}) varied from −0.017 ± 0.011 in Istrian Sheep to 0.022 ± 0.018 in Rab Island Sheep. The results of observed and expected heterozygosity are similar to those found by Al-Mamun et al. (2015) in five Australian sheep breeds using a 50k SNP chip and Edea et al. (2017) in Ethiopian sheep using a 50 and 600K SNP chip. Similar values of expected heterozygosity were reported for native Italian breeds (Ciani et al., 2014), southern and western European sheep breeds (Kijas et al., 2012) and native Russian sheep breeds (Deniskova et al., 2018).

Genetic relationships between individuals of multiple sheep breeds were assessed by principal component analysis (PCA). The PCA plot of the first and second principal components (PC) is shown in **Figure 2B**. The breeds were generally grouped according to their geographic origin. Similar results were also presented in Rochus et al. (2020) and Ciani et al. (2020). Total variation explained by the PC components was 94.52%. The largest component (8.6% of the total variation) showed an east-west cline between the Western European sheep and the Balkan Pramenka sheep. The second PC separated Croatian island breeds (ISS, CIS, RIS, KIS, PIS) from the rest of the Pramenka cluster and highlighted a division of native Croatian breeds into mainland breeds (DRS, LPS, and DPS) and island breeds. Two Croatian breed groups have contrasting structures. In the mainland group, breeds were clearly separated and individuals had clear cluster assignments, indicating greater differentiation between breeds. In contrast, the island breeds showed less differentiation with tighter clusters and breed overlapping. A similar differentiation in the structure of the two French sheep groups was observed in Rochus et al. (2018). Two Italian breeds occupy a central position between Balkan Pramenka sheep and French and Spanish sheep. The

genetic proximity of Pramenka sheep, Italian and Spanish breeds is interpreted as a possible sheep migration route along the Mediterranean coast from the Balkans to Spain *via* Italy, as presented in Ciani et al. (2020).

To better estimate the relationships between native Croatian sheep breeds PCA was performed only on the Croatian dataset (**Supplementary Figure S1**). The first two components together explained 14.9% of the total variation, representing a large portion of the variability. Principal component analysis showed a clear separation of Istrian Sheep and Dubrovnik Sheep, while Lika Pramenka and Dalmatian Pramenka created one close cluster and Rab Island Sheep, Pag Island Sheep, Krk Island Sheep, and Cres Island Sheep formed a second cluster. Mainland and island cluster are not well defined here. The result shows a north-south cline following the geographical breeding area of each breed. A similar pattern following the geographical distribution of breeds in the PCA graph was found in Ciani et al. (2014) and Edea et al. (2017). The largest PC (8.1% of the total variation) positioned Dubrovnik Sheep apart from the rest of the Croatian native sheep breeds. The second PC (6.8% of the total variation) separated the Istrian Sheep from all others. The genetic uniqueness of Istrian Sheep was observed earlier in Ćinkulov et al. (2007), Lawson Handley et al. (2007), and Salamon et al. (2014) using microsatellite markers. Istrian Sheep and Dubrovnik Sheep are also geographically isolated from other native Croatian sheep breeds, but they are also phenotypically the most diverse.

The relationship between the Mediterranean sheep breeds and the mouflon was determined using genetic distances according to Nei (1972), which include a correction for drift and mutations. Higher genetic distances also indicate a longer temporal divergence of the breeds. The European mouflon was set as an outgroup. Neighbor-net analysis of Mediterranean sheep (**Figure 2C**) confirms the separation of the Pramenka from the rest of the European sheep population. The genetic uniqueness of the Croatian island breeds is emphasized here by their separation from the Pramenka group. Analysis by Ciani et al. (2020) using a 50K SNP chip shows similar separation of Croatian island sheep breeds from other European breeds and their positioning between Pramenka and Italian sheep. The Lika and Dalmatian Pramenka were in a group with the Serbian Pramenka and the Ukrainian Mountain Carpathian sheep, confirming the relationships identified by Ciani et al. (2020). The separation of Croatian island and continental breeds suggests that geographic barriers have an influence on shaping the current structure of native Croatian sheep breeds. Greater geographic distances with barriers such as the sea and mountains have reduced the exchange of genetic material between these two groups of breeds, resulting in greater separation of these breeds. Earlier gene flows and population mixing, which most likely occurred between the Balkan Pramenka populations, are still visible in the genetic structure of the Croatian mainland breeds and should not be neglected here. In the past, a Pramenka was bred on the territory of today's Republic of Croatia, the so-called Balkan Pramenka, which was divided into different types depending on where it was bred (e.g., Travnik type, Sjenica type . . .). Later, with the disintegration of Yugoslavia, the gene flow between Pramenka types decreased and new breeds were formed, which still show a great genetic connection. The more frequent and stronger introduction of Italian breeds into island populations, especially Pag Island Sheep, is reflected in the proximity of these two sheep groups in the graph, although in this dataset were no representatives of Italian Merino sheep involved in breeding native Croatian sheep breeds.

The level of genetic differentiation based on F_{ST} estimation between breeds were calculated. Genetically similar populations will have lower F_{ST} values, while breeds that are more genetically diverse should have higher F_{ST} values (Berendse et al., 2009). F_{ST} values for Croatian breeds and selected breed pairs are presented in **Table 2**. The breeds that were marginal on the PCA plot were selected. The lowest F_{ST} value was observed between Lika Pramenka and Dalmatian Pramenka (F_{ST} = 0.003), indicating high genetic similarity between the two breeds. The most divergent breeds among the selected populations were Istrian Sheep and North-Macedonian Pramenka (F_{ST} = 0.081). The significant genetic distance of Istrian Sheep and Dubrovnik Sheep observed in PCA was consistent with the higher F_{ST} values obtained when comparing Istrian Sheep and Dubrovnik Sheep with native Croatian sheep breed. The mean F_{ST} value was calculated for each selected breed. The lowest mean F_{ST} was observed for Dalmatian Pramenka (MF_{ST} = 0.029). The breed with the lowest mean F_{ST} is considered as the central breed, since it has the highest genetic similarity with all other breeds.

Population Structure

The population structure of Mediterranean breeds was identified using a model-based STRUCTURE analysis with an assumed number of populations set to K = 30. The Evanno ΔK method did not reveal a clear number of genetic clusters, peaking at K = 4, K = 6, and K = 11 (**Supplementary Figure S2**), indicating a hierarchical population structure. The results of model-based population structuring revealed a complex genetic structure of native Croatian sheep breeds when placed in the Mediterranean context (**Figure 2D**). At K = 2, we observed the separation of Balkan Pramenka from Western European breeds. A similar split is presented in the genetic diversity results of 57 European and Middle Eastern sheep breeds by Peter et al. (2007). At K = 3, the French breeds Causse du Lot (CDL) and Manech Tête Rouge (MTR) showed different genetic backgrounds. The separation of breeds continued as the number of ancestral populations increased, with the separation of Dubrovnik Sheep at K = 9 (data not shown) and Istrian Sheep at K = 11. Dalmatian Pramenka and Lika Pramenka have similar population structure as North Macedonian Pramenka (NMS), Sumavska sheep (SUM) and Mountain Carpathian sheep (MKS). They show a very complex ancestral structure. Very similar results were obtained by Ciani et al. (2020) FineStructure analysis of Eastern and Southeastern European sheep breeds on a 50K SNP chip, showing incomplete differentiation of the Balkan Pramenka sheep and a common ancestral genome. This result is an indication that the defined breeds do not match the actual population boundaries and suggests an exchange of genetic material between these populations. Considering the political situation in which Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia and Slovenia were united

TABLE 2 | Genetic differentiation (F_{ST}) between native Croatian sheep breeds and selected European breeds. MF_{ST}—average F_{ST} estimates among breeds.

Breed	DPS	TAR	NMS	RIS	PIS	NVE	MER	LPS	LEC	LAC	KIS	ISS	DRS	CIS	MF_{ST}
DPS-Dalmatian Pramenka															0.029
TAR-Tarasconnaise	0.033														0.048
NMS-North Macedonian Sheep	0.042	0.067													0.064
RIS-Rab Island Sheep	0.016	0.041	0.056												0.034
PIS-Pag Island Sheep	0.015	0.039	0.053	0.011											0.033
NVE-Noire du Velay	0.037	0.044	0.072	0.045	0.043										0.052
MER-Merino	0.033	0.040	0.062	0.042	0.039	0.046									0.048
LPS-Lika Pramenka	0.003	0.036	0.045	0.021	0.020	0.041	0.037								0.032
LEC-Leccese	0.023	0.039	0.060	0.032	0.030	0.045	0.040	0.027							0.041
LAC-Lacaune (Dairy)	0.041	0.047	0.074	0.049	0.048	0.047	0.049	0.045	0.048						0.055
KIS-krk Island Sheep	0.029	0.055	0.070	0.024	0.025	0.059	0.056	0.033	0.046	0.063					0.047
ISS-Istrian Sheep	0.040	0.066	0.081	0.040	0.041	0.070	0.068	0.044	0.057	0.074	0.051				0.058
DRS-Dubrovnik Sheep	0.026	0.053	0.064	0.041	0.039	0.058	0.051	0.030	0.045	0.061	0.053	0.065			0.050
CIS-Cres Island Sheep	0.031	0.056	0.072	0.020	0.022	0.060	0.058	0.036	0.046	0.064	0.034	0.053	0.055		0.047
ALT-Altamurana	0.034	0.051	0.070	0.043	0.041	0.055	0.050	0.038	0.030	0.058	0.056	0.068	0.055	0.057	0.050

until 25 years ago, the exchange of genetic material between Pramenka sheep was very likely, and since there is no organized selection in the populations of Dalmatian Pramenka and Lika Pramenka, five generations are not sufficient for genetic differentiation of these breeds.

PCA and population structure analysis have clearly shown that among the native Croatian sheep breeds there is a certain degree of population mixing, which is lowest in Istrian Sheep and Dubrovnik Sheep.

To identify population structure within native Croatian sheep breeds on a finer scale, we performed another STRUCTURE analysis only on the Croatian dataset with an assumed number of populations (K) between one and ten. The most informative number of ancestral populations (K = 3) was estimated using the Evanno ΔK method (**Supplementary Figure S3**). The results for K = 2–6 are presented in **Supplementary Figure S4**. The clearest ancestry components when assuming only two ancestral populations showed a clear separation of the Dubrovnik Sheep. At K = 4, the Istrian Sheep are completely separated. The specific genetic structure of the Istrian Sheep was also shown in the research of Ciani et al. (2014), where it stands out as a distinct breed at a very low K = 8. At K = 5, Dalmatian Pramenka and Lika Pramenka form a clearly separate group. These two breeds are not separated even at K = 10, the maximum number of hypothesized groups for this data set that suggest common ancestry. At K = 6, the Krk Island Sheep is divided into three subpopulations, with two subpopulations showing the dominance of a single ancestral genome, while the third population is genetically similar to other island breeds. The cause of such separation within the Krk Island Sheep breed could be due to reduced exchange of genetic material within Krk Island Sheep populations, when under the influence of inbreeding and genetic drift, gene fixation occurs and population structure changes.

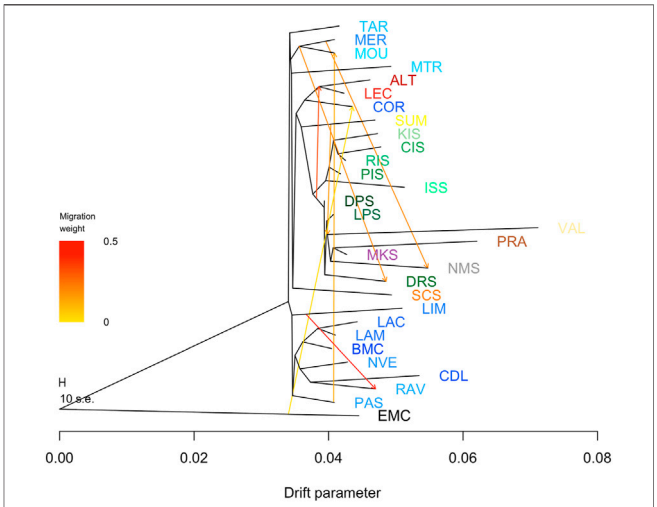


FIGURE 3 | Inferred maximum likelihood population graph considering 12 migration events. Migration arrows are coloured according to their weight. The length of the horizontal branches is proportional to the amount of genetic drift that occurred on each branch. The tree was derived using TreeMix v.1.12. The European mouflon was set as the outgroup. Each country is presented with different colour: Croatia—green; France—blue; Spain—orange; Italy—red; North-Macedonia—grey; Ukraine—purple; Czech Republic—yellow; Serbia—brown. Within a country, each breed is represented with a different main colour and three-letter coding: DRS, Dubrovnik Sheep; LPS, Lika Pramenka; DPS, Dalmatian Pramenka; PIS, Pag Island Sheep; RIS, Rab Island Sheep; CIS, Cres Island Sheep; KIS, Krk Island Sheep; ISS, Istrian Sheep; PRA, Serbian Pramenka; VAL, Valachian; MKS, Mount Carpatian Sheep; NMS, North Macedonian Pramenka; SUM, Sumavaska; ALT, Altamunrana; LEC, Leccese; SCS, Churra; COR, Corse; MER, Merino; NVE, Noire du Velay; CDL, Causse du Lot; RAV, Rava; BMC, Blanche du Massif Central; LAM, Meat Lacaune; LAC, Milk Lacaune; LIM, Limousine; TAR, Tarasconnaise; MTR, Manech Tete Rouge; PAS, Preaples du Sud; MOU, Mourerous.

Gene Flow

Migration patterns between native Croatian and Mediterranean sheep breeds were studied using population tree models accounting for migration events and implemented in the program TreeMix. The maximum likelihood tree, created without assuming migration events, was calculated and rooted in the European Mouflon (EMC) (**Supplementary Figure S5**). The ML tree generally confirmed the relationships revealed by the

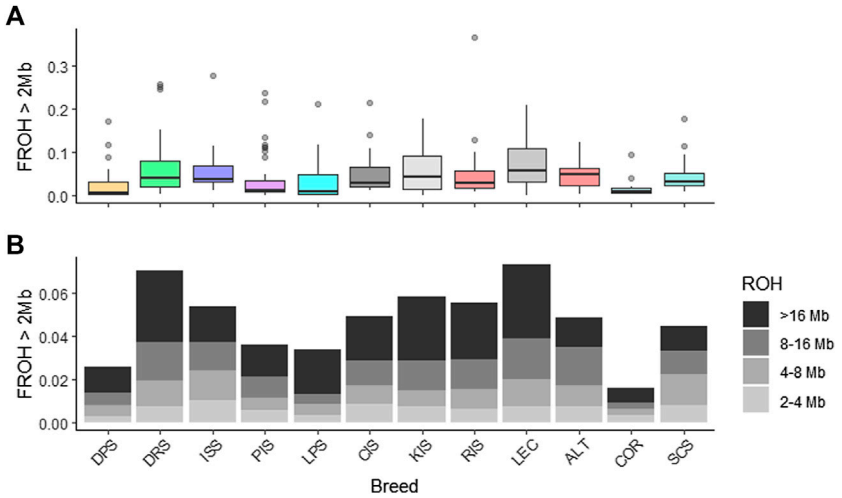


FIGURE 4 | Levels of genomic inbreeding. The genomic inbreeding coefficient (F_{ROH}) was estimated as the proportion of runs of homozygosity (ROH) in the total length of autosomes. **(A)** Individual F_{ROH} are averaged per breed. **(B)** The population mean F_{ROH} is presented for each ROH length category. Breeds are represented with a three letter coding: DRS, Dubrovnik Sheep; LPS, Lika pramenka; DPS, Dalmatian pramenka; PIS, Pag Island Sheep; RIS, Rab Island Sheep; CIS, Cres Island Sheep; KIS, Krk Island Sheep; ISS, Istrian Sheep; ALT, Altamunrana; LEC, Leccese; SCS, Churra; COR, Corse.

PCA and STRUCTURE analysis, as in the Mediterranean context the Croatian native sheep breeds were clustered with the other Balkan Pramenka sheep. The Croatian island breeds additionally separated from the other Balkan Pramenka breeds, creating a new branch. As in the PCA graph, the Italian breeds together with the Corse sheep represent the connection between the Balkan and West-European breeds. The West-European sheep breeds show a similar phylogenetic structure as shown in Rochus et al. (2018). In the population tree, the sheep breeds, including the native Croatian breeds, generally had short branch lengths, indicating higher genetic diversity within the breed. By adding migration events, the population tree showed a more detailed view of the genetic relationships between the Mediterranean and Croatian breeds. Estimation of the optimal number of migration events using the ΔM method showed that the population tree with m = 7 migrations best explained the genetic relationships between the breeds (**Figure 3**). Inference of population trees with m = 7 showed that the migration event with the greatest weight was from the root of the Balkan Pramenka sheep breeds to the root of the two Italian sheep breeds. There are no historical records of the importation of Balkan sheep into Italy, so it is assumed that this is an ancient migration, and it would be interesting to estimate how old it is. The migration edge from the French Merino branch to the Dubrovnik Sheep branch confirms the evolutionary history of the Dubrovnik Sheep, which has a higher presence of the Merino genotype than other native Croatian sheep breeds. The Dubrovnik Sheep is also known by the name Ruda, which was given to the Dubrovnik Sheep in reference to its finer wool. The migration edge with low weight from European mouflon to Corse sheep is presented. This migration is explained by the knowledge that European mouflon living in European countries are imported from populations in Sardinia and Corsica (Guerrini et al., 2015) and that the introduction of mouflon into domestic sheep populations has been recorded in Sardinia and Corsica (Barbato et al., 2017). Migrations between French breeds are the result of breed formation or crossbreeding and are discussed in detail in Rochus et al. (2018).

Genomic Inbreeding

In sheep populations, estimation of inbreeding coefficient based on pedigree information has not been accurate due to incomplete pedigrees, low generations and frequent errors in recording (Ferenčaković et al., 2013), so estimation of inbreeding coefficients based on high-density SNP markers by ROHs gives us more accurate results.

The overall genomic inbreeding ($F_{ROH>2\ Mb}$) for Croatian native sheep breeds and three selected breeds from the Mediterranean dataset (Altamurana, Leccese, Churra and Corse sheep) was estimated as well as inbreeding for each of the defined categories ($F_{ROH2-4\ Mb}$, $F_{ROH4-8\ Mb}$, $F_{ROH8-16\ Mb}$, $F_{ROH>16\ Mb}$, **Figure 4**).

Inbreeding coefficients ($F_{ROH>2\ Mb}$) between sheep populations ranged from 0.025 to 0.070 (**Table 1**), with lower inbreeding coefficients observed in Dalmatian Pramenka and Pag Island Sheep. Similar inbreeding coefficients were reported for native Russian breeds (Deniskova et al., 2019) and most Italian sheep breeds (Mastrangelo et al., 2018), while higher levels of inbreeding were observed in Barbaresca, Delle Langhe, Valle del Belice (Mastrangelo et al., 2017), Leccese (Mastrangelo et al., 2018, this study) and African sheep breeds (Dzomba et al., 2021). The low F_{ROH} values observed for almost all native Croatian sheep breeds suggest that the animals in this study are not highly related. During recombination, long ROH segments are interrupted. Long ROH segments are associated with recent inbreeding, while short ROH segments indicate ancient inbreeding (Ferenčaković et al., 2013; Curik et al., 2014). The largest inbreeding values for $F_{ROH2-4\ Mb}$ = 0.010 and $F_{ROH4-8\ Mb}$ = 0.013 were observed in Istrian Sheep, indicating the presence of IBD individuals and more ancient

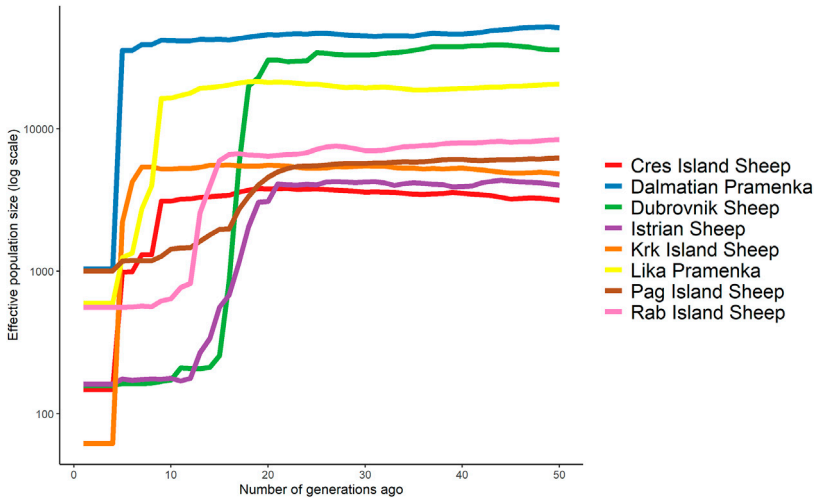


FIGURE 5 | Historical effective population size calculated with the software GONE for native Croatian sheep breeds during 50 generations.

relatedness, as suggested by Howrigan et al. (2011). In contrast, the highest inbreeding values for $F_{ROH8–16\ Mb}$ = 0.018 and $F_{ROH>16\ Mb}$ = 0.033 were observed in Dubrovnik Sheep, indicating recent inbreeding. The Dubrovnik Sheep experienced a drastic decline in the number of individuals in the 1990s. The revival of the breed, which began in 2003, included 112 individuals (Mioč et al., 2003), which represent the genetic base of the current Dubrovnik Sheep population. The highest degree of inbreeding in Dubrovnik Sheep (F_{ROH} = 0.070) is a consequence of the low number of animals available for revitalization. The results show an increase in the number of homozygous regions from the 20th generation to the third generation in all breeds. Accordingly, there is an increase in the average inbreeding coefficient from 0.005 to 0.021, respectively from the 25th to the third generation. The increase in the recent inbreeding coefficient could be due to the extensive use of few rams within herds, as suggested by Mastrangelo et al. (2017). In Croatia, as in Sicily, livestock breeders use natural mating where rams are used in herds with closely related individuals for several years or one ram is used in several herds, what leads to increased inbreeding and consequently lower variability.

Effective Population Size

Effective population size (N_e) is an important parameter used in population genetics. It is useful for monitoring genetic diversity and explaining population trends. In addition, effective population size is an indicator of the risk of genetic erosion (Tenesa et al., 2007). It has been known for many years that monitoring effective population size is an important tool for the long-term conservation of endangered populations (Notter, 1999; Gandini et al., 2004; Biscarini et al., 2015).

The current effective population size for Croatian sheep breeds estimated using the software GONE ranged from 61 to 1039 for Krk Island Sheep and Dalmatian Pramenka, respectively (**Table 1**). Higher N_e values were also observed for Pag Island Sheep (1005), Lika pramenka (598) and Rab Island Sheep (558). These four breeds (DPS, PIS, LPS, and RIS) show similar N_e values as the two Russian Gissar and Aykol (Deniskova et al., 2019). The very low effective population size identified for the Dubrovnik Sheep is similar to that observed for the Russian Kuchugur breed (N_e = 65) and the Swedish Gute sheep (N_e = 68) and is most likely the result of the Homeland War when the Dubrovnik Sheep population was drastically reduced. The software GONE identified several sudden changes in historical effective population size (**Figure 5**). The Dalmatian Pramenka and Lika Pramenka, as well as the Dubrovnik Sheep, reached a high historical effective population size and experienced a sharp decline about 25 generations ago. The lowest current effective population size was observed in Krk Island Sheep (N_e = 61).

When considering historical effective population size, the highest N_e across all generations was observed in Dalmatian Pramenka. Fifty generations ago, the lowest N_e values were 3159 (Cres Island Sheep), 4030 (Istrian Sheep) and 4814 (Krk Island Sheep) and the highest were 51537 (Dalmatian Pramenka) and 35988 (Dubrovnik Sheep). These values are higher than those determined by Kijas et al. (2012) and Ciani et al. (2014) for the European sheep breeds and Deniskova et al. (2018) for the Russian breeds. Very high values of historical effective population size for certain breeds indicate that these breeds were very widespread on the territory of Croatia in the past and confirm the historical records of a large number of sheep (1,105,078 head) bred in the area of present Dalmatia (Defilipis, 1966). Also, a very high linear trend in the effective population size of Pag Island Sheep can be attributed to the systematic activities of breeders since 1870, when the Gregge Modella Society was established for the improvement of sheep breeding (Pavlinić, 1936). The much higher values obtained for the Croatian sheep breeds compared to other European breeds could be due to the application of different software. The presence of population substructuring into herds could also affect the estimate of genomic effective population size. In the simulations, the software GONE showed an accurate estimation of effective population size up to 200

generations. Unlike other programs that are based on LD for N_e estimation, GONE very accurately reflects changes in N_e over history, even when there were substantial increases or decreases in value (Wang et al., 2016; Santiago et al., 2020). All N_e estimates based on LD are sensitive to population mixing because the presence of mixing in populations increases N_e values. In addition, migration affects N_e estimates, as the exchange of rams between flocks and the natural mating systems in sheep breeding can lead to an overestimation of N_e values.

CONCLUSION

The present study is the first detailed analysis of genomic diversity and population structure of eight native Croatian sheep breeds using high-density SNP markers. PCA, Neighbour network, Maximum likelihood trees and STRUCTURE analyses consistently revealed a common clustering pattern showing a clear separation between the island and mainland breeds. The eight Croatian breeds were generally unique. Some degree of admixture was observed in the population structure, confirming common ancestry. For all Croatian breeds except Krk Island Sheep, the estimated effective population size was above 100 - a threshold above which the breed is sustainable. Inbreeding coefficients estimated from runs of homozygosity were low to moderate. We compared the native Croatian breeds with the other Mediterranean and Pramenka breeds genotyped with high-density markers. In the global context, Croatian sheep clustered with other Pramenka type breeds and differed from West-European breeds. We observed gene flow from the mouflon population to domestic sheep and from Croatian native sheep breeds to Italian breeds. This migration is very interesting and it would be useful to determine the age of the migration as it could be confirmations of the Mediterranean migration route from the domestication center. This study contributes to a better understanding of the genetic background of Croatian native sheep breeds and provides information to support the genomic improvement of these local breeds.

DATA AVAILABILITY STATEMENT

Genotypic data of 201 animals representing eight Croatian sheep breeds are deposited and available at https://doi.org/10.5061/dryad.pg4f4qrsn.

ETHICS STATEMENT

The animal study was reviewed and approved by the Bioetic Committee for Animal Welfare and Protection of the University of Zagreb, Faculty of Agriculture. Written informed consent for participation was not obtained from the owners because Animal samples were collected within National Gene Bank procedure in Croatia.

AUTHOR CONTRIBUTIONS

ID, VC-C, and IC conceived and designed the study; ID, VC-C, JK, M-HL, SM, EC, and BL organized sample collection, extraction and genotyping; ID and MS performed the analysis with the supervision of IC, VC-C, and JAL; ID wrote the manuscript with the supervision of VC-C and IC. All authors read and approved the final manuscript.

FUNDING

This study was founded by Croatian Science Foundation project IP-2018-01-8708 Application of NGS methods in the assessment of genomic variability in ruminants—ANAGRAMS and the National Natural Science Foundation of China (Nos. 31661143014, 31825024, 31972527, 31660651, and 31760661).

ACKNOWLEDGMENTS

This study was supported by a project Sir je IN: KK.01.1.1.04.0058. We thank to the National Gene Bank, Ministry of Agricultre of Croatia for providing some of the samples. The authors would like to thank the University of Zagreb University Computing Centre (SRCE)—Danijel Vrčić, Emir Imamagić and Tomislav Stilinović for providing computational facilities and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.940736/full#supplementary-material

REFERENCES

Al-Mamun, H. A., Clark, S. A., Kwan, P., and Gondro, C. (2015). Genome-Wide Linkage Disequilibrium and Genetic Diversity in Five Populations of Australian Domestic Sheep. *Genet. Sel. Evol.* 47, 90. doi:10.1186/s12711-015-0169-6

Andersson, L. (2012). How Selective Sweeps in Domestic Animals Provide New Insight into Biological Mechanisms. *J. Intern. Med.* 271, 1–14. doi:10.1111/j.1365-2796.2011.02450.x

Barbato, M., Hailer, F., Orozco-terWengel, P., Kijas, J., Mereu, P., Cabras, P., et al. (2017). Genomic Signatures of Adaptive Introgression from European Mouflon into Domestic Sheep. *Sci. Rep.* 7, 7623. doi:10.1038/s41598-017-07382-7

Barker, J. S. (2001). Conservation and Management of Genetic Diversity: A Domestic Animal Perspective. *Can. J. For. Res.* 31, 588–595. doi:10.1139/x00-180

Barendse, W., Harrison, B. E., Bunch, R. J., Thomas, M. B., and Turner, L. B. (2009). Genome Wide Signatures of Positive Selection: The Comparison of Independent Samples and the Identification of Regions Associated to Traits. *BMC Genomics* 10, 178. doi:10.1186/1471-2164-10-178

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

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, 33. doi:10.3389/fgene.2015.00033

Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., et al. (2020). Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. *Mol. Biol. Evol.* 38, 838–855. doi:10.1093/molbev/msaa236

Ciani, E., Crepaldi, P., Nicoloso, L., Lasagna, E., Sarti, F. M., Moioli, B., et al. (2014). Genome-wide Analysis of Italian Sheep Diversity Reveals a Strong Geographic Pattern and Cryptic Relationships between Breeds. *Anim. Genet.* 45, 256–266. doi:10.1111/age.12106

Ciani, E., Mastrangelo, S., Mastrangelo, S., Da Silva, A., Marroni, F., Ferenčaković, M., et al. (2020). On the Origin of European Sheep as Revealed by the Diversity of the Balkan Breeds and by Optimizing Population-Genetic Analysis Tools. *Genet. Sel. Evol.* 52, 25. doi:10.1186/s12711-020-00545-7

Ćinkulov, M., Popovski, Z., Porcu, K., Tanaskovska, B., Hodžić, A., Bytyqi, H., et al. (2007). "Genetic Diversity of the West Balkan Pramenka Sheep," in *Book of Abstract of the Conference on Native Breeds and Varieties as Part of Natural and Cultural Heritage* (Šibenik, Croatia: State Institute for Nature Protection), 65–66.

Cubric-Curik, V., Feligni, M., Lukac-Havranek, J., Curik, I., and Enne, G. (2002). Genetic Polymorphism of β-Lactoglobulin in Native Sheep from the Island of Pag. *Food Techol. Biotechnol.* 40, 75–78.

Cubric-Curik, V., Feligini, M., Ferencakovic, M., Dzidic, A., Salajpal, K., Ambriovic-Ristov, A., et al. (2009). Sequence Polymorphism of PrP Exon 3 Gene in Istrian and Crossbred Sheep. *Italian J. Animal Sci.* 8, 86–88. doi:10.4081/ijas.2009.s3.86

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and Runs of Homozygosity: A Possible Solution to an Old Problem. *Livest. Sci.* 166, 26–34. doi:10.1016/j.livsci.2014.05.034

Defilipis, J. (1966). *Neki Aspekti Regionalne Razvijenosti I Struktura Poljoprivrede Dalmacije S Posebnim Osvrtom Na Ovčarstvo*. Master Thesis. Zemun, Jugoslavija: Faculty of Agriculture.

Deng, J., Xie, X.-L., Wang, D.-F., Zhao, C., Lv, F.-H., Li, X., et al. (2020). Paternal Origins and Migratory Episodes of Domestic Sheep. *Curr. Biol.* 30, 4085–4095. doi:10.1016/j.cub.2020.07.077

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

Deniskova, T., Dotsev, A., Lushihina, E., Shakhin, A., Kunz, E., Medugorac, I., et al. (2019). Population Structure and Genetic Diversity of Sheep Breeds in the Kyrgyzstan. *Front. Genet.* 10, 1311. doi:10.3389/fgene.2019.01311

Dzomba, E. F., Chimonyo, M., Pierneef, R., and Muchadeyi, F. C. (2021). Runs of Homozygosity Analysis of South African Sheep Breeds from Various Production Systems Investigated Using OvineSNP50k Data. *BMC Genomics* 22, 7. doi:10.1186/s12864-020-07314-2

Edea, Z., Dessie, T., Dadi, H., Do, K.-T., and Kim, K.-S. (2017). Genetic Diversity and Population Structure of Ethiopian Sheep Populations Revealed by High-Density SNP Markers. *Front. Genet.* 8, 218. doi:10.3389/fgene.2017.00218

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Feligini, M., Vlaco, S., Curik, V. C., Parma, P., Greppi, G., and Enne, G. (2005). A Single Nucleotide Polymorphism in the Sheep κ-casein Coding Region. *J. Dairy Res.* 72, 317–321. doi:10.1017/S0022029905000932

Ferenčaković, M., Sölkner, J., and Curik, I. (2013). Estimating Autozygosity from High-Throughput Information: Effects of SNP Density and Genotyping Errors. *Genet. Sel. Evol.* 45, 42. doi:10.1186/1297-9686-45-42

Ferencakovic, M., Curik, I., Pérez-Pardal, L., Royo, L. J., Cubric-Curik, V., Fernández, I., et al. (2012). Mitochondrial DNA and Y-Chromosome Diversity in East Adriatic Sheep. *Anim. Genet.* 44, 184–192. doi:10.1111/j.1365-2052.2012.02393.x

Frankham, R., Briscoe, D. A., and Ballou, J. D. (2002). *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Gandini, G. C., and Villa, E. (2003). Analysis of the Cultural Value of Local Livestock Breeds: A Methodology. *J. Anim. Breed. Genet.* 120, 1–11. doi:10.1046/j.1439-0388.2003.00365.x

Gandini, G. C., Ollivier, L., Danell, B., Distl, O., Georgoudis, A., Groeneveld, E., et al. (2004). Criteria to Assess the Degree of Endangerment of Livestock Breeds in Europe. *Livest. Prod. Sci.* 91, 173–182. doi:10.1016/j.livprodsci.2004.08.001

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic Diversity in Farm Animals - A Review. *Anim. Genet.* 41, 6–31. doi:10.1111/j.1365-2052.2010.02038.x

Guerrini, M., Forcina, G., Panayides, P., Lorenzini, R., Garel, M., Anayiotos, P., et al. (2015). Molecular DNA Identity of the Mouflon of Cyprus (*Ovis orientalis* Ophion, Bovidae): Near Eastern Origin and Divergence from Western Mediterranean Conspecific Populations. *Syst. Biodivers.* 13, 472–483. doi:10.1080/14772000.2015.1046409

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting Autozygosity through Runs of Homozygosity: A Comparison of Three Autozygosity Detection Algorithms. *BMC Genomics* 12, 460. doi:10.1186/1471-2164-12-460

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

Ivanković, A., Dovč, P., Kavar, T., Caput, P., Mioč, B., Pavić, V., et al. (2005). Genetic Characterisation of the Pag Island Sheep Breed Based on Microsatellite and mtDNA Data. *Small Ruminant Res.* 57, 167–174. doi:10.1016/j.smallrumres.2004.07.002

Kasap, A., Ramljak, J., and Špehar, M. (2021). Estimation of Population-specific Genetic Parameters Important for Long-Term Optimum Contribution Selection-Case Study on a Dairy Istrian Sheep Breed. *Animals* 11, 2356–2415. doi:10.3390/ani11082356

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, e1001258. doi:10.1371/journal.pbio.1001258

Larson, G., and Burger, J. (2013). A Population Genetics View of Animal Domestication. *Trends Genet.* 29, 197–205. doi:10.1016/j.tig.2013.01.003

Lawson Handley, L.-J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M. W., et al. (2007). Genetic Structure of European Sheep Breeds. *Heredity* 99, 620–631. doi:10.1038/sj.hdy.6801039

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., et al. (2007). Runs of Homozygosity Reveal Highly Penetrant Recessive Loci in Schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19942–19947. doi:10.1073/pnas.0710021104

Li, Y.-L., and Liu, J.-X. (2018). StructureSelector : A Web-Based Software to Select and Visualize the Optimal Number of Clusters Using Multiple Methods. *Mol. Ecol. Resour.* 18, 176–177. doi:10.1111/1755-0998.12719

Lv, F.-H., Cao, Y.-H., Liu, G.-J., Luo, L.-Y., Lu, R., Liu, M.-J., et al. (2021). Whole-Genome Resequencing of Worldwide Wild and Domestic Sheep Elucidates Genetic Diversity, Introgression, and Agronomically Important Loci. *Mol. Biol. Evol.* 39, msab353. [Epub Online ahead of print]. doi:10.1093/molbev/msab353

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

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., Portolano, B., et al. (2018). Runs of Homozygosity Reveal Genome-wide Autozygosity in Italian Sheep Breeds. *Anim. Genet.* 49, 71–81. doi:10.1111/age.12634

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. doi:10.1016/j.ajhg.2008.08.007

Mioč, B., Ivanković, A., Pavić, V., Barač, Z., Sinković, K., and Marić, I. (2003). Odlike Eksterijera I Polimorfizmi Proteina Krvi Dubrovačke Ovce. *Stočarstvo* 57, 3–11.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge: MIT Press.

Nei, M. (1972). Genetic Distance Between Populations. *Am. Nat.* 106, 283–292. doi:10.1086/282771

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genet* 89, 583–590. doi:10.1093/genetics/89.3.583

Notter, D. R. (1999). The Importance of Genetic Diversity in Livestock Populations of the Future. *J. Anim. Sci.* 77, 61–69. doi:10.2527/1999.77161x

Pavlinić, P. (1936). *Paška Ovca*, 6. Zagreb: Veterinarski Arhiv.

Pembleton, L. W., Cogan, N. O. I., and Forster, J. W. (2013). St AMPP: An R Package for Calculation of Genetic Differentiation and Structure of Mixed-

ploidy Level Populations. *Mol. Ecol. Resour.* 13, 946–952. doi:10.1111/1755-0998.12129

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G., Erhardt, G., et al. (2007). Genetic Diversity and Subdivision of 57 European and Middle-Eastern Sheep Breeds. *Anim. Genet.* 38, 37–44. doi:10.1111/j.1365-2052.2007.01561.x

Pickrell, J. K., and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-wide Allele Frequency Data. *PLoS Genet.* 8, e1002967. doi:10.1371/journal.pgen.1002967

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genet* 155, 945–959. doi:10.1093/genetics/155.2.945

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

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

Rochus, C. M., Jonas, E., and Johansson, A. M. (2020). Population Structure of Five Native Sheep Breeds of Sweden Estimated with High Density SNP Genotypes. *BMC Genet.* 21, 27. doi:10.1186/s12863-020-0827-8

Salamon, D., Gutierrez-Gil, B., Arranz, J. J., Barreta, J., Batinic, V., and Dzidic, A. (2014). Genetic Diversity and Differentiation of 12 Eastern Adriatic and Western Dinaric Native Sheep Breeds Using Microsatellites. *Animal* 8, 200–207. doi:10.1017/S1751731113002243

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., and Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol. Biol. Evol.* 37, 3642–3653. doi:10.1093/molbev/msaa169

Shihabi, M., Lukic, B., Cubric-Curik, V., Brajkovic, V., Oršanić, M., Ugarković, D., et al. (2022). Identification of Selection Signals on the X-Chromosome in East Adriatic Sheep: A New Complementary Approach. *Front. Genet.* 13, 887582. doi:10.3389/fgene.2022.887582

Špehar, M., Ramljak, J., and Kasap, A. (2022). Estimation of Genetic Parameters and the Effect of Inbreeding on Dairy Traits in Istrian Sheep. *Italian J. Animal Sci.* 21, 331–342. doi:10.1080/1828051X.2022.2031320

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, 275–284. doi:10.1111/j.1365-294X.2007.03475.x

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent Human Effective Population Size Estimated from Linkage Disequilibrium. *Genome Res.* 17, 520–526. doi:10.1101/gr.6023607

Wang, J., Santiago, E., and Caballero, A. (2016). Prediction and Estimation of Effective Population Size. *Heredity* 117, 193–206. doi:10.1038/hdy.2016.43

Zeder, M. A. (2008). Domestication and Early Agriculture in the Mediterranean Basin: Origins, Diffusion, and Impact. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11597–11604. doi:10.1073/pnas.0801317105

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* 28, 3326–3328. doi:10.1093/bioinformatics/bts606

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 | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 August 2022
DOI 10.3389/fgene.2022.886781

OPEN ACCESS

EDITED BY
Klaus Wimmers,
Leibniz Institute for Farm Animal Biology (FBN), Germany

REVIEWED BY
Abiodun Joseph Fatoba,
University of KwaZulu-Natal, South Africa
Mahmoud Kandeel,
Kafrelsheikh University, Egypt
Xiaokai Song,
Nanjing Agricultural University, China

*CORRESPONDENCE
Jun Heon Lee,
junheon@cnu.ac.kr

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

RECEIVED 28 February 2022
ACCEPTED 12 July 2022
PUBLISHED 08 August 2022

CITATION
Kim M, Chung Y, Manjula P, Seo D, Cho S, Cho E, Ediriweera TK, Yu M, Nam S and Lee JH (2022), Time-series transcriptome analysis identified differentially expressed genes in broiler chicken infected with mixed Eimeria species.
Front. Genet. 13:886781.
doi: 10.3389/fgene.2022.886781

Time-series transcriptome analysis identified differentially expressed genes in broiler chicken infected with mixed *Eimeria* species

Minjun Kim[1], Yoonji Chung[1], Prabuddha Manjula[2], Dongwon Seo[3,4], Sunghyun Cho[1], Eunjin Cho[4], Thisarani Kalhari Ediriweera[4], Myunghwan Yu[1], Sunju Nam[1] and Jun Heon Lee[1,4]*

[1]Division of Animal and Dairy Science, Chungnam National University, Daejeon, Korea, [2]Department of Animal Science, Uva Wellassa University, Badulla, Sri Lanka, [3]Research Institute TNT Research Company, Jeonju, Korea, [4]Department of Bio AI Convergence, Chungnam National University, Daejeon, Korea

Coccidiosis caused by the *Eimeria* species is a highly problematic disease in the chicken industry. Here, we used RNA sequencing to observe the time-dependent host responses of *Eimeria*-infected chickens to examine the genes and biological functions associated with immunity to the parasite. Transcriptome analysis was performed at three time points: 4, 7, and 21 days post-infection (dpi). Based on the changes in gene expression patterns, we defined three groups of genes that showed differential expression. This enabled us to capture evidence of endoplasmic reticulum stress at the initial stage of *Eimeria* infection. Furthermore, we found that innate immune responses against the parasite were activated at the first exposure; they then showed gradual normalization. Although the cytokine-cytokine receptor interaction pathway was significantly operative at 4 dpi, its downregulation led to an anti-inflammatory effect. Additionally, the construction of gene co-expression networks enabled identification of immunoregulation hub genes and critical pattern recognition receptors after *Eimeria* infection. Our results provide a detailed understanding of the host-pathogen interaction between chicken and *Eimeria*. The clusters of genes defined in this study can be utilized to improve chickens for coccidiosis control.

KEYWORDS

chicken, host-pathogen interaction, innate immunity, anti-inflammation, transcriptome analysis, gene co-expression network, Eimeria

Introduction

Chicken coccidiosis is a disease caused by protozoan parasites of the genus *Eimeria* (Blake and Tomley, 2014). These parasites invade the small or large intestine of chickens and cause tissue damage, leading to malnutrition, diarrhea, bloody stools, and (in severe cases) death (Lillehoj et al., 2007). Chicken coccidiosis is thus an important livestock disease that causes great economic loss to the poultry industry; anticoccidial drugs and vaccines are continuously used for prevention (Chapman et al., 2010). The establishment of a coccidiosis-resistant chicken variety would be an effective strategy for overcoming coccidiosis because it is consistent with the recent trend of reducing antibiotics; it would also reduce the cost of disease prevention (Jeffers, 1981). To achieve this goal, there is a need to identify physiological functions in chickens that are important under coccidial conditions. *Eimeria* infection promotes cell-mediated immunity in chickens; several immune cells (e.g., T lymphocytes and macrophages) have major roles in such immunity (Lillehoj and Trout, 1996). Previous studies have used a quantitative reverse transcription polymerase chain reaction (qRT-PCR) to identify differentially expressed genes (DEGs) during *Eimeria* infection by observing the immune response or using RNA sequencing (RNA-Seq) technology (Laurent et al., 2001; Wang et al., 2019).

The host response to coccidiosis parasites appears to differ over time (Bremner et al., 2021; Choi et al., 2021; Yu et al., 2021). For a more detailed understanding of the host response, there is a need to understand the biological functions performed in the body at different stages of infection. In this study, chronological DEGs in the cecum of chickens infected with multiple *Eimeria* species were identified via RNA-Seq. Clustering analysis was performed and gene co-expression networks were constructed to observe the main functions using enrichment analyses.

Materials and methods

Sample preparation and RNA-Seq

Thirty-nine 1-day-old (Indian River; Aviagen™) male broiler chickens were randomly allocated to 13 oocyst-free cages. After 14 days, 21 chickens (PC) were orally administrated with 1 ml of Livacox® T™ (Biopharm Co., Prague, Czech Republic) as positive control chickens (PC). A 10-fold dosage containing 3–5 × 10^3 active oocysts of each of *Eimeria acervulina*, *Eimeria tenella*, and *Eimeria maxima* were used for infection. Eighteen non-challenged negative control (NC) birds were equally inoculated with distilled water to simulate inoculation stress. All PC and NC chickens were humanely euthanized at 4, 7, and 21 days post-infection (dpi), respectively. Cecum tissue samples of all chickens were collected in tubes with RNA-later and stored at −80°C for RNA extraction. Finally, 3 samples of NC and 5 samples of PC were randomly selected for sequencing at each time point. The overall experimental design is shown in Figure 1.

Total RNA was isolated from tissue samples using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) and processed with an RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany). Total RNA integrity was determined using a 2,100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States). All samples with RNA integrity numbers ≥ 7 were used for cDNA library construction, which was performed using a TruSeq stranded mRNA sample preparation kit (Illumina, San Diego, CA, United States) in accordance with the manufacturer's guidelines. After library construction, all samples were sequenced using the Illumina NovaSeq 6,000 platform and 101 bp paired-end reads were generated.

Sequencing data processing and quantification of differentially expressed genes

Raw sequence read quality control was performed using the FastQC software v0.11.9. Sequencing adaptors and low-quality reads were trimmed using Cutadapt software v1.15 and Trimmomatic software v0.39. Before analysis, pre-processed reads were re-checked with FastQC, then mapped to the chicken reference genome (GRCg6a, GCA_000002315.5) using STAR software v2.7.7a. The reference genome was downloaded from the Ensemble genome browser FTP site (ftp.ensembl.org/pub/release-104/fasta/gallus_gallus/dna/); the index file of the reference genome was built using bowtie2 software v2.4.2 and samtools software v1.11. STAR software was used to count reads matching the genes, based on the exons in *Gallus gallus* GRCg6a v102 GTF (Ensembl) as a genomic annotation reference file. The EdgeR package v3.30.3 in Bioconductor software was used to quantify the reads mapped to each gene. Genes with a total read count ≤ 8 for all genes were excluded to avoid statistical bias in the identification of DEGs. Read count normalization was conducted with the trimmed mean of M-value (TMM) method. Differential expression analysis of two treatments (PC versus NC) at three time points (4, 7, and 21 dpi) was performed; the resulting *p*-values were corrected using the Benjamini-Hochberg procedure. The DEGs were determined with a level of absolute $\log_2$ fold-change (FC) ≥ 1 on an adjusted false discovery rate (FDR) corrected *p*-value of < 0.05.

Gene clustering analysis and gene co-expression network construction

After DEGs had been identified, their expression changes over time were analyzed via gene clustering. Genes were classified with the k-means clustering algorithm using each normalized TMM count. The k value that best distinguished the features of the gene expression

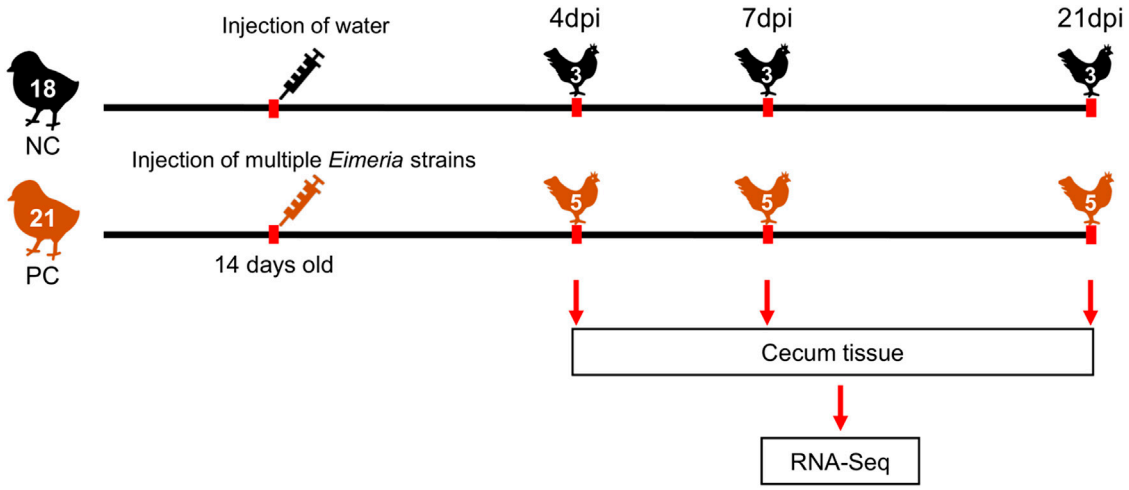


FIGURE 1
Overview of the experimental design. NC, Negative control; PC, Positive control; dpi, days post-infection.

pattern was used for subsequent analysis. MeV v4.9.0 software was used for the analysis and visualization of the results.

GCNs were established to enable analysis of the functional associations of genes. DEGs identified at least once at each time point were collected; instances of significant co-expression between genes were determined using the partial correlation coefficient with information theory (PCIT) algorithm. Associations with an absolute co-expression correlation ≥0.90 between DEGs were used to construct GCNs. After construction, GCNs containing at least 10 genes were considered meaningful. Cytoscape v3.8.2 software was used for network visualization.

Functional enrichment analysis

Annotations of gene ontology (GO) terms and a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were performed for the enrichment analysis of genes clustered in gene groups. The biological process (BP) and molecular function (MF) databases were used for GO analysis. ClusterProfiler (Yu et al., 2012) R packages in Bioconductor were used for GO and KEGG analysis, and with a significant cut-off of $p < 0.05$. All annotation procedures were performed in *Gallus gallus*.

To identify significantly different enrichment terms between the PC and NC treatments at each time point, gene set enrichment analysis (GSEA) was performed via GSEA v4.1.0 software (Subramanian et al., 2005; Luo and Brouwer, 2013) using $\log_2$-normalized TMM counts of genes at each time point. Gene sets in the BP, cellular component, MF, and KEGG databases were used for the analysis; significant pathways (FDR-corrected $p < 0.05$) in the gene sets were listed as results.

Validation of RNA-Seq analysis via qRT-PCR

RNA isolation and cDNA synthesis were performed using cecum tissue samples from each individual of the treatment at each of the three time points. A Qiagen RNeasy kit was used for isolation; mRNA was converted into cDNA in the presence of Oligo d(T) primer (SuperScript IV, Invitrogen). Nine DEGs from each expression type were selected for validation; primers for qRT-PCR were designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/-tools/primer-blast/) and are listed (Supplementary Table S1 of Supplementary Material S1). qRT-PCR was carried out on a CFX Connect Real-Time System (Bio-Rad, Hercules, CA, United States) using SYBR green master mix (GeNetBio, Daejeon, Korea), in accordance with the manufacturer's instructions. Relative gene expression was calculated using the delta-delta-Ct method. Finally, correlations of gene expression between the qRT-PCR and RNA-Seq analyses were estimated using R^2 value.

Results

Gene expression analysis

In total, 983 million paired-end reads were generated using cecum tissue samples from 24 chickens. Read data were processed

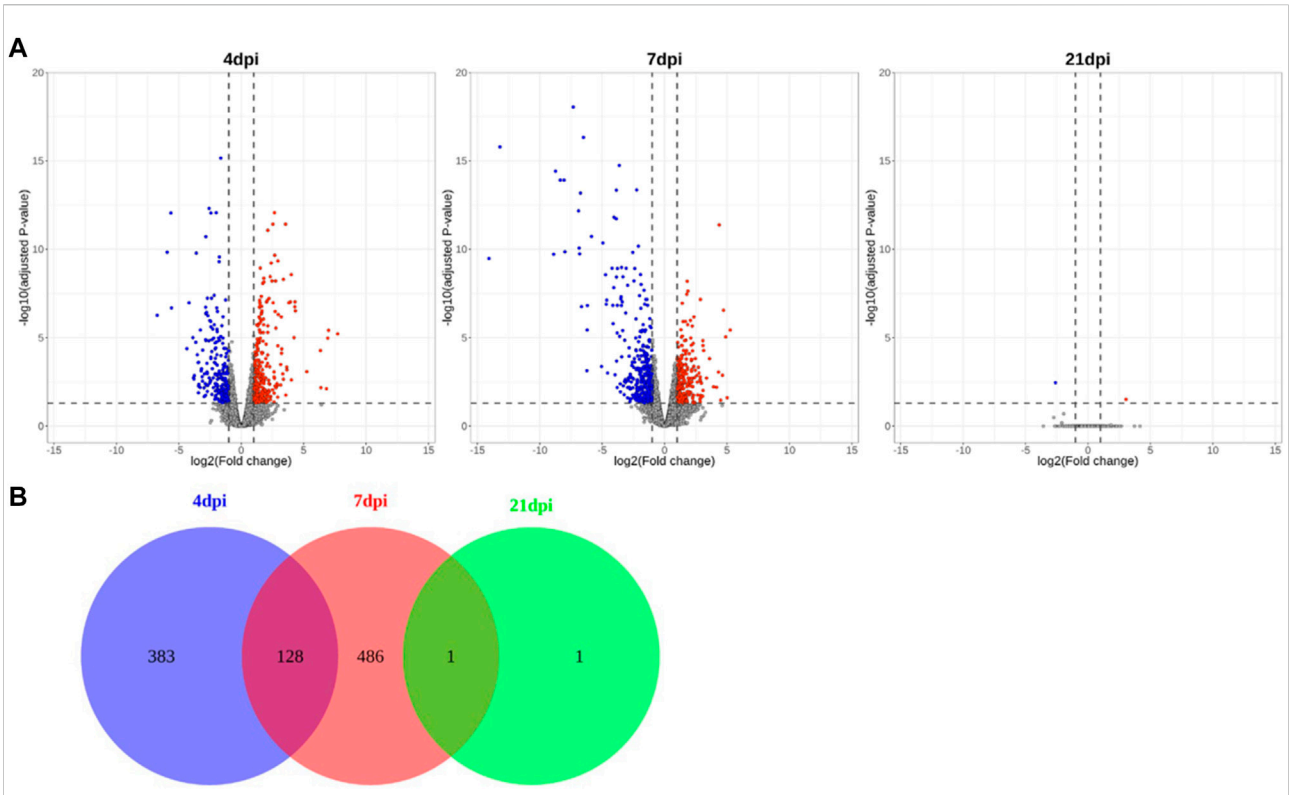


FIGURE 2
Volcano plots and Venn diagram showing significant DEGs obtained via time-series comparison between PC and NC treatments. **(A)** Volcano plots showing different time points, **(B)** Venn diagram indicating numbers of DEGs.

(processing overview is shown in Supplementary Materials S2) and mapped to the chicken reference genome (GRCg6a). Based on the calculated FC, 511 and 615 significant DEGs were detected at 4 and 7 dpi, respectively, whereas only 2 DEGs were found at 21 dpi; these results indicated a stark contrast, as demonstrated in volcano plots and a Venn diagram (Figure 2). Over time, the trends in gene regulation changed. At 4 dpi, 53% of the DEGs were upregulated, while this value was only 41% at 7 dpi. The number of DEGs indicates that the difference in gene expression between the NC and PC treatments was small at 21 dpi. Analysis of genes from all time points together revealed 999 unique DEGs in *Eimeria*-infected chickens.

Clustering genes and gene co-expression network analysis

To analyze the biological regulation of the host response to *Eimeria* infection, DEGs were classified based on their expression patterns at each of the three time points. Overall, 999 genes were clustered into three groups, which comprised 436, 414, and 149 genes, respectively. The groups were referred to as Types 1 to 3 because of their distinct expression patterns, as shown in Figure 3A. The mean log_2FC values of Type I genes over time were 0.87, −1.87, and −0.07, respectively; thus, expression level tended to sharply decrease after an initially high level in the initial stage of infection. In contrast, Type 2 genes were not active in the early stage of infection, then rebounded at 7 dpi and gradually normalized; the mean log_2FC values over time were -1.13, 1.06, and 0.08, respectively. The mean log_2FC values of Type 3 genes were 1.65, 0.65, and -0.31, respectively; these values tended to be high at 4 dpi, then decreased over time. Notably, 106 Type 1 genes were significantly upregulated at 4 dpi, then significantly downregulated at 7 dpi. Among Type 2 genes, 66 exhibited inverted regulation from 4 to 7 dpi.

For additional clustering, GCNs were constructed using 371 genes with 2,656 connections after data had been passed through the PCIT algorithm. Three meaningful networks were formed; they were named GCN278, GCN53, and GCN40, based on the number of genes included in each network. Co-expression findings did not match the expression pattern clusters, as shown in Figure 3B. GCN278 comprised 111 Type 1 genes, 126 Type 2 genes, and 41 Type 3 genes. GCN53 comprised 22 Type 1 genes, 28 Type 2 genes, and three Type 3 genes (i.e., only a small number of Type 3 genes). GCN40 consisted of 21 Type 1 genes, 13 Type 2 genes, and six Type 3 genes; thus, it was characterized by a high proportion of Type 1 genes. All expression values, expression type group, and GCNs assigned to each DEGs are summarized in Supplementary Materials S3.

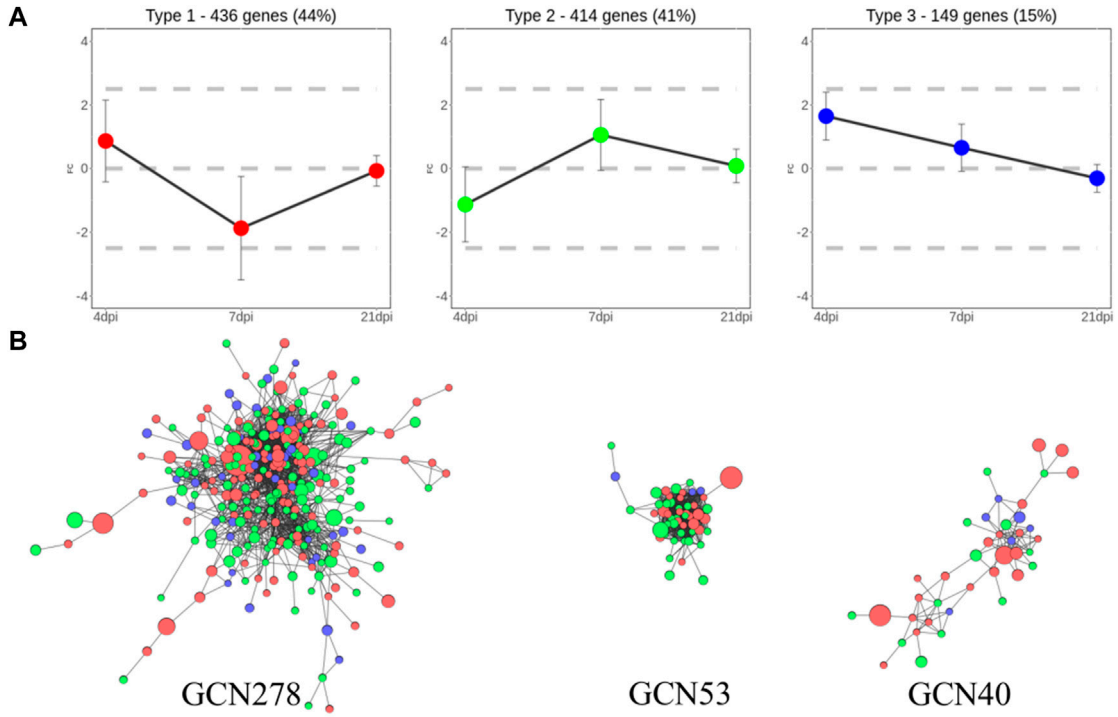


FIGURE 3
Clustered DEGs using two different grouping methods **(A)** Three types of clusters separated based on their gene expression patterns. In total, 436, 414, and 149 genes were assigned to Types 1, 2, and 3, respectively. **(B)** GCNs clustered using gene co-expression values. Each node is a gene colored with its expression type. Node size reflects absolute $\log_2$FC. Edges between nodes represent significant co-expression of genes. Network names reflect the number of genes they contain.

Functional enrichment analyses

GO and KEGG enrichment analyses were performed on the three types of expression pattern clusters and three GCNs. The results of the analyses are presented in Figures 4, 5, and Supplementary Materials S4. The top eight BP terms for the GO analysis of Type 1 genes involved lipid production and utilization processes; other functions were also observed (e.g., organic anion transport, homeostatic process, regulation of biological quality, and antibiotic metabolic process). Among Type 3 genes, BP terms related to the innate immune response (e.g., regulation of defense response, cytokine secretion, and inflammatory response) exhibited high significance; tissue regeneration and circulatory system development terms were also observed. No significant ($p < 0.05$) terms were found among Type 2 genes (Figure 4D,ED).

Characteristic results of KEGG analysis were identified in expression pattern clusters. Type 1 terms included steroid biosynthesis and cytokine-cytokine receptor interaction, while Type 2 terms included ABC transporters, nitrogen metabolism, and intestinal immune network for IgA production. Type 3 terms included nutrient metabolism, such as starch and sucrose metabolism, and phagosome and extracellular matrix–receptor interaction functions (Figure 4A–C).

The results of GCN functional analysis were as follows. The BP terms for GCN278 were cellular homeostasis, antibiotic metabolic process, and epithelium development; the MF terms were signaling receptor activator activity, hormone activity, and oxidoreductase activity. The top BP terms for GCN53 were innate immune response-related functions; nucleotide binding-related terms were identified in the MF database. In the GO analysis of GCN40, positive regulation of immune system process was the most significant BP term, together with leukocyte migration and cell-cell adhesion. Among the MF terms, functions related to protein tyrosine kinase activity were identified (Figure 5D–FD).

KEGG analysis of GCNs revealed distinct biological functions. In GCN278, there were oxidative phosphorylation and PPAR signaling pathway terms; in GCN53, influenza A and several types of receptor signaling pathways were enriched. Finally, in GCN40, the cytokine-cytokine receptor interaction, lipid metabolism, and intestinal immune network for IgA production terms were identified (Figure 5A–C).

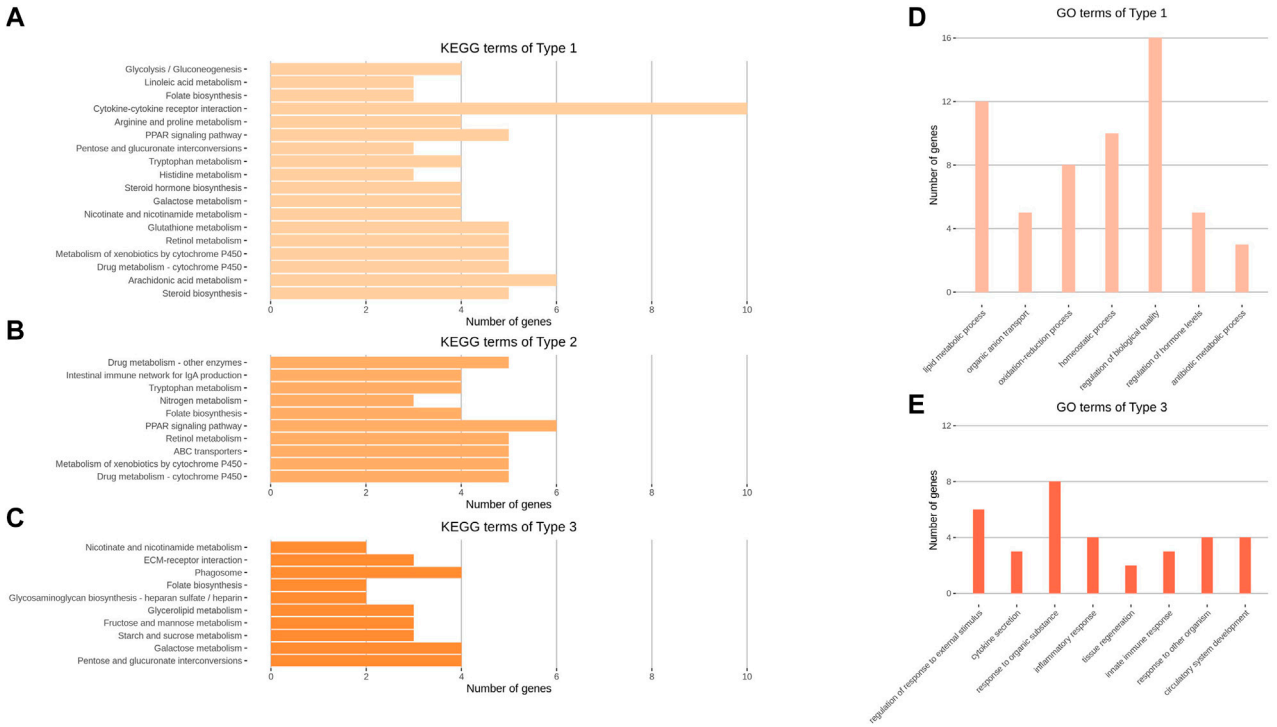


FIGURE 4
KEGG and GO analysis results of Type 1, 2, and 3 gene clusters. **(A–C)** Enriched KEGG pathways of Type 1, 2, and 3 genes clustered using the k-means clustering method. **(D–E)** Enriched GO terms of the biological process category of Types 1 and 3 gene clusters. Representative GO terms were selected from 53 to 188 significant terms of Types 1 and 3, respectively. GO analysis of Type 2 genes did not show any significant terms. **(A–E)** All enrichment terms passed a significance cut-off ($p < 0.05$) and were sorted in the order of largest significance levels.

Gene set enrichment analysis results associated with the time-series host response to *Eimeria* infection

GSEA was performed on the GO and KEGG databases to focus on the functions performed over time during *Eimeria* infection (Supplementary Material S5). At 21 dpi, genes with a small differential expression level were excluded from the analysis. Among the KEGG gene sets upregulated at 4 dpi after *Eimeria* infection, the most enriched term was steroid biosynthesis, consistent with the KEGG analysis results for Type 1 DEGs. There were 12 core enrichment genes in steroid biosynthesis; seven were DEGs. In addition, inflammatory response-related signaling pathways were upregulated terms. In terms of the normalized enrichment score, six of the top 10 GO terms were related to lipid or sterol synthesis and metabolic functions. In addition, the keratinization function had a high significance, such that 19 core enriched genes (including four DEGs) constituted the gene set.

Nutrient metabolic function terms such as glutathione, tyrosine, nitrogen, and phenylalanine were found in the KEGG terms upregulated at 7 dpi, consistent with the KEGG analysis results for Type 2 and Type 3 clusters. In addition, epithelial cell signaling in the *Helicobacter pylori* infection KEGG term was upregulated at 7 dpi, with 12 core enriched genes containing nine DEGs. GO analysis revealed many oxidative phosphorylation-related terms with a high normalized enrichment score; among the terms belonging to BP, the phagosome acidification term contained 10 core enrichment genes (including six DEGs) in the gene set. GSEA confirmed the functions downregulated during *Eimeria* infection. KEGG analysis at 7 dpi showed that the natural killer cell-mediated cytotoxicity pathway had 40 core genes, including three DEGs. GO analysis identified the phosphotyrosine residue binding term in the MF database, with 22 core-enriched genes including two DEGs.

qRT-PCR

To validate the differential expression indicated in RNA-Seq analysis, qRT-PCR was conducted using nine genes that were selected according to their gene expression types. The

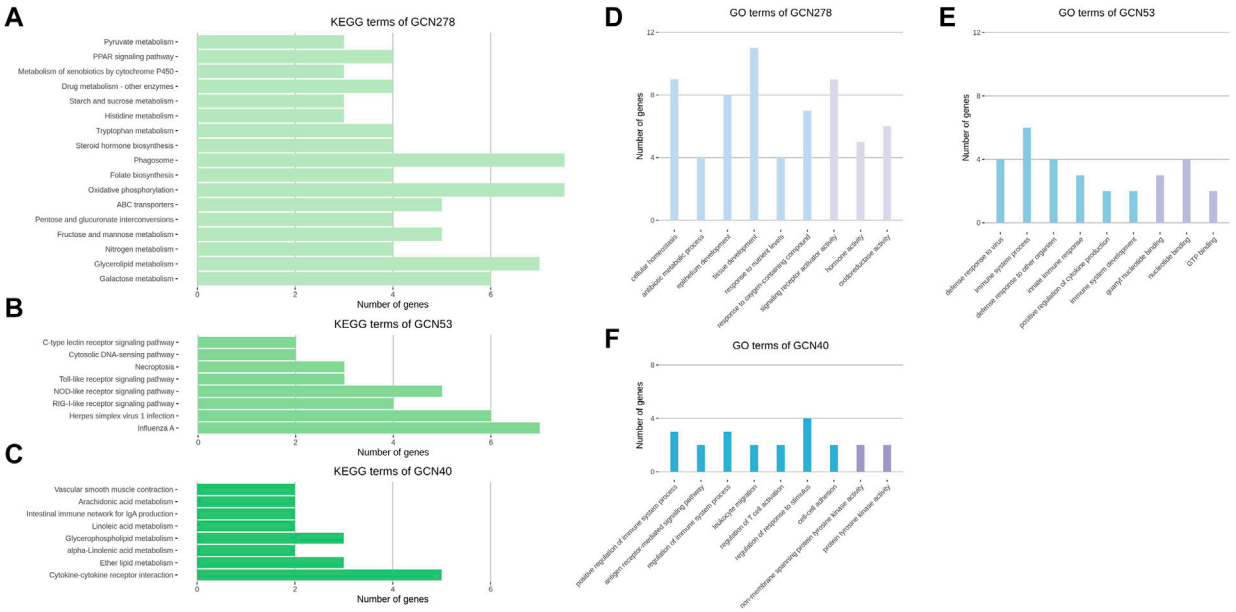


FIGURE 5
KEGG and GO analysis results of three GCNs. **(A–C)** Enriched KEGG pathways of the GCN278, GCN53, and GCN40 networks. **(D–F)** Enriched GO terms of the GCN278, GCN53, and GCN40 networks. GO terms were categorized as a biological process and molecular functions (colored blue and purple, respectively). Representative GO terms were selected from 26, 42, and 70 significant biological process terms and 12, 15, and 6 significant molecular functions terms for GCN278, GCN53, and GCN40, respectively. **(A–F)** All enrichment terms passed a significance cut-off ($p < 0.05$) and were sorted in the order of largest significance levels.

target genes were *FABP2* (fatty acid binding protein 2), *IFI6* (interferon alpha inducible protein 6), *INSIG1* (insulin induced gene 1), *KRT40* (keratin 40), *IL13RA2* (interleukin 13 receptor subunit alpha 2), *ART7B* (GPI-anchored ADP-ribosyltransferase), *IRF9* (interferon regulatory factor 9), *CCKAR* (cholecystokinin A receptor), and *XKR9* (XK related 9). Relative quantification of gene expression was performed with *ACTB* (actin, beta) and *YWHAZ* (tyrosine 3-monooxygenase-/tryptophan 5-monooxygenase activation protein zeta) as control genes. Experimental samples were selected among the same treatment accordant with those used in RNA-Seq analysis. Throughout the three time points, expression values were calculated (Supplementary Table S2 in Supplementary Material S1) and the mean correlation of the expression patterns of the genes between RNA-Seq and qRT-PCR reached 0.811 ($R^2 = 0.658$). Therefore, statistical analysis showed good correspondence between the qRT-PCR and RNA-Seq results.

Discussion

Coccidiosis is a critical parasitic disease in the chicken industry. To overcome the situation, recent studies indicated that chages from traditional methods to develop more effective anticoccidial vaccine (Venkatas and Adeleke, 2019). As another approach to the promising strategies to prevent the disease, this study aimed to characterize the host immune response that occurs in chickens under coccidiosis infection.

Blended effects on gene expression of cecum by multiple *Eimeria* species infection

Eimeria species is known as habitat-specific (Li et al., 2020). Accordingly, *Eimeria acervulina*, *Eimeria tenella*, and *Eimeria maxima* inhabit basically in duodenum, cecum, and jejunum respectively (Raman et al., 2011; Quiroz-Castañeda and Dantán-González, 2015). As the mRNA derived from the cecum tissue has been studied here, most of the effects on the identified responses are attributed to *E. tenella*. However, Li et al. (2019) observed that *E. maxima*-infection also affected the differential gene expression in the cecum tissue of chicken. Nonetheless, DEGs and the corresponding functions tended to be similar to those of chickens infected with *E. tenella*. Further, in the cecum, the transcriptome levels of genes responsive to *E. tenella* is known to be higher (Li et al., 2019). Hence, it manifests that the major effects described in this study are due to *E. tenella* and while some minor effects are shared with the other two strains.

Changes in number of differentially expressed genes over time

Whole serial transcriptomes for three time points (4, 7, and 21 dpi) after mixed strains of *Eimeria* infection were compared and integrated. The number of DEGs increased from 4 to 7 dpi, but DEGs almost disappeared at 21 dpi. According to Morris et al. (2007) and Long et al. (1981), coccidiosis-damaged cecal mucosa expresses inflammatory symptoms before 8 to 10 dpi; the mucosal surface then becomes normal. After 7 dpi, the degree of damage is alleviated and the host response decreases; thus, fewer significant differences were observed at 21 dpi.

An increase in the number of downregulated genes was observed at 7 dpi, consistent with the previous finding that strong downregulation occurred in chicken immune cells after infection (Sandholt et al., 2021); the effect of self-regulation was evident at 7 dpi. Self-regulation may be an immune mechanism to alleviate inflammation; alternatively, *Eimeria* may induce downregulation of host pathways as an immune evasion mechanism (Sandholt et al., 2021).

Functional analysis of endoplasmic reticulum stress response under *Eimeria* infection

GO, KEGG, and GSEA assessments of the Type 1 expression pattern cluster in this study suggested that lipid and sterol synthesis and metabolism are highly active at 4 dpi during *Eimeria* infection. Furthermore, some studies have shown that host cells can undergo ER stress because of intestinal parasite infection (e.g., coccidiosis), leading to an unfolded protein response (Galluzzi et al., 2017). A previous microarray analysis of intraepithelial lymphocytes from *Eimeria*-infected chickens revealed activation of the ER stress response mechanism (Liu et al., 2021). The ER is an organelle that synthesizes lipids; when homeostasis is disrupted because of continuous ER stress and excessive unfolded protein response, lipid synthesis regulation fails, eventually leading to metabolic disorders and apoptosis (Han and Kaufman, 2016). Therefore, the rapid increases in lipid synthesis and metabolism in the chickens at 4 dpi in this study were caused by *Eimeria* infection-related ER stress. To overcome the risk of ER homeostasis disruption, cells reduce the ribosome and protein synthesis pathways by downregulating the transcription of rRNA; this reduces the overall burden on the ER (Hotamisligil, 2010; Han and Kaufman, 2016). Thus, the GSEA results in this study indicated that downregulation of the ribosome term at 4 dpi occurred in response to ER stress. Furthermore, GSEA of the data from 7 dpi indicated upregulation of intracellular ribosome function, suggesting that ER homeostasis had been restored by that time. Key Type 1 genes for lipid biosynthesis (e.g., *FDPS*, *BC O 1*, and *EXFABP*) showed significant downregulation at 7 dpi; this presumably aided in ER homeostasis.

Innate immune response after *Eimeria* infection

The findings in GO analysis of Type 3 genes could be explained by the innate immune response to *Eimeria* infection, including both the direct defense mechanisms and cytokines released *via* leukocyte activity and inflammatory responses. Innate immune responses were initially active and relieved afterward, following the chronological expression of Type 3 genes. In addition, functional enrichment analysis indicated that genes in the GCN53 network were mainly responsible for the innate immune mechanism. In the KEGG analysis of GCN53, DEGs included in influenza A and herpes simplex virus 1 infection-related terms showed high significance; these genes also have a protective function during coccidiosis. The activation of necroptosis and pattern recognition receptor (RIG-I-like receptor, NOD-like receptor, and Toll-like receptor) signaling pathways leads to the release of pathogen-associated molecular patterns or danger-associated molecular patterns through the death of infected cells. This release is likely to promote innate immune system activation and the upregulation of inflammatory responses (Franchi et al., 2009; Haunshi et al., 2017). In addition, among the terms upregulated at 4 dpi in the GSEA results, the virus defense response, MDA5 signaling, interferon I, response to cytokine stimulus, and adipocytokine signaling pathway terms supported the results of the analysis described above. Because the keratinization, cornification, epidermis development, and cornified envelope terms exhibited concomitant upregulation at 4 dpi, the development and differentiation of intestinal epithelial cells were presumed to defend against parasitic stimuli. Furthermore, the formation of a physical barrier through keratinization, such as the response of gizzard epithelial cells to parasitic infection in waterfowl (Padilla-Aguilar et al., 2020) provided a potential defense against *Eimeria* invasion. For the same purpose, bicellular tight junction assembly, a GO term upregulated at 7 dpi according to GSEA, forms a gut barrier through cell proliferation, differentiation, and organization (von Buchholz et al., 2021).

Cytokine-cytokine receptor interaction and the anti-inflammatory process in response to *Eimeria* infection

The cytokine-cytokine receptor interaction term had the largest number of genes (10 genes) among the KEGG enrichment analysis results of the Type 1 cluster. Genes related to the cytokine-cytokine receptor interaction term belonged to the Type 1 cluster; these included *IL13RA2* (interleukin 13 receptor subunit alpha 2), *CXCL13* (CXC motif chemokine ligand 13), *CXCR5* (CXC motif chemokine receptor 5), *GDF15* (growth differentiation factor 15), *IL7R* (interleukin 7 receptor), *IL12RB2* (interleukin 12 receptor subunit beta 2), *BMP7* (bone morphogenetic protein 7), *TNFRSF11B* (TNF receptor superfamily member 11b), *IL22* (interleukin 22), and *IFNGR1* (interferon-gamma receptor 1). These genes were involved in the inflammatory pathway; they were also responsible for activating the immune response through the migration of T and B lymphocytes. Importantly, growth differentiation factor 15 levels have been shown to increase in injured tissues; this gene has a role in inflammatory processes (Wischhusen et al., 2020). Moreover, interleukin 12 receptor subunit beta 2 contributes to the inflammatory response and host defense (Zou et al., 1997). All genes except *IFNGR1* were upregulated at 4 dpi during *Eimeria* infection, although they were significantly downregulated at 7 dpi. The *IFNGR1* gene codes interferon-gamma receptor 1; it interacts with interferon-gamma molecules. Interferon-gamma, a cytokine produced by macrophages, has an important role in cell-mediated acquired immunity; it functions in a synergistic manner with major histocompatibility complex (MHC) class II molecules (Yun et al., 2000). There have also been reports of the protective immune functions of interferon-gamma against *Eimeria* (Shah et al., 2010). However, in this study, significant downregulation of the *IFNGR1* gene continued at 4 and 7 dpi. The interferon-gamma-related function may have been activated in chickens immediately after infection (i.e., before the first sampling), and the downregulation began prior to 4 dpi. Similar to the immune systems described by Sandholt et al. (2021), the cytokine-cytokine receptor interaction pathway exhibits continuous downregulation to suppress the *Eimeria*-induced inflammatory response and autoimmunity after the initial stage of innate immune activity.

KEGG analyses of GCN40 revealed that the cytokine-cytokine receptor interaction pathway exhibited the highest enrichment score; BP terms related to the regulation of the immune system process were identified in the GO analysis of GCN40. The *BTK*, *SYK*, and *FAM65B* genes were the enriched core genes for the BP terms. *BTK* and *SYK* together have critical roles in the immune response that involve the activation of PLC-gamma, which is necessary for the activity and migration of B and T lymphocytes (Kurosaki and Tsukada, 2000). In the present study, *BYK* and *SYK* genes were significantly downregulated at 4 and 7 dpi, respectively. In contrast, the *FAM65B* gene was significantly upregulated at 7 dpi, in association with negative regulation of the leukocyte activation GO term. Therefore, while an active innate immune response was underway against *Eimeria* invasion, opposing mechanisms to inhibit lymphocyte activity were activated for anti-inflammatory effects, as described above.

Conclusion

In this study, time-series host response pathways in *Eimeria*-infected broiler chicken were observed via RNA-Seq. From 4dpi to 21dpi, *Eimeria*-infected chickens experienced dynamic changes; a total of three types of gene expression patterns, which are characterized by fluctuant up and down regulation and the different number of DEGs. These transitions reflected the real-time state of immune and homeostasis mechanisms against parasitic invasion. Upon closer inspection, organelle malfunction and activated innate immune response occurred at 4 dpi, and the next, recovery from the impairment and inflammation control were dominant at 7 dpi. Eventually, there were few significant DEGs between infected and uninfected birds at 21dpi. In other words, we summarized that the main host responses were related to ER stress-induced functional changes and the signaling systems responsible for the innate immune response. Additionally, gene clustering analysis and GCN analysis revealed networks of genes that play important roles during *Eimeria* infection. For example, the Type 3 cluster and GCN53 genes are mostly involved with innate immune response functions, and the part of Type 1 cluster and GCN40 genes mainly work for inflammation and anti-inflammation respectively. The identification of significant DEGs and the results of the gene grouping analysis in this study will help to improve disease control by aiding in the selection of chickens.

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: NCBI SRA BioProject, accession no: PRJNA844579.

Ethics statement

The animal study was reviewed and approved by the Animal Ethics Committee of Chungnam National University.

Author contributions

JL provided experimental conception and journal selection. MK carried out data analysis and manuscript writing. PM and MY contributed to the experimental design and sample

preparation. YC and DS contributed to supervision of data analysis and preparation of figures. SC and EC involved in interpretation of the analysis results. TE participated in manuscript preparation. SN assisted with qRT-PCR experiment. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by a grant from the National Research Foundation, Republic of Korea (grant number MHC-2019R1F1A1061670), and an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (grant number 2020-0-01441, Artificial Intelligence Convergence Research Center, Chungnam National University).

Acknowledgments

We would like to thank Byeonghwi Lim, Chiwoong Lim, and Jun-Mo Kim of Department of Animal Science and Technology in Chung-Ang University for consultations and help with analyses.

Conflict of interest

Author DS was employed by the company Research Institute TNT Research 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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.886781/full#supplementary-material

References

Blake, D. P., and Tomley, F. M. (2014). Securing poultry production from the ever-present Eimeria challenge. *Trends Parasitol.* 30 (1), 12–19. doi:10.1016/j.pt.2013.10.003

Bremner, A., Kim, S., Morris, K. M., Nolan, M. J., Borowska, D., Wu, Z., et al. (2021). Kinetics of the cellular and transcriptomic response to Eimeria maxima in relatively resistant and susceptible chicken lines. *Front. Immunol.* 12, 653085. doi:10.3389/fimmu.2021.653085

Chapman, H., Jeffers, T., and Williams, R. (2010). Forty years of monensin for the control of coccidiosis in poultry. *Poult. Sci.* 89 (9), 1788–1801. doi:10.3382/ps.2010-00931

Choi, J., Ko, H., Tompkins, Y. H., Teng, P.-Y., Lourenco, J. M., Callaway, T. R., et al. (2021). Effects of Eimeria tenella infection on key parameters for feed efficiency in broiler chickens. *Animals.* 11 (12), 3428. doi:10.3390/ani11123428

Franchi, L., Warner, N., Viani, K., and Nuñez, G. (2009). Function of Nod-like receptors in microbial recognition and host defense. *Immunol. Rev.* 227 (1), 106–128. doi:10.1111/j.1600-065X.2008.00734.x

Galluzzi, L., Diotallevi, A., and Magnani, M. (2017). Endoplasmic reticulum stress and unfolded protein response in infection by intracellular parasites. *Future Sci. OA* 3 (3), FSO198. doi:10.4155/fsoa-2017-0020

Han, J., and Kaufman, R. J. (2016). The role of ER stress in lipid metabolism and lipotoxicity. *J. Lipid Res.* 57 (8), 1329–1338. doi:10.1194/jlr.R067595

Haunshi, S., Burramsetty, A. K., Kannaki, T., Ravindra, K. R., and Chatterjee, R. (2017). Pattern recognition receptor genes expression profiling in indigenous chickens of India and White Leghorn. *Poult. Sci.* 96 (9), 3052–3057. doi:10.3382/ps/pex113

Hotamisligil, G. S. (2010). Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. *Cell* 140 (6), 900–917. doi:10.1016/j.cell.2010.02.034

Jeffers, T. (1981). Resistance and cross-resistance studies with narasin, a new polyether antibiotic anticoccidial drug. *Avian Dis.* 25, 395–403.

Kurosaki, T., and Tsukada, S. (2000). Blnk: connecting syk and btk to calcium signals. *Immunity* 12 (1), 1–5. doi:10.1016/s1074-7613(00)80153-3

Laurent, F., Mancassola, R., Lacroix, S., Menezes, R., and Naciri, M. (2001). Analysis of chicken mucosal immune response to Eimeria tenella and Eimeria maxima infection by quantitative reverse transcription-PCR. *Infect. Immun.* 69 (4), 2527–2534. doi:10.1128/IAI.69.4.2527-2534.2001

Li, C., Yan, X., Lillehoj, H. S., Oh, S., Liu, L., Sun, Z., et al. (2019). Eimeria maxima-induced transcriptional changes in the cecal mucosa of broiler chickens. *Parasit. Vectors* 12 (1), 285. doi:10.1186/s13071-019-3534-4

Li, W., Wang, M., Chen, Y., Chen, C., Liu, X., Sun, X., et al. (2020). EtMIC3 and its receptors BAG1 and ENDOUL are essential for site-specific invasion of Eimeria tenella in chickens. *Vet. Res.* 51 (1), 90. doi:10.1186/s13567-020-00809-6

Lillehoj, H., Kim, C.-H., Keeler, C., Jr, and Zhang, S. (2007). Immunogenomic approaches to study host immunity to enteric pathogens. *Poult. Sci.* 86 (7), 1491–1500. doi:10.1093/ps/86.7.1491

Lillehoj, H. S., and Trout, J. M. (1996). Avian gut-associated lymphoid tissues and intestinal immune responses to Eimeria parasites. *Clin. Microbiol. Rev.* 9 (3), 349–360. doi:10.1128/CMR.9.3.349-360.1996

Liu, B., Ma, X., and Cai, J. (2021). Construction and analysis of coexpression network to understand biological responses in Chickens infected by Eimeria tenella. *Front. Vet. Sci.* 8, 688684. doi:10.3389/fvets.2021.688684

Long, P., Johnson, J., and Wyatt, R. (1981). Pathological and clinical effects of Eimeria tenella in partially immune chickens. *J. Comp. Pathol.* 91 (4), 581–587. doi:10.1016/0021-9975(81)90087-6

Luo, W., and Brouwer, C. (2013). Pathview: an R/bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29 (14), 1830–1831. doi:10.1093/bioinformatics/btt285

Morris, G. M., Woods, W. G., Richards, D. G., and Gasser, R. B. (2007). Investigating a persistent coccidiosis problem on a commercial broiler–breeder farm utilising PCR-coupled capillary electrophoresis. *Parasitol. Res.* 101 (3), 583–589. doi:10.1007/s00436-007-0516-9

Padilla-Aguilar, P., Romero-Callejas, E., Ramírez-Lezama, J., Osorio-Sarabia, D., García-Prieto, L., Manterola, C., et al. (2020). Gastrointestinal helminths of waterfowl (Anatidae: Anatinae) in the Lerma marshes of central Mexico: Some pathological aspects. *Int. J. Parasitol. Parasites Wildl.* 13, 72–79. doi:10.1016/j.ijppaw.2020.07.008

Quiroz-Castañeda, R. E., and Dantán-González, E. (2015). Control of avian coccidiosis: Future and present natural alternatives. *Biomed. Res. Int.* 2015, 430610. doi:10.1155/2015/430610

Raman, M., Banu, S. S., Gomathinayagam, S., and Raj, G. D. (2011). Lesion scoring technique for assessing the virulence and pathogenicity of Indian field isolates of avian Eimeria species. *Veterinarski arh.* 81 (2), 259–271.

Sandholt, A. K., Xu, F., Söderlund, R., Lundén, A., Troell, K., Svärd, S. G., et al. (2021). Dual RNA-Seq transcriptome analysis of chicken macrophage-like cells (HD11) infected *in vitro* with Eimeria tenella. *Parasitology* 148 (6), 712–725. doi:10.1017/S0031182021000111

Shah, M. A. A., Song, X., Xu, L., Yan, R., Song, H., Ruirui, Z., et al. (2010). The DNA-induced protective immunity with chicken interferon gamma against poultry coccidiosis. *Parasitol. Res.* 107 (3), 747–750. doi:10.1007/s00436-010-1940-9

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Venkatas, J., and Adeleke, M. (2019). A review of Eimeria antigen identification for the development of novel anticoccidial vaccines. *Parasitol. Res.* 118 (6), 1701–1710. doi:10.1007/s00436-019-06338-2

von Buchholz, J. S., Bilic, I., Aschenbach, J. R., Hess, M., Mitra, T., Awad, W. A., et al. (2021). Establishment of a novel probe-based RT-qPCR approach for detection and quantification of tight junctions reveals age-related changes in the gut barriers of broiler chickens. *Plos one* 16 (3), e0248165. doi:10.1371/journal.pone.0248165

Wang, X., Zou, W., Yu, H., Lin, Y., Dai, G., Zhang, T., et al. (2019). RNA sequencing analysis of chicken cecum tissues following Eimeria tenella infection *in vivo*. *Genes* 10 (6), 420. doi:10.3390/genes10060420

Wischhusen, J., Melero, I., and Fridman, W. H. (2020). Growth/differentiation factor-15 (GDF-15): from biomarker to novel targetable immune checkpoint. *Front. Immunol.* 11, 951. doi:10.3389/fimmu.2020.00951

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, H., Zou, W., Mi, C., Wang, Q., Dai, G., Zhang, T., et al. (2021). Research Note: Expression of T cell-related cytokines in chicken cecal and spleen tissues following Eimeria tenella infection *in vivo*. *Poult. Sci.* 100 (7), 101161. doi:10.1016/j.psj.2021.101161

Yun, C., Lillehoj, H., and Lillehoj, E. (2000). Intestinal immune responses to coccidiosis. *Dev. Comp. Immunol.* 24 (2-3), 303–324. doi:10.1016/s0145-305x(99)00080-4

Zou, J., Presky, D. H., Wu, C.-Y., and Gubler, U. (1997). Differential associations between the cytoplasmic regions of the interleukin-12 receptor subunits beta1 and beta2 and JAK kinases. *J. Biol. Chem.* 272 (9), 6073–6077. doi:10.1074/jbc.272.9.6073

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 August 2022
DOI 10.3389/fgene.2022.909472

OPEN ACCESS

EDITED BY
Natalia A. Zinovieva,
L. K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

REVIEWED BY
Ottmar Distl,
University of Veterinary Medicine Hannover, Germany
Eui-Soo Kim,
Recombinetics, United States

*CORRESPONDENCE
T. C. Chokoe,
TlouC@dalrrd.gov.za

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

RECEIVED 31 March 2022
ACCEPTED 27 June 2022
PUBLISHED 09 August 2022

CITATION
Chokoe TC, Hadebe K, Muchadeyi FC, Nephawe KA, Dzomba EF, Mphahlele TD, Matelele TC and Mtileni BJ (2022), Conservation status and historical relatedness of South African communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker. *Front. Genet.* 13:909472. doi: 10.3389/fgene.2022.909472

Conservation status and historical relatedness of South African communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker

T. C. Chokoe[1,2]*, K. Hadebe[3], F. C. Muchadeyi[3], K. A. Nephawe[4], E. F. Dzomba[5], T. D. Mphahlele[1], T. C. Matelele[1] and B. J. Mtileni[4]

[1]Farm Animal Genetic Resources, Department of Agriculture, Land Reform and Rural Development, Pretoria, South Africa, [2]School of Agriculture & Environmental Sciences, University of Limpopo, Polokwane, South Africa, [3]Biotechnology Platform, Agricultural Research Council, Pretoria, South Africa, [4]Department of Animal Sciences, Tshwane University of Technology, Pretoria, South Africa, [5]Discipline of Genetics, School of Life Sciences, University of Kwazulu-Natal, Scottsville, South African

Indigenous goats form the majority of populations in smallholder, low input, low output production systems and are considered an important genetic resource due to their adaptability to different production environments and support of communal farming. Effective population size (N_e), inbreeding levels, and the runs of homozygosity (ROHs) are effective tools for exploring the genetic diversity and understanding the demographic history in efforts to support breeding strategies to use and conserve genetic resources. Across populations, the current N_e of Gauteng was the lowest at 371 animals, while the historical N_e across populations suggests that the ancestor N_e has decreased by 53.86%, 44.58%, 42.16%, and 41.16% in Free State (FS), North West (NW), Limpopo (LP), and Gauteng (GP), respectively, over the last 971 generations. Genomic inbreeding levels related to ancient kinship (F_{ROH} > 5 Mb) were highest in FS (0.08 ± 0.09) and lowest in the Eastern Cape (EC) (0.02 ± 0.02). A total of 871 ROH island regions which include important environmental adaptation and hermo-tolerance genes such as *IL10RB, IL23A, FGF9, IGF1, EGR1, MTOR,* and *MAPK3* were identified (occurring in over 20% of the samples) in FS (n = 37), GP (n = 42), and NW (n = 2) populations only. The mean length of ROH across populations was 7.76 Mb and ranged from 1.61 Mb in KwaZulu-Natal (KZN) to 98.05 Mb (GP and NW). The distribution of ROH according to their size showed that the majority (n = 1949) of the detected ROH were > 5 Mb in length compared to the other categories. Assuming two hypothetical ancestral populations, the populations from KZN and LP are revealed, supporting PC 1. The genomes of KZN and LP share a common origin but have substantial admixture from the EC and NW populations. The findings revealed that the occurrence of high N_e and autozygosity varied largely across breeds in communal indigenous goat populations at recent and ancient events when a genome-wide single-nucleotide polymorphism (SNP) marker was used. The use of Illumina goat

SNP50K BeadChip shows that there was a migration route of communal indigenous goat populations from the northern part (LP) of South Africa to the eastern areas of the KZN that confirmed their historical relatedness and coincides with the migration periods of the Bantu nation.

KEYWORDS

runs of homozygosity, effective population size, populations, genomic inbreeding, conservation

Introduction

There are currently 15 South African goat genetic resources listed on the DAD-IS of FAO and 13 on the Domestic Animal Genetic Resources Information System (DAGRIS) of the International Livestock Research Institute, including those listed in DAD-IS. In the country, indigenous goat ecotypes have been used as triple-purpose animals (e.g., skin, milk, and meat); and depending on the region, the animal characteristics, and the geographical isolation, they have begun to diverge into breeds/populations (Visser et al., 2016). These ecotypes have generally been named after their place of origin (e.g., Northern Cape Skilder) and/or their prominent characteristics (e.g., Xhosa lobbed ear) and on the basis of the people who own them (e.g., Nguni) (Morrison, 2007). These ecotypes are widely spread across all major agro-ecological regions of South Africa, displaying adaptability traits to a specific habitat or production environment and represent a significant resource to satisfy present and future demands for sustainable farming in a changing environment.

Improvement of indigenous livestock has been practiced through the introduction of high-performing breeds (exotic and improved breeds), and as a result of indiscriminate mating and breeding, the majority of communal indigenous goat populations are crossbreds (Mdladla et al., 2016a). The majority of the smallholder farmers have small herds or flocks where herd sizes could be less than five animals, with the majority of these herds being non-descript, crossbred or indigenous cattle, sheep, and goats (Mthi et al., 2017; Nyamushamba et al., 2017). The reduction in local indigenous populations suggests a need for the conservation of local genetic resources through the implementation of a national conservation strategy. Various studies conducted in the smallholder communal areas showed average flock sizes of between 1 and 120 goats (Webb and Mamabolo, 2004; Dube, 2015). Furthermore, about 70% of the goats are kept under traditional management systems where the farm structure comprises of about twenty (±20) indigenous goats (Monau et al., 2020). A detailed information on the phenotypic, genetic diversity and population structure of goats ecotype populations become important to guide conservation strategies through utilization of these populations. According to Food and Agriculture Organization of the United Nations, (2015b), conservation and characterization of animal genetic resources is critical because of their contribution to the sustainable livelihoods of rural communities that depend on them for food security. Conservation frameworks should incorporate both genetic diversity and breed merits for prioritizing breeds/populations from community to national level to support breeding programs of current populations.

In South Africa, more extensive research studies on genetic diversity analyses were done using microsatellite markers that were instrumental in providing an insight into the genetic structure and variation among South African goat populations (Kotze et al., 2004; Visser et al., 2004; Kotzé et al., 2014). (Kotzé et al., 2014) observed average heterozygosity of 63% in Kalahari Reds using 18 microsatellite markers, nine of which were used in the study by (Visser et al., 2004). Recently, microsatellites have been used to study genetic variation of the Tankwa feral goat population, which showed it to be highly divergent from the other farmed populations (Kotzé et al., 2014). In spite of their common use in most livestock diversity studies, microsatellites are often criticized for their usual location in the non-coding regions of the genome and for not being directly associated with genes that affect phenotypes. This has led to low-density microsatellites finding little application in studies of the adaptive genetic diversity of local breeds.

The completion of the first draft of the goat genome (Dong et al., 2013) made it possible for the development of high-density markers (Tosser-Klopp et al., 2014). The Illumina goat SNP50K BeadChip includes 53 347 SNPs (Tosser-Klopp et al., 2014) that have found utility in South African population genetic studies in Angora (Lashmar et al., 2016), commercial, indigenous, and village goat populations (Mdladla et al., 2017), as well as investigating genetic adaptation to environmental pressures (Mdladla et al., 2018). The use of the tool has been described in other African countries (Zvinorova, 2017; Onzima et al., 2018) and in specialized breeds (Martin et al., 2016). In South Africa, research work on the use of Illumina goat SNP50K BeadChip to determine the differences of indigenous communal goats at a point of genetic background is limited as compared to the studies where microsatellites were used. Therefore, while much work on South African commercial, indigenous, and village goat populations has focused on genetic studies and investigations on genetic adaptation, less work has focused on the conservation status and historical relatedness of communal indigenous goat populations.

The presence of the extent of an effective population size (*Ne*) is an important population genetic parameter that has

recently received a great deal of research attention (Zhao et al., 2014), determining population demographic development (Deng et al., 2019) and demographic processes such as migration and admixture (Nicoloso et al., 2015), and having profound implications for understanding the architecture of the animal genome (Lashmar et al., 2016; Alvarenga et al., 2018). In addition, *Ne* is widely regarded as one of the most critical population parameters because it measures the rates of genetic drift and inbreeding as well as affects the efficacy of systematic evolutionary forces such as mutation, selection, and migration (Shin et al., 2013; Silió et al., 2013). It also helps to discover population demographic history and allows for the prediction of the behavior of genetic diversity through time. The *Ne* is estimated using the r^2 coefficient and measures the observed range and the amount of genetic variation within a frame of population genetics (Berihulay et al., 2019). It also provides information on the degree of inbreeding of the population under consideration (Flury et al., 2010). The *Ne* determines the amount of genetic variation, genetic drift, and linkage disequilibrium (LD) in populations (Liu et al., 2017). Implementation of a national conservation strategy for FAnGR must be based on a better understanding of the degree of inbreeding of the populations, genetic variation, genetic drift, and linkage disequilibrium (LD) in those populations.

An increase in inbreeding (*F*) over generations leads to a reduction in genetic diversity (Onzima et al., 2018). Higher frequency of homozygous genotypes for deleterious alleles with a reduction in individual performance (inbreeding depression) and lower population viability (Ouborg et al., 2010). Offspring may inherit autozygotic chromosomal segments from both parents that are identical by descent (IBD) when they are inbred, i.e., segments that are derived from a common ancestor (Broman and Weber, 1999). The result is the runs of homozygosity (ROH), also known as the continuous homozygous segments in the genome. The ROHs are contiguous lengths of homozygous segments of the genome where the two haplotypes inherited from the parents are identical (Gibson et al., 2006).

The extent of ROH can be used to estimate the inbreeding coefficient (Bosse et al., 2012; Martin et al., 2016; Peripolli et al., 2018), to disclose the genetic relationships among individuals, usually estimating with high accuracy the autozygosity at an individual and/or population level (Ferenčaković et al., 2011; Ferenčaković et al., 2013). Autozygosity is the homozygous state of identical by descent (IBD) alleles, which can result from several phenomena such as genetic drift, population bottlenecks, mating of close relatives, and natural and artificial selection (Curik et al., 2014; Lashmar et al., 2016). The level of selection pressure on the populations can also be established by the use of ROH (Zhang et al., 2015). Distant from more recent inbreeding may also be distinguished by the length and frequency of ROHs, since the length of IBD segments follows an inverse exponential distribution with a mean of ½ *g* Morgans, where *g* is the number of generations from a common ancestor (Howrigan et al., 2011). Effective populations (*Ne*) and ROH have been studied in humans (Gibson et al., 2006), cattle (Ferenčaković et al., 2011; Marras et al., 2014; Mastrangelo et al., 2016), pigs (Ai et al., 2013; Shin et al., 2013), and sheep (Mastrangelo et al., 2017), but less comprehensively in other livestock species, such as goats, for designing conservation strategies, especially for the communal indigenous goats of South Africa. The objective of this study was to determine the conservation status and historical relatedness of South African communal indigenous goat populations using genome-wide SNP markers.

Methods

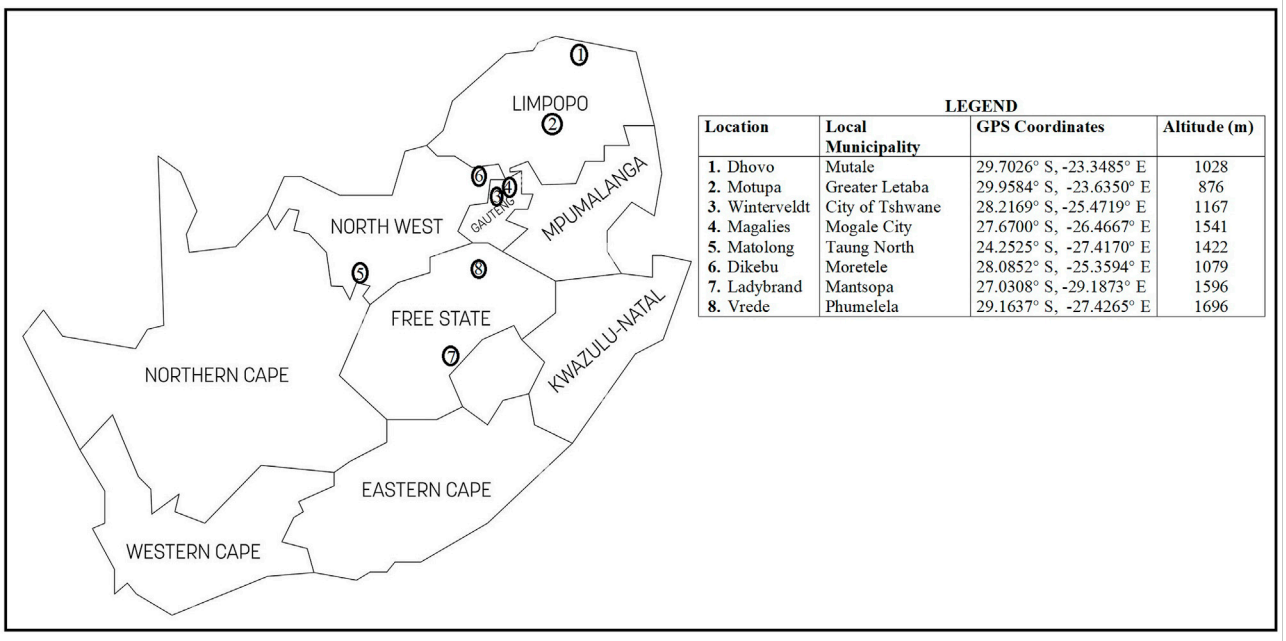
Ethics approval

Permission for the study and ethical approval were obtained from the animal research ethics committees of both the University of Limpopo and the Department of Agriculture, Land Reform and Rural Development. Furthermore, verbal consent was given by the goat owners.

Sample collection and genotyping

A total of 117 communal indigenous goat populations were sampled from the Free State ($n = 24$), Gauteng ($n = 28$), Limpopo ($n = 30$) and North West ($n = 35$) provinces of South Africa (Figure 1). Additionally, genotypes of communal indigenous goat populations that are under extensive production system (Mdladla, 2016) representing Eastern Cape ($n = 20$), KwaZulu-Natal ($n = 30$), Limpopo ($n = 30$) and North West ($n = 20$) provinces were included for further analysis.

All animals were genotyped using the Illumina GoatSNP50 BeadChip (Illumina Inc., San Diego, CA, United States) using the Infinium assay compatible with the Illumina HiScanSQ genotyping platform at the Agricultural Research Council—Biotechnology Platform in South Africa. A number of quality control measures were applied to all SNPs as follows: SNPs were removed if they had a call rate < 95%, a minor allele frequency (MAF) < 0.05, and if they deviated from the Hardy-Weinberg equilibrium (for a *p*-value cut-off of 0.00001), had no assigned genomic locations, and on sex chromosomes were also excluded from the analysis. The parameter "–cow" was used to indicate the number (i.e., 29) of autosomes in the goat genome since cows and goats have the same number of autosomes. The final dataset included 47,778 SNPs and 207 individuals.



LEGEND

Location	Local Municipality	GPS Coordinates	Altitude (m)
1. Dhovo	Mutale	29.7026° S, -23.3485° E	1028
2. Motupa	Greater Letaba	29.9584° S, -23.6350° E	876
3. Winterveldt	City of Tshwane	28.2169° S, -25.4719° E	1167
4. Magalies	Mogale City	27.6700° S, -26.4667° E	1541
5. Matolong	Taung North	24.2525° S, -27.4170° E	1422
6. Dikebu	Moretele	28.0852° S, -25.3594° E	1079
7. Ladybrand	Mantsopa	27.0308° S, -29.1873° E	1596
8. Vrede	Phumelela	29.1637° S, -27.4265° E	1696

FIGURE 1
South African map showing the geographical locations of communal indigenous goat populations.

Genetic diversity indices

Historical and recent effective population sizes (N_e) for each breed were estimated with the *SNeP* (Barbato et al., 2015), which is based on the relationship between linkage disequilibrium (LD), N_e and recombination rate. The different SNP marker distance bins used for r^2 analysis were used to obtain different estimates of *Ne,* at different time points by calculating the number of generations (t) in the past as ½c. To verify the accuracy of the coefficient of inbreeding, the genomic coefficient was estimated *via* two methods. 1) PLINK (Purcell et al., 2007) was used to measure the inbreeding coefficient based on the difference between the observed and expected numbers of homozygous genotypes (F_{HOM}) using the function –het. The calculation formula was as follows:

$$F_{HOM} = (E_{HOM} - O_{HOM})/(L - E_{HOM}),$$

where L is the number of genotyped autosomal SNPs, E_{HOM} is the number of expected homozygous genotypes, and O_{HOM} is the number of observed homozygous genotypes. The inbreeding coefficient based on the proportion of autosomes covered in runs of homozygosity per individual (F_{ROH}) was determined using *detectRUNs* (Biscarini et al., 2018). F_{ROH} was calculated as follows:

$$F_{ROH} = L_{ROH}/L_{AUTO},$$

where L_{ROH} is the total length of ROH on autosomes and L_{AUTO} is the total length of the autosomes covered by SNPs, which was 2450.71 Mb. Furthermore, the correlation between F_{ROH} and F_{HOM} was calculated for the four populations.

Distribution of runs of homozygosity

Runs of homozygosity (ROH) were identified in every individual using *detectRUNS* (Biscarini et al., 2018) using a sliding window of a specified length or number of homozygous SNPs to scan along with each individual's genotype at each SNP marker position to detect homozygous segments. The parameters and thresholds applied to define a ROH were: 1) a sliding window of 50 SNPs across the genome; 2) the minimum number of consecutive SNPs included in a ROH was 50; 3) the minimum length of a ROH was set to 1 Mb to avoid short and common ROH that occur throughout the genome due to LD (Purfield et al., 2017); and 4) a maximum of five SNPs with missing genotypes were allowed in a ROH to eliminate false positives. ROH was divided into five physical length categories (1–5 Mb, 5–10 Mb, 10–20 Mb, 20–30 Mb and < 40 Mb). The mean number of ROH per individual, the average length of ROH, the total number of ROH per animal, the percentage of chromosomes covered by ROH, and mean ROH were calculated on *detectRUNS*. The genomic inbreeding coefficient based on ROHs (F_{ROH}) was also calculated as the sum of the length of the autosome covered by ROHs divided by the overall length of the autosome covered by SNPs as described by

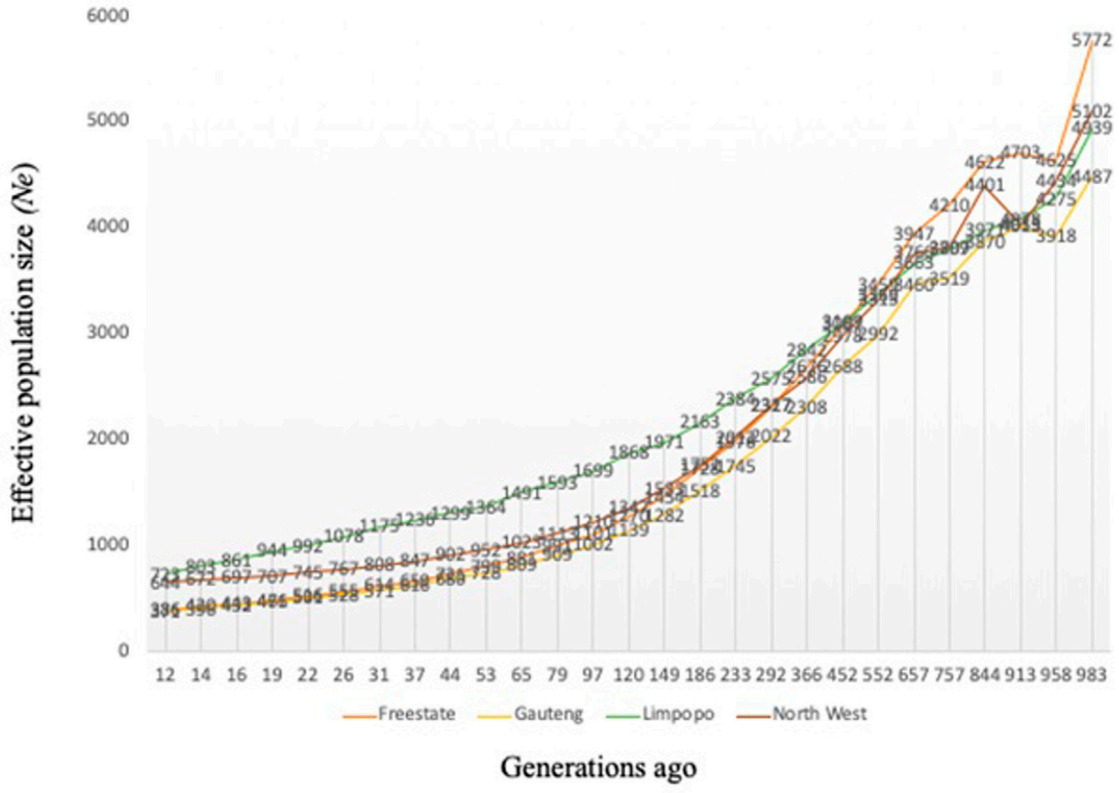


FIGURE 2
Effective population size (Ne) for the communal indigenous goat populations from Free State, Gauteng, Limpopo and North West.

(Mastrangelo et al., 2017). The means and standard deviation (sd) of F_{ROH} were calculated as the sum of the lengths of $F_{ROH1–5\ Mb}$, $F_{ROH5–10\ Mb}$, $F_{ROH10–20\ Mb\ or\ <\ 20\ Mb}$.

To identify the genomic regions that were most commonly associated with ROH, the percentage of the occurrences of a SNP in ROH was calculated by counting the number of times the SNP was detected in those ROH across individuals, and this was plotted against the position of the SNP along the chromosome. This percentage threshold was normalized to 70%, 50%, and 20% of individuals per population to be an indication of a possible hotspot of ROH in the genome. The functions of these genes and pathways in which they are involved were assessed using the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) database and literature search.

Population structure

Principal component analysis (PCA) was calculated and plotted in Golden Helix SNP and Variation Suite (SVS) V8.1 (Golden Helix, Inc. 2012). For the analysis of ancestry proportions (admixture) with K set from 2 to 10, the ADMIXTURE *v1.3.0* program (Alexander et al., 2009) was used. The default parameter of PLINK *v1.9* (50 SNPs step 5 SNPs, r^2 0.5) was used to subject the whole genotype data set to linkage disequilibrium (LD) pruning (Pemberton et al., 2012) prior to use in admixture analysis. Finally, to visualize admixture plots, GENESIS software was used (Buchmann and Hazelhurst, 2014).

Results

Genetic diversity indices

Variation of the estimated N_e at t generations ago (from 12 to 983) is presented in Figure 2. As expected, N_e decreased progressively across generations, however, N_e was higher than 150 for all breeds 12 generations ago. The variation in N_e across generations was smallest for Gauteng (N_e = 371) and the Free State (N_e = 386), whilst Limpopo had the highest (N_e = 723). Ancestral populations exhibited considerably larger N_e values, with the largest historical N_e values (N_e = 5,772).

TABLE 1 Distribution of runs of homozygosity inbreeding coefficients (F_{ROH}) within each population.

Class (Mb)	Eastern cape (n = 20)	Free state (n = 24)	Gauteng (n = 28)	KwaZulu-natal (n = 25)	Limpopo (n = 55)	North west (n = 55)
0–5 Mb	0.02 ± 0.02	0.08 ± 0.09	0.07 ± 0.09	0.04 ± 0.07	0.03 ± 0.04	0.08 ± 0.10
5–10 Mb	0.02 ± 0.01	0.06 ± 0.09	0.06 ± 0.09	0.05 ± 0.08	0.03 ± 0.04	0.07 ± 0.09
10–20 Mb	0.02 ± 0.01	0.06 ± 0.08	0.05 ± 0.09	0.08 ± 0.09	0.03 ± 0.05	0.07 ± 0.09
>20 Mb	0.03 ± 0.01	0.06 ± 0.04	0.08 ± 0.12	0.06 ± 0.05	0.07 ± 0.03	0.08 ± 0.09

TABLE 2 Number of runs of homozygosity (nROH) and length (in Mb) categorized by ROH length class ($ROH_{1-5\ Mb}$, $ROH_{5-10\ Mb}$, $ROH_{10-20\ Mb}$, $ROH_{20-40\ Mb}$ and $ROH_{>40\ Mb}$).

Class (Mb)	Eastern cape (n = 20)	Free state (n = 24)	Gauteng (n = 28)	KwaZulu-natal (n = 25)	Limpopo (n = 55)	North west (n = 55)
0–5 Mb	130	311	390	151	322	645
5–10 Mb	34	146	193	46	100	271
10–20 Mb	8	77	68	34	47	163
20–40 Mb	7	34	19	27	25	72
>40 Mb	1	6	12	6	8	30
Total	180	574	682	264	502	1181

The average inbreeding coefficient (F_{IS}) was lowest in Free State (F_{IS} = 0.03 ± 0.09), followed by 0.04 ± 0.09 in Gauteng, 0.05 ± 0.01 in the North West and the highest in Limpopo (F_{IS} = 0.09 ± 0.05). Overall, the inbreeding level was 0.06 ± 0.08. The average F_{ROH}, its range of variation across populations and its distribution are summarized in Table 1. Genomic inbreeding coefficients (F_{ROH}) based on the distribution of the length of runs of homozygosity by class are described in Table 2 and by chromosome in Figure 3. F_{ROH} differed significantly among populations across the length categories. The genomic inbreeding coefficients of the shortest ROH (0–5 Mb; related to ancient kinship) per animal ranged from 0.02 ± 0.02 in the Eastern Cape population to 0.08 ± 0.10 in the North West population. The F_{ROH} of Eastern Cape, Limpopo populations increased with category size, whilst decreased in Free State. Gauteng F_{ROH} decreased from 0.07 ± 0.09 to 0.05 ± 0.09 for F_{ROH} >20 Mb and increased at >20 Mb. In KwaZulu-Natal, F_{ROH} increased to up to 0.08 ± 0.09 at <20 Mb and decreased for > 20 Mb category.

Chromosomal distribution of inbreeding showed higher inbreeding levels in chromosome 15, in Gauteng chromosome 19, in KwaZulu-Natal in chromosome 19 followed by chromosome 16. In the North West, chromosomes 22 and 19 had the highest inbreeding coefficients compared to other chromosomes.

Distribution of runs of homozygosity

A total of 3383 ROH were identified across populations, although the frequency and length of ROH differed per population. Among the 207 animals, only 1 animal in the Eastern Cape population was lacking ROH, whilst 206 (99.52%) had at least one ROH longer than 1 Mb. The mean number of ROH per population was 24.36, 23.92, 21.47, 9.13, and 9 in Gauteng, Free State, North West, KwaZulu-Natal, Limpopo, and Eastern Cape, respectively (Figure 4).

The mean ROH length was 7.76 Mb across the population. The longest segment (SNP position 39467151–137516937) was observed in the Gauteng and North West populations and was 98.05 Mb in length (1992 SNPs) found on chromosome 1. In KwaZulu-Natal, Free State and Eastern Cape, the longest segments were found on chromosome 9 (83.54 Mb; 1691 SNPs), chromosome 7 (73.33 Mb; 1449 SNPs) and chromosome 7 (51.81 Mb; 1034 SNPs), respectively. ROH shorter than 5 Mb predominated (n = 1949) across all populations (Table 1), accounting for 57.61% of all detected segments and differed per population. These short segments accounted for 72.22% of the Eastern Cape population, followed by Limpopo (n = 322; 64%), KwaZulu-Natal (n = 151; 57.20%), Gauteng (n = 390; 57.18%), North West (n = 645; 54.61%), and Free State (n = 311; 54.18%).

The relationship between the mean number of ROH and the length of the genome covered by ROH per individual varies considerably among animals and populations. The number of ROH per chromosome displayed a specific pattern with the larger numbers found for chromosome 1, 2, and 3, a number that tended to decrease with chromosome length, and the smallest number on chromosome (Keller et al., 2011) with 44 segments. Chromosome 1 had the highest number of ROH and the Eastern Cape did not have ROH on chromosomes 23 and 28 (Figure 5).

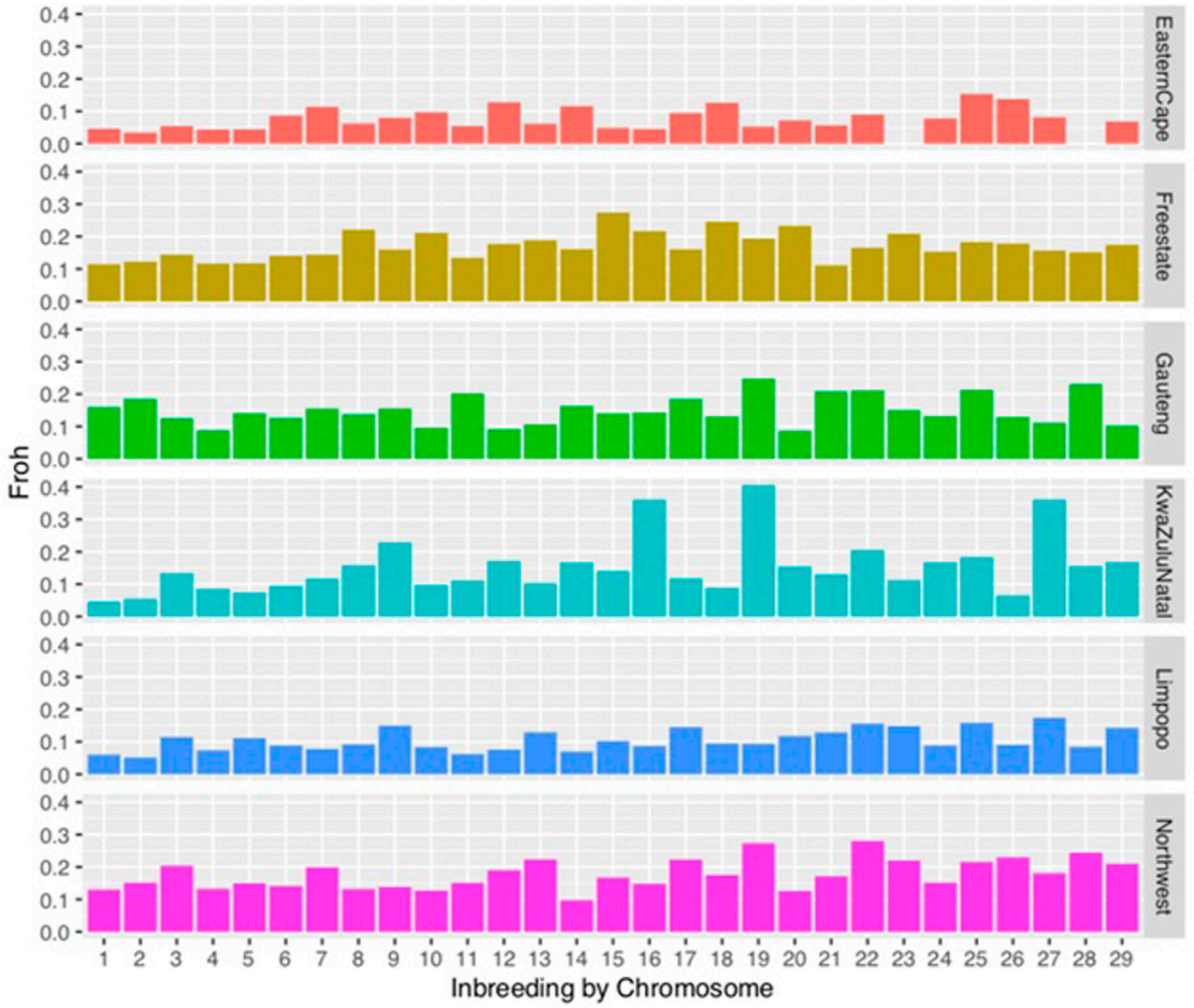


FIGURE 3
Distribution of inbreeding coefficients (FROH) based on runs of homozygosity (ROH) for each chromosome.

The proportion of chromosomes covered by ROH is illustrated in Figure 6. Overall, the highest coverage by ROH was observed on chromosomes 1, 2, 6, and 8 at 0.37, 0.31, 0.38, and 0.33, respectively. For the Eastern Cape population, chromosomes 4 and 8 had 0.07%, whilst Free State, Limpopo and North West had the highest proportion on chromosome 1 with 0.06%. Gauteng and KwaZulu-Natal had the highest on chromosome 6 with 0.06 and 0.08%, respectively.

To identify the genomic regions that were most commonly associated with ROH, the percentage of SNPs in ROH was assessed by analyzing the frequency of a SNP occurring in those ROH across different individuals (20%), and this was plotted against the position of the SNP along the chromosome (Figure 7). The threshold of 70% and 50% of the individuals did not yield ROH islands across populations, so 20% was used. Several genomic regions were identified that frequently appeared in ROH within individual animals (Table 3). We identified 58 ROH islands at the 20% threshold in the Free State (n = 28) and Gauteng (n = 29) provinces. No ROH islands were observed in the Eastern Cape, KwaZulu-Natal, North West, and Limpopo at the set thresholds. The ROH hotspot with the highest occurrences (SNPs = 149) in Gauteng was located on chr7 (7.69 Mb). Chromosome position, start and end position of ROH, ROH length and number of SNPs within the genomic regions of extended homozygosity are reported in Table 3 and Additional file 1, Supplementary Table S1. A total of 871 genes are inside the ROH islands, which include important environmental adaptation and hermo-tolerance genes such as

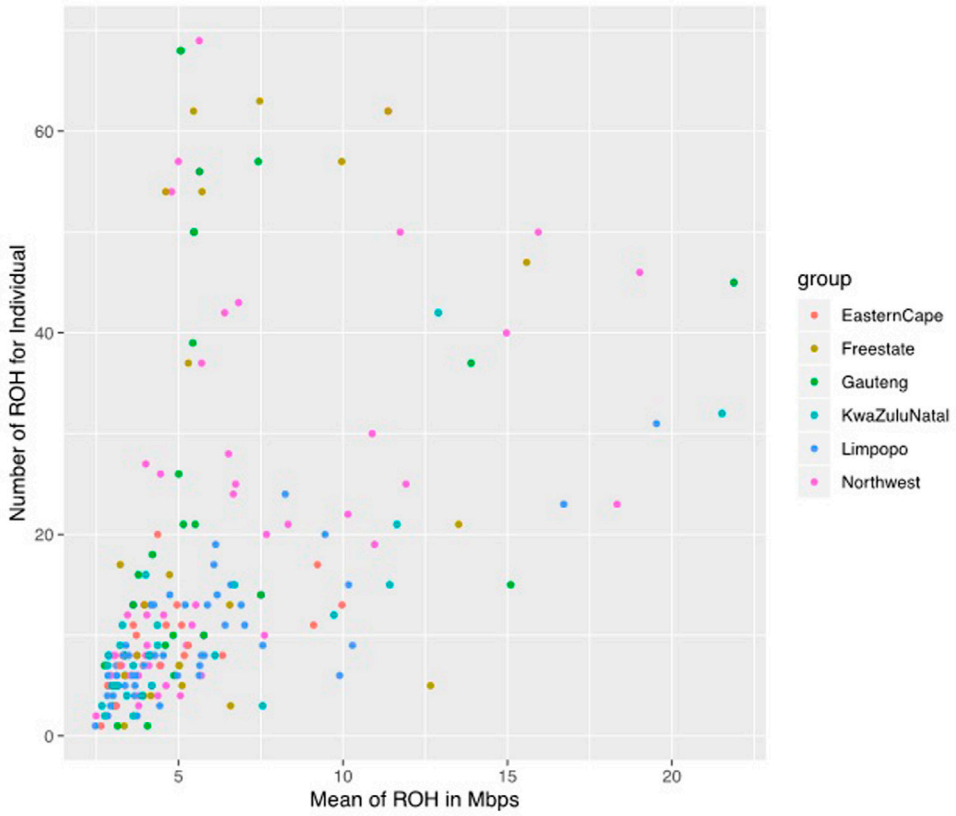


FIGURE 4
The genomic length with mean runs of homozygosity (ROH) per individual and the number of ROH for individuals.

IL10RB, IL23A, FGF9, IGF1, EGR1, MTOR and *MAPK3*. An additional file 1 shows the KEGG 292 pathways associated with the genes (Supplementary Table S2) that include Fructose and mannose metabolism, Starch and sucrose metabolism, Vitamin B6 metabolism, Protein export and Phototransduction.

Population structure

For population structure analysis, further quality control parameters were effected in PLINK v1.90 (Peripolli et al., 2018): to remove linked SNPs using the *--indep-pairwise 50 5 0.5* command and related individuals (IBS > 0.65). The database for population structure included 32886 SNPs and 189 individuals across populations. The plot of the first (PC1) and second eigenvectors (PC2) (Figure 8) shows weak differentiation among the clusters of admixed populations. PC1 shows clusters 1 and 2 consisting of populations from KwaZulu-Natal and Limpopo, respectively. The Eastern Cape population (Cluster 3) revealed a separate cluster (with three outliers). Clusters 4 consists of the majority of the North West and Gauteng, whilst cluster 5, was Free State.

For the further understanding of the degree of admixture within the populations, the ADMIXTURE 1.3 (Alexander et al., 2009) software was used for K = 2 to 10 hypothetical ancestral populations, only K = 2 to K = 6 is plotted since no further clusters were observed after K = 6 (Figure 9). Assuming two hypothetical ancestral populations, the populations from KwaZulu-Natal and Limpopo are revealed, supporting PC 1. The genomes of KwaZulu-Natal and Limpopo share an origin but have substantial admixture from the Eastern Cape and North West populations. K = 3 as the most likely number of genetically distinct groups within our populations, presenting the lowest cross-validation error (0.4617), reflecting a clear cluster of the Limpopo populations. The Free State, Gauteng, and North West showed similar genetic heterogeneity patterns with a considerable level of admixture. The North West revealed a high level of within-population genetic differentiation as there are individuals closer to the Eastern Cape and another subpopulation closer or clustering with Free State. This is also in agreement with the second PC coordinate analysis in showing genetic heterogeneity within the population. Moreover, with the increment of the value of K (K = 4 to

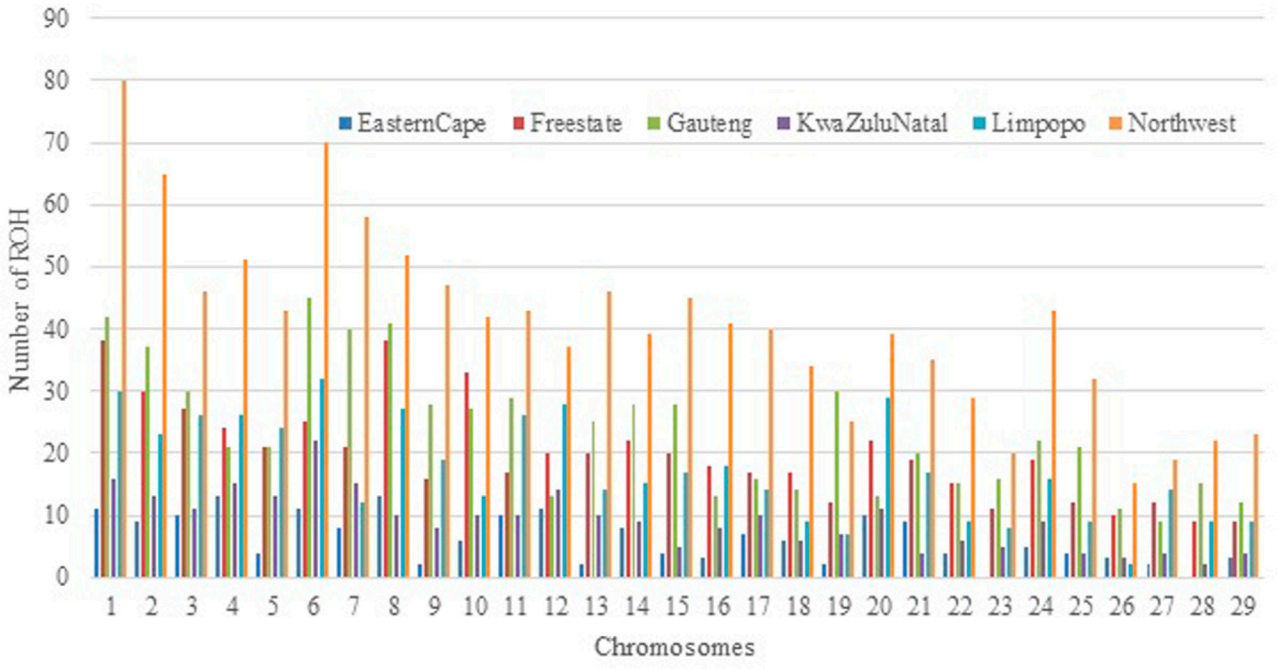


FIGURE 5
Number of runs of homozygosity (ROH) per chromosome identified across all populations.

K = 6), Free State and Gauteng show a higher level of genetic heterogeneity than the other populations.

Discussion

In this study, the genetic diversity and population structure of goat populations in South Africa were revealed. The main links to the development of commercial goat populations are represented by indigenous goat populations and may potentially be relevant as a future source of untapped adaptable genetic material (Webb and Mamabolo, 2004). Therefore, improving our understanding of within-ecotype relationships among the major goat producing provinces in South Africa offers a rare opportunity to enhance efficient use of the breeds and implement conservation programs. This study investigated the indications for population status on inbreeding and runs of homozygosity in the indigenous goat population. Data from a previous study (Mdladla et al., 2018) enabled a broad geographical coverage of South Africa and represented populations from the major goat producing provinces within the country.

Effective population size (N_e) is a crucial population genetic parameter because of its relationship to the loss of genetic variation, increases in inbreeding, the accumulation of mutations, and the determination of the accuracy of genomic selection (Goddard, 2009; Berihulay et al., 2019). Gauteng had the smallest estimated N_e among the population, and Limpopo had the highest. It was also observed in other studies that effective population size (N_e) showed a reduction to 132 in the Kingdom of Eswatini and the highest in South Africa 12 generations ago (Webb and Mamabolo, 2004). It is recommended that to prevent a reduction of the adaptive value in populations, N_e values between 50 and 100 animals (Meuwissen and Goddard, 2007) and 50 to avoid inbreeding depression. A study by (Colli et al., 2018), reported a large N_e in local goat breeds, such as those from Africa, Spain, and Central-Southern Italy, local goats breeds and a small *Ne* in the Angora, Boer, Nubian, Cashmere, Saanen, and Alpine populations.

The rapid increase pattern in *Ne* may also include bottlenecks associated with domestication, selection, and breed formation, as well as the endangerment of the breed (Shin et al., 2013). A study by (Mdladla, 2016) based on SNP data and using the same method, reported large N_e for all investigated breeds (ranging from 140 to 348). Furthermore, a study by (Mdladla et al., 2016a) revealed that the ecotype goat had a slightly higher effective population size than the Tankwa and commercial goat populations across generations. From a conservation standpoint, the indigenous goat population should be top priority in the population studied due to their diminishing effective population size and increased inbreeding coefficients.

Runs of homozygosity (ROH) can disclose the genetic relationships among individuals, estimating with high accuracy

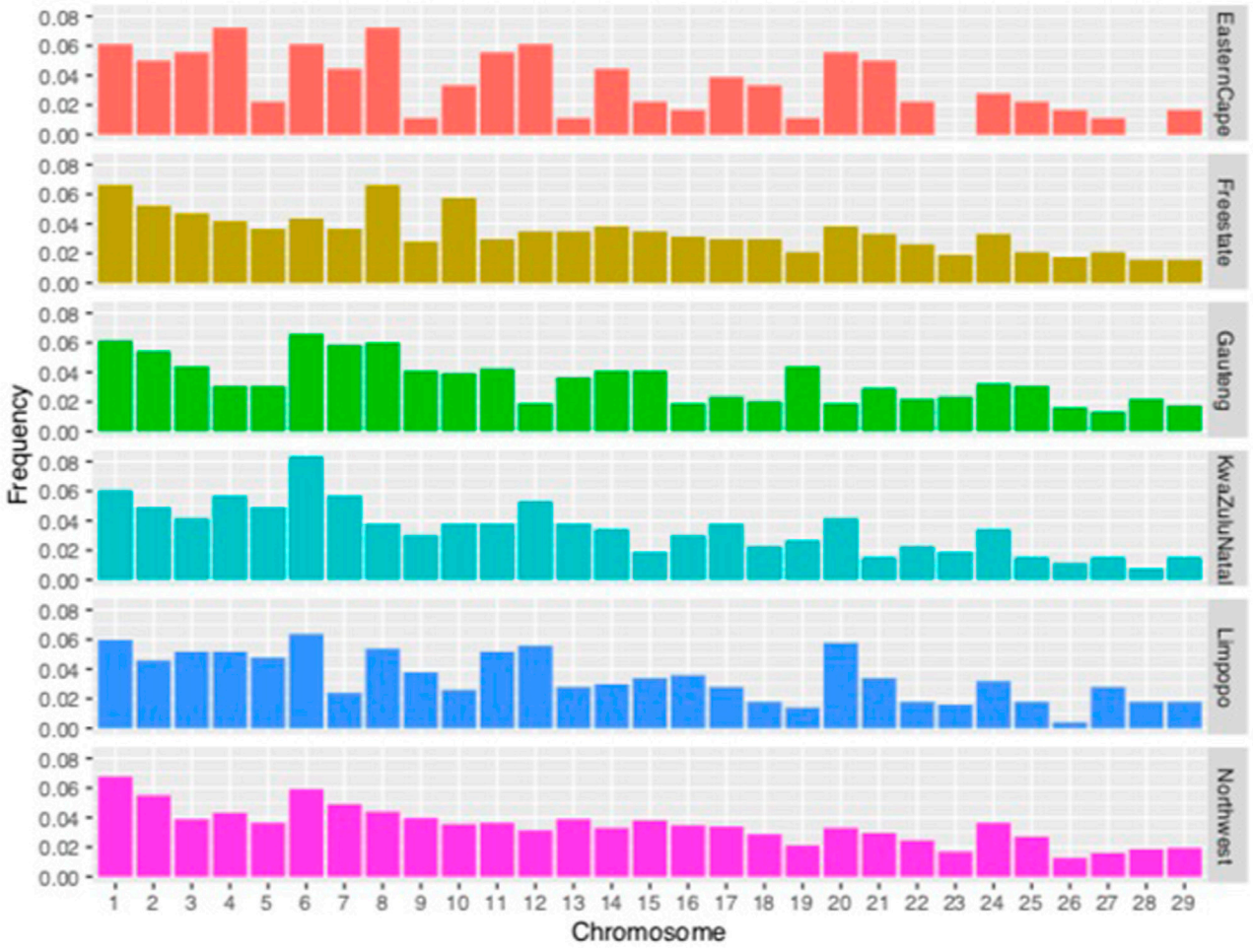


FIGURE 6
Frequency of runs of homozygosity (ROH) per chromosome per population.

the autozygosity at the individual and population levels and can elucidate about selection pressure events (Purfield et al., 2012). If long ROH accumulates in the genomes of some individuals, they could seriously impact the overall biological fitness (Manunza et al., 2016). Therefore, it was an important objective to investigate and understand the level of homozygosity among the populations. In this study, only 1 animal was lacking ROH, whilst 206 (99.52%) had at least one ROH longer than 1 Mb. The genomic inbreeding coefficients (F_{ROH}) values found in the study for the Ethiopian goats were $F_{ROH\ >\ 1\ Mb}$ values (Nandolo et al., 2019). Similar results were found by (Nandolo et al., 2019) with more African goats (Cameroon, Ethiopia, Kenya, Madagascar, Malawi, Mali, Mozambique, Nigeria, Tanzania, Uganda, and Zimbabwe) using a clustering algorithm. In this study, differences in terms of total number and length of ROH were short (>5 Mb) were more abundant (57.61%) (Onzima et al., 2018) reported results that showed lower than the $F_{ROH\ >\ 2\ Mb}$ for Kenya, Uganda, and Mozambique goat breeds when the Goat 50K BeadChip was used. On the other hand, the Eastern Cape population showed very low amounts of ROH. This has been suggested to be consistent with recent admixture in the individuals of Chinese cattle (Xu et al., 2019). Long segments were abundant in the North West population (Islam et al., 2019). A recent study revealed a high mean ROH in the long length category (>30 Mb), and their study suggested that inbreeding is more recent and is indicative of demographic decline. Considering the extensive management systems of goats in the region, these results might be likely even if some researchers have argued that such extensive systems may lead to inbreeding, especially where goats are on an extensive system or are shepherded with other flocks for some part of the year (Tefera et al., 2004; Rumosa Gwaze et al., 2009b). The lower inbreeding levels in African goats could be due to the openness of the breeding systems in most

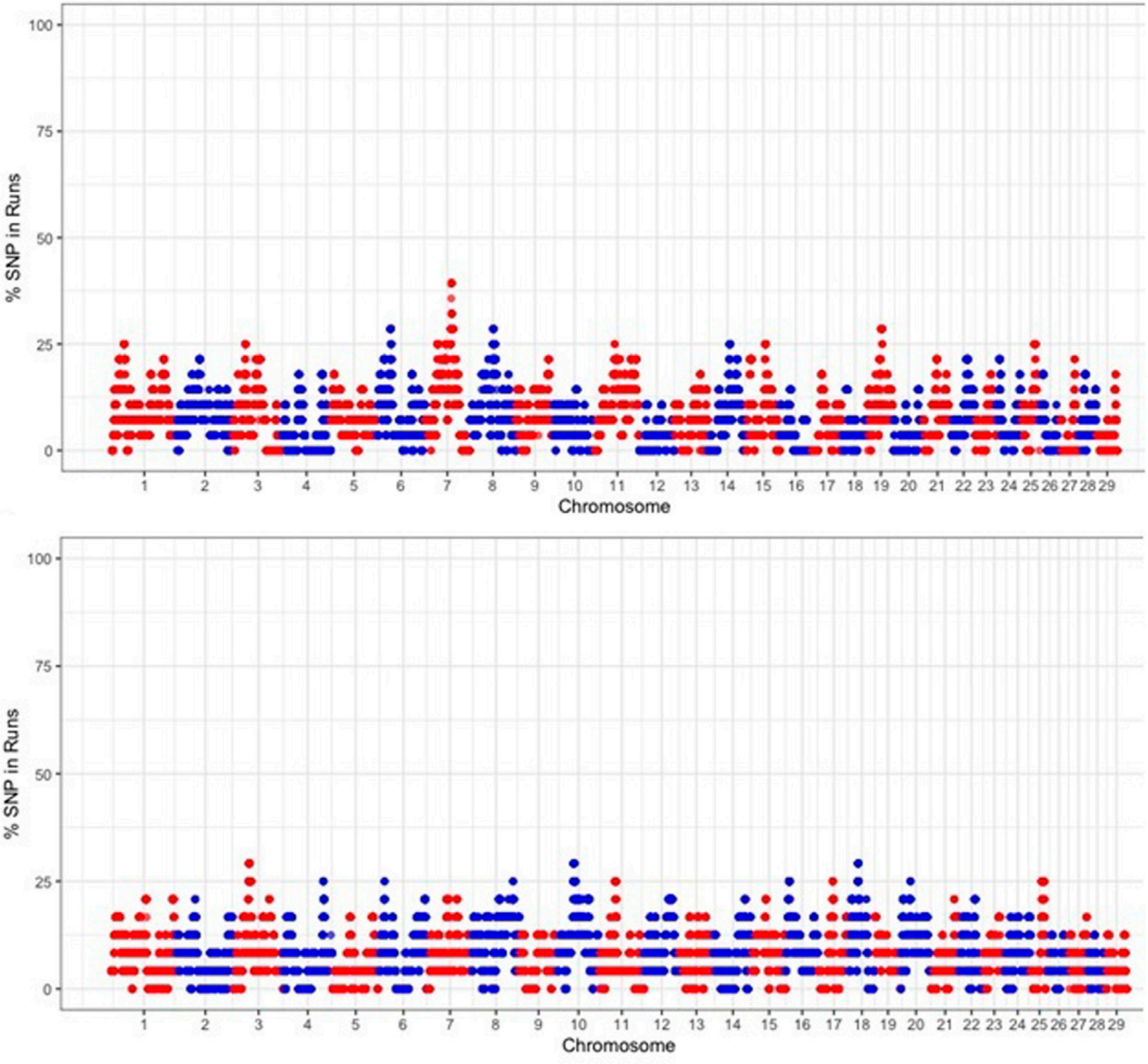


FIGURE 7
Manhattan plot of occurrences (%) of a SNP in ROH across populations.

of Africa that leads to the loose definition of livestock breeds in the region (Manunza et al., 2016).

One of the main advantages of genomic coefficients is the availability of chromosomal inbreeding coefficients (McQuillan et al., 2008; Mastrangelo et al., 2016). ROHs, representing the level of genomic autozygosity, are continuous homozygous segments at the individual and population levels that can be used as a measurement of inbreeding; more in-depth ROHs are the result of demography, natural and artificial selection, and inbreeding (Shi and Manley, 2007). In this study, we do not discuss in detail all the genomic regions associated with ROH but focus on some selected regions that show associations with several specific traits related to livestock breeding. We identified five genes reported to be associated with the important traits of goats (Figure 6) identified by the selection signature. Overall, the highest coverage by ROH was observed on chromosomes 1, 2,and 6, respectively. Gene INHA, located on chromosome 2, was reported as a candidate gene for litter size in goats (Hou et al., 2012). Significant QTL for milk production traits such as milk yield and milk protein have been reported on chromosome 2 in sheep

TABLE 3 Regions of the ROH islands at 20% across Gauteng and Free State populations by length.

Population	Position (chr: SNP 1: SNP 2)	nSNP	Length (Mb)
Gauteng	3:54056151:55070625	22	1.01
Gauteng	15:7897116:9088595	28	1.19
Gauteng	9:76995912:78321101	30	1.33
Gauteng	8:53874425:55257845	34	1.38
Gauteng	19:27892936:29379671	33	1.49
Gauteng	24:7056817:8823189	37	1.77
Gauteng	1:122842771:124683449	34	1.84
Gauteng	15:43908135:45815862	33	1.91
Gauteng	21:30374560:32314473	45	1.94
Gauteng	14:51115518:53088572	38	1.97
Gauteng	25:29189441:31200199	43	2.01
Gauteng	2:53997579:56229103	44	2.23
Gauteng	11:85024580:87384856	47	2.36
Gauteng	8:60260123:62640344	51	2.38
Gauteng	7:29457232:31898006	46	2.44
Gauteng	11:45829634:48502615	52	2.67
Gauteng	1:17079512:19890667	60	2.81
Gauteng	11:95518465:98340959	43	2.82
Gauteng	19:31681378:34691035	57	3.01
Gauteng	1:28400001:31438929	61	3.04
Gauteng	6:15586763:18769652	60	3.18
Gauteng	3:62095554:65301414	62	3.21
Gauteng	8:55933868:59160945	67	3.23
Gauteng	11:50602742:54104265	72	3.50
Gauteng	15:9294238:12900069	70	3.61
Gauteng	22:35705179:39950926	81	4.25
Gauteng	7:45438722:51833004	126	6.39
Gauteng	7:60187788:67883046	149	7.70
Freestate	25:28641340:31924694	69	3.28
Freestate	7:48193142:49278113	19	1.08
Freestate	20:21671108:22811848	22	1.14
Freestate	15:30999298:32153016	30	1.15
Freestate	20:4743026:6211601	32	1.47
Freestate	6:11346351:12878262	32	1.53
Freestate	20:19597454:21199311	37	1.60
Freestate	14:73908424:75554996	30	1.65
Freestate	22:36144242:37987490	36	1.84
Freestate	6:107507640:109477575	40	1.97
Freestate	25:25914911:28121785	48	2.21
Freestate	10:73412812:75622539	47	2.21
Freestate	4:94620783:96868758	55	2.25
Freestate	10:70170521:72471193	48	2.30
Freestate	1:81289909:83642225	50	2.35
Freestate	18:6154744:8563747	54	2.41
Freestate	1:145037345:147663352	55	2.63
Freestate	8:62328607:64991613	62	2.66
Freestate	11:36630830:39372620	53	2.74
Freestate	17:32672670:35522389	54	2.85
Freestate	8:94585706:97590188	52	3.00
Freestate	12:57673212:60742379	61	3.07
Freestate	16:5685888:9117064	69	3.43
Freestate	10:48160488:51636811	64	3.48
Freestate	18:20922080:24775492	76	3.85
Freestate	12:63521630:67564200	87	4.04
Freestate	10:36878056:43333293	124	6.46
Freestate	3:33683337:40785725	140	7.10

(Garcia-Gámez et al., 2013). Gene INHA, located on chromosome 2, was reported as a candidate gene for litter size in goats. The *PPP1R36* and Heat Shock Protein A2 (*HSPA2*) (*CHI10*, 26.402–26.719 Mb) identified in these communal indigenous goats are involved in heat stress response and, in other studies, *HSPA2*, *DNAJC24*, and *DNAJC*13 are associated with the heat shock family of genes (Shi and Manley, 2007). The presence of multiple genes associated with heat stress would seem to suggest that the trait is under intense selection pressure in tropically adapted breeds (Onzima et al., 2018).

In accordance with (Nothnagel et al., 2010), these regions in humans, when they are present in more than 50% of the individuals of a population, can indicate a strong selection occurrence. The occurrence of ROH hotspots in genomic regions that harbor candidate genes may be involved in selection pressure in response to production and environmental conditions. This study identified 58 ROH hotspots in Gauteng and Free State populations and revealed 871 genes and 292 KEGG pathways. This threshold did not yield any results, and only about 20% of ROH islands were detected in the communal indigenous goat population. The 20% threshold has been used in indigenous Chinese pigs, Jinhua (Xu et al., 2019). ROH islands can be defined as genomic regions with reduced genetic diversity and, consequently, high homozygosity around the selected locus that might harbor targets of positive selection and are under strong selective pressure (Pemberton et al., 2012; Peripolli et al., 2018). The minimum expected length of homozygous DNA segments is based on the time frame of approximately 25 generations, over which goats are believed to have been characterized into separate breeds (Onzima et al., 2018).

This study explored the population genetic structure of the indigenous goat population in the context of all South African goat populations. In accordance with our earlier studies (Lashmar et al., 2016; Mdladla et al., 2017), the principal component analysis (PCA), and the ADMIXTURE analyses

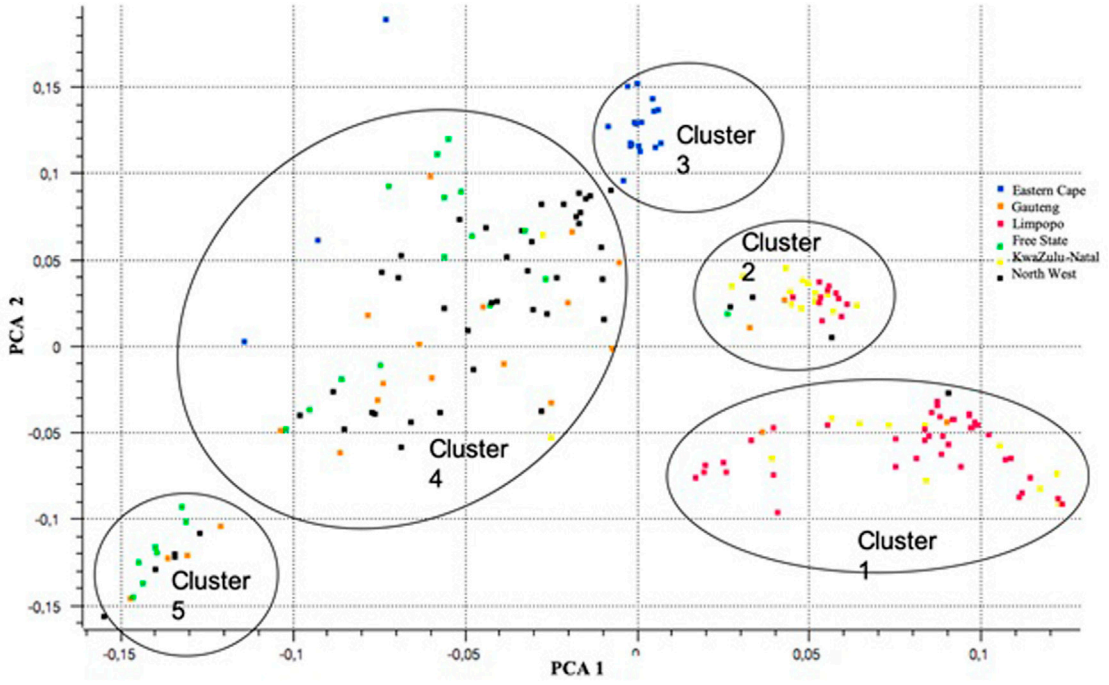


FIGURE 8
A plot of principal components (PCA) analysis showing differentiations among the clusters of admixtures across six communal indigenous goat populations.

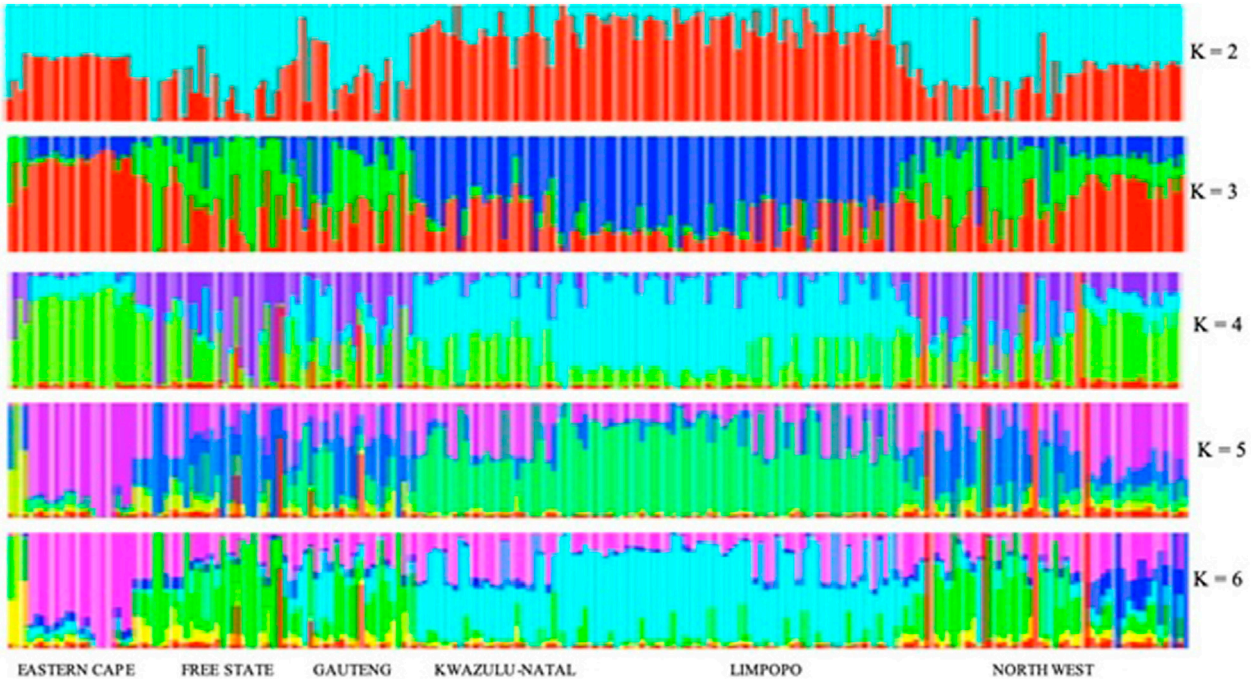


FIGURE 9
Clusters inferred from ADMIXTURE at K = 2–6.

based on the SNP array and sequence data sets, capitulated the major genetic division among the South African goat populations from two large geographic regions: Eastern Cape and Limpopo. The KwaZulu-Natal and Limpopo populations' genomes shared an origin yet with significant admixture from the Eastern Cape and North West populations. Some signals of admixture and underlying genetic relationships among the populations were generated by analysis of population admixture (Hou et al., 2012). A migration route of ancient goat from the northern part of South Africa to the eastern areas of the KZN, during their migration periods of the Bantu nation, is supported by this study as observed.

Conclusion

The results of this study indicated a greater negative impact of inbreeding in recent times, which is important for planning conservation strategies. It was revealed that the occurrence of high N_e and autozygosity varied largely across ecotypes in communal indigenous goat populations at recent and ancient events when a genome-wide SNP marker was used. The use of Illumina goat SNP50K BeadChip shows that there was a migration route of communal indigenous goat populations from the northern part (Limpopo province) of South Africa to the eastern areas of KwaZulu-Natal, which confirmed their historical relatedness and coincides with the migration periods of the Bantu nation. The communal-traditional indigenous goat farming system and adaptation to different climatic conditions had an influence on the results in this study. The study deepened the understanding of the conservation status and selection mechanisms of goats in a communal indigenous goat production setting. For effective conservation programs and utilization of South African communal indigenous goat populations, effort should be made to establish a conservation program for the unique genetic resources of indigenous goat populations.

Data availability statement

The datasets presented in this study can be found in online repositories. The name of the repository and link to the data can be found at: Dryad; https://doi.org/10.5061/dryad.931zcrjnh.

Ethics statement

The animal study was reviewed and approved by the DALRRD Animal Ethics Committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

Conceptualization: TC, KH, and BM. Formal analysis: TC, KH, FM, ED, KN, and BM. Investigation: TC, KH, TDM, and TCM. Project administration: TC. Resources: TC and FM. Supervision: BM, KH, and FM. Writing—review, and editing: TC, KH, and BM. All authors read and approved the final manuscript.

Funding

This study was supported by a research grant from the Department of Agriculture, Land Reform and Rural Development's Directorate of Genetic Resources, Directorate of Policy Research and Support and The South African National Research Foundation to fund the fieldwork data from the project "Characterization of indigenous goats" as well as full genome sequences of indigenous goats. Department of Agriculture, Land Reform and Rural Development Grant No: 2 GR ANIMAL (NATIONAL FUNCTION). National Research Foundation Grant No: 129349.

Acknowledgments

The first author would like to acknowledge the support from the University of Limpopo by allowing him to be their Ph.D. student.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.909472/full#supplementary-material

References

Ai, H., Huang, L., and Ren, J. (2013). Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One* 8 (2), e56001.

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

Alvarenga, A. B., Rovadoscki, G. A., Petrini, J., Coutinho, L. L., Morota, G., Spangler, M. L., et al. (2018). Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. *Sci. Rep.* 8, 8851. doi:10.1038/s41598-018-27259-7

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

Berihulay, H., Islam, R., Jiang, L., and Ma, Y. (2019). Genome-wide linkage disequilibrium and the extent of effective population sizes in six Chinese goat populations using a 50K single nucleotide polymorphism panel. *Animals* 9, 350. doi:10.3390/ani9060350

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. (2018). *bioinformatics*. Available at: https://github.com/bioinformatics-ptp/detectRUNS/tree/master/detectRUNS.

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L. A. F., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: Consequence of demography and the recombination landscape. *PLOS Genet.* 8, e1003100. doi:10.1371/journal.pgen.1003100

Broman, K. W., and Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Étude du Polymorphisme humain. *Am. J. Hum. Genet.* 65, 1493–1500. doi:10.1086/302661

Buchmann, R., and Hazelhurst, S. (2014). *Genesis manual*. Johannesburg: University of the Witwatersrand.

Colli, L., Milanesi, M., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., et al. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet. Sel. Evol.* 50, 58. doi:10.1186/s12711-018-0422-x

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi:10.1016/j.livsci.2014.05.034

Deng, T., Liang, A., Liu, J., Hua, G., Ye, T., Liu, S., et al. (2019). Genome-wide SNP data revealed the extent of linkage disequilibrium, persistence of phase and effective population size in purebred and crossbred buffalo populations. *Front. Genet.* 9, 688. doi:10.3389/fgene.2018.00688

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, 135–141. doi:10.1038/nbt.2478

Dube, K. (2015). *Characterization of goat production systems in selected coastal areas of the Eastern Cape Province, South Africa*. PhD Thesis.

Ferenčaković, M., Hamzic, E., Gredler, B., Curik, I., and Sölkner, J. (2011). Runs of homozygosity reveal genome-wide autozygosity in the Austrian fleckvieh cattle. *Agric. Conspec. Sci.* 76, 325–328.

Ferenčaković, M., Hamzić, 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

Flury, C., Tapio, M., Sonstegard, T., Drögemüller, C., Leeb, T., Simianer, H., et al. (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet.* 127, 339–347. doi:10.1111/j.1439-0388.2010.00862.x

Food and Agriculture Organization of the United Nations (2015b). *Draft second report on the state of the world's animal genetic resources for food and agriculture (Part3), commission on genetic resources for food and agriculture, (CGRFA-15/15/Inf.17.2)*. Rome. From: Available at: http://www.fao.org/3/a-mm310e.pdf (Accessed August 24, 2018).

Garcia-Gámez, E., Gutiérrez-Gil, B., Suarez-Vega, A., de la Fuente, L. F., and Arranz, J. J. (2013). Identification of quantitative trait loci underlying milk traits in Spanish dairy sheep using linkage plus combined linkage disequilibrium and linkage analysis approaches. *J. Dairy Sci.* 96, 6059–6069.

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

Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi:10.1007/s10709-008-9308-0

Hou, J., An, X., Li, G., Wang, Y., Song, Y., and Cao, B. (2012). Exploring polymorphisms and their effects on reproductive traits of the INHA and INHβA genes in three goat breeds. *Anim. Sci. J.* 83, 273–278. doi:10.1111/j.1740-0929.2011.00968.x

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics* 12, 460. doi:10.1186/1471-2164-12-460

Islam, R., Li, Y., Liu, X., Berihulay, H., Abied, A., Gebreselassie, G., et al. (2019). Genome-wide runs of homozygosity, effective population size, and detection of positive selection signatures in six Chinese goat breeds. *Genes.* 10, 938. doi:10.3390/genes10110938

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

Kotzé, A., Grobler, J., Van Marle-Köster, E., Jonker, T., and Dalton, D. (2014). The Tankwa karoo national park feral goat population: A unique genetic resource. *SA J. An. Sci.* 44 (1), 43–48. doi:10.4314/sajas.v44i1.6

Kotze, A., Swart, H., Grobler, J. P., and Nemaangani, A. (2004). A genetic profile of the Kalahari Red goat breed from Southern Africa. *S A. J. Anim. Sci.* 34, 10–12.

Lashmar, S. F., Visser, C., and Marle-Köster, E. v. (2016). SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. *Small Ruminant Res.* 136, 65–71. doi:10.1016/j.smallrumres.2016.01.006

Liu, S., He, S., Chen, L., Li, W., Di, J., and Liu, M. (2017). Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (Xinjiang type) sheep by genome-wide SNPs. *Genes. Genom* 39, 733–745. doi:10.1007/s13258-017-0539-2

Manunza, A., Noce, A., Serradilla, J. M., Goyache, F., Martínez, A., Capote, J., et al. (2016). A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genet. Sel. Evol.* 48, 52. doi:10.1186/s12711-016-0229-6

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., et al. (2014). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* 46, 110–121. doi:10.1111/age.12259

Martin, P. M., Palhière, I., Ricard, A., Tosser-Klopp, G., and Rupp, R. (2016). Genome wide association study identifies new loci associated with undesired coat color phenotypes in saanen goats. *PLoS ONE* 11, e0152426. doi:10.1371/journal.pone.0152426

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M. T., and Portolano, B. (2016). Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal* 10, 746–754. doi:10.1017/s1751731115002943

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 (1), 84. doi:10.1186/s12711-017-0360-z.hal-01635193

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 (3), 359–372. doi:10.1016/j.ajhg.2008.08.007

Mdladla, K., Dzomba, E. F., Huson, H. J., and Muchadeyi, F. C. (2016a). Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Anim. Genet.* 47, 471–482. doi:10.1111/age.12442

Mdladla, K., Dzomba, E. F., and Muchadeyi, F. C. (2017). Characterization of the village goat production systems in the rural communities of the eastern Cape, KwaZulu-natal, Limpopo and North West provinces of South Africa. *Trop. Anim. Health Prod.* 49, 515–527. doi:10.1007/s11250-017-1223-x

Mdladla, K., Dzomba, E. F., and Muchadeyi, F. C. (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* 120 (4), 369–378. doi:10.1038/s41437-017-0044-z

Mdladla, K. (2016). Landscape genomic approach to investigate genetic adaptation in South African indigenous goat populations. *PhD thesis*.

Meuwissen, T. H. E., and Goddard, M. E. (2007). Multipoint identity-by-descent prediction using dense markers to map quantitative trait loci and estimate effective population size. *Genetics* 176, 2551–2560. doi:10.1534/genetics.107.070953

Monau, P., Raphaka, K., Zvinorova-Chimboza, P., and Gondwe, T. (2020). Sustainable utilization of indigenous goats in southern Africa. *Diversity* 12, 20. doi:10.3390/d12010020

Morrison, J. W. (2007). *A guide to the identification of the natural indigenous goats of Southern Africa*. From: http://www.landbou.com/wp.../03/f2297405-a93f-4399-bdb7-6f3de538d75d.pdf (Accessed April 20, 2018).

Mthi, S., Skenjana, A., and Fayemi, P. O. (2017). Characteristics of small-scale sheep production systems in some communal areas of the Eastern Cape Province, South Africa. *Int. J. Livest. Prod.* 8, 199–206.

Nandolo, W., Mészáros, G., Banda, L. J., Gondwe, T. N., Lamuno, D., Mulindwa, H. A., et al. (2019). Timing and extent of inbreeding in african goats. *Front. Genet.* 10, 537. doi:10.3389/fgene.2019.00537

Nicoloso, L., Bomba, L., Bomba, L., Colli, L., Negrini, R., Milanesi, M., et al. (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet. Sel. Evol.* 47, 62. doi:10.1186/s12711-015-0140-6

Nothnagel, M., Lu, T. T., Kayser, M., and Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol.Genet.* 19, 2927–2935. doi:10.1093/hmg/ddq198

Nyamushamba, G. B., Mapiye, C., Tada, O., Halimani, T. E., and Muchenje, V. (2017). Conservation of indigenous cattle genetic resources in southern africa's smallholder areas: Turning threats into opportunities - a review. *Asian-Australas J. Anim. Sci.* 30, 603–621. doi:10.5713/ajas.16.0024

Onzima, R. B., Upadhyay, M. R., Doekes, H. P., Brito, L. F., Bosse, M., Kanis, E., et al. (2018). Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. *Front. Genet.* 9, 318. doi:10.3389/fgene.2018.00318

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R., and Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends Genet.* 26, 177–187. doi:10.1016/j.tig.2010.01.001

Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., and Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* 91, 275–292. doi:10.1016/j.ajhg.2012.06.014

Peripolli, E., Stafuzza, N. B., Munari, D. P., Lima, A. L. F., Irgang, R., Machado, M. A., et al. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics* 19, 34. doi:10.1186/s12864-017-4365-3

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., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13, 70. doi:10.1186/1471-2156-13-70

Purfield, D. C., McParland, S., Wall, E., and Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* 12, e0176780. doi:10.1371/journal.pone.0176780

Rumosa Gwaze, F., Chimonyo, M., and Dzama, K. (2009b). Communal goat production in southern Africa: A review. *Trop. Anim. Health Prod.* 41, 1157–1168. doi:10.1007/s11250-008-9296-1

Shi, Y., and Manley, J. L. (2007). A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. *Mol. Cell.* 28, 79–90. doi:10.1016/j.molcel.2007.08.028

Shin, D.-H., Cho, K.-H., Park, K.-D., Lee, H.-J., and Kim, H. (2013). Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. *Asian Australas. J. Anim. Sci.* 26, 1672–1679. doi:10.5713/ajas.2013.13320

Silió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., and Óvilo, C. (2013). Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J. Anim. Breed. Genet.* 130, 349–360.

Tefera, A. N., Mekala, D. G., Mnisi, P. E., Mukisira, M., Clerkson, M., Murungweni, C., et al. (2004). Goat production and livelihood systems in the Sekhukhune district of the Limpopo Province, South Africa: Opportunities for commercializing goats and their by-products. *Work. Doc. Ser.* 118. Available at; http://search.ebscohost.com/login.aspx?direct=true&db=lah&AN=20043182451&site=ehost-live (accessed October 25, 2019).

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., et al. (2014). Design and characterization of a 52K SNP chip for goats. *PloS One* 9, e86227. doi:10.1371/journal.pone.0086227

Visser, C., Hefer, C. A., van Marle-Köster, E. V., and Kotze, A. (2004). Genetic variation of three commercial and three indigenous goat populations in South Africa. *S A. J. Anim. Sci.* 34, 24–27.

Visser, C., Lashmar, S. F., Van Marle-Köster, E., Poli, M. A., and Allain, D. (2016). Genetic diversity and population structure in South African, French and argentinian Angora goats from genome-wide SNP data. *PLoS One* 11–e0154353. doi:10.1371/journal.pone.0154353

Webb, E. C., and Mamabolo, M. J. (2004). Production and reproduction characteristics of South African indigenous goats in communal farming systems. *Sa. J. Anim. Sci.* 34 (1), 236–239.

Xu, L., Zhao, G., Yang, L., Zhu, B., Chen, Y., Zhang, L., et al. (2019). Genomic patterns of homozygosity in Chinese local cattle. *Sci. Rep.* 9 (1), 16977. doi:10.1038/s41598-019-53274-3

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S., and Sahana, G. (2015). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16, 542. doi:10.1186/s12864-015-1715-x

Zhao, F., Wang, G., Zeng, T., Wei, C., Zhang, L., Wang, H., et al. (2014). Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations. *Livest. Sci.* 170, 22–29. doi:10.1016/j.livsci.2014.10.015

Zvinorova, P. I. (2017). *A genome-wide association study on mechanisms underlying genetic resistance to gastrointestinal parasites in goats*. Ph.D. thesis. Stellenbosch.

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fvets.2022.976603

OPEN ACCESS

EDITED BY
Natalia A. Zinovieva,
L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

REVIEWED BY
Xiao Li,
Northwest A&F University, China
Hui Li,
Guangxi University, China

*CORRESPONDENCE
Xiangyang Miao
miaoxy32@163.com;
mxy32@sohu.com

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

RECEIVED 23 June 2022
ACCEPTED 01 August 2022
PUBLISHED 25 August 2022

CITATION
Feng H, Liu T, Yousuf S, Zhang X, Huang W, Li A, Xie L and Miao X (2022) Identification of potential miRNA-mRNA regulatory network and the key miRNAs in intramuscular and subcutaneous adipose.
Front. Vet. Sci. 9:976603.
doi: 10.3389/fvets.2022.976603

Identification of potential miRNA-mRNA regulatory network and the key miRNAs in intramuscular and subcutaneous adipose

Hui Feng, Tianyi Liu, Salsabeel Yousuf, Xiuxiu Zhang, Wanlong Huang, Ai Li, Lingli Xie and Xiangyang Miao*

State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

Intramuscular fat (IMF) is an important indicator for evaluating meat quality. Breeds with high IMF content are often accompanied by high subcutaneous fat (SCF), severely affecting the meat rate of pigs. Studying the mechanisms of miRNAs related to lipogenesis and lipid metabolism has important implications for pig breeding. We constructed two small RNA libraries from intramuscular and subcutaneous fat to evaluate the patterns of lipogenesis in Laiwu pig, a Chinese breed. A total of 286 differentially expressed miRNAs (DEmiRNAs), including 193 known miRNA and 93 novel miRNAs, were identified from two types of adipose. GO and KEGG enrichment analysis for DEmiRNAs showed that their target genes involved in many adipogenesis and lipid metabolism biological processes and signaling pathways, such as Wnt signaling pathway,MAPK signaling pathway, Hippo signaling pathway, PI3K-Akt signaling pathway, Melanogenesis, Signaling pathways regulating pluripotency of stem cells and so on. Then, we constructed a miRNA-mRNA interaction network to find out which miRNAs were the key miRNAs of regulation in Wnt signaling pathway. In this pathway, miR-331-3p, miR-339-5p, miR-874 and novel346_mature target PPARD, WNT10B, RSPO3, WNT2B. This study provides a theoretical basis for further understanding the post-transcriptional regulation mechanism of meat quality formation and predicting and treating diseases caused by ectopic fat.

KEYWORDS

miRNA, intramuscular fat, subcutaneous fat, pig, high-throughput sequencing

Introduction

Intramuscular fat (IMF) is one of the important factors that affect the quality of pork, however, subcutaneous fat (SCF) is considered as wastes (1). Thus, exploring how to optimize fat deposition is necessary. Besides, excessive fat deposition can lead to many diseases. Previous studies have found that SCF tissue is a relatively safe fat depot, compared with ectopic fat such as IMF and visceral fat. Increased ectopic fat deposition

leads to decreased tissue function and even disease (2, 3). For instance, excessive IMF deposition cause skeletal muscle aging (4), decreasing muscle strength (5), systemic insulin resistance (6) and increasing the risk of multiple diseases, such as COVID-19 (7), rheumatoid joints (8), diabetes (9). Pigs are ideal models for exploring human diseases, with similar physiological structures to human. Understanding mechanism of adipose deposition in different tissues is a cornerstone of improving the quality of pork and exploring how to cure diseases related to ectopic fat deposition.

There is a kind of special cell called fibro-adipogenic progenitors (FAPs), with the capability of adipogenic and fibrogenic differentiation, which is the main source of IMF. Platelet-derived growth factor receptor α (also named CD140a)-positive (PDGFRA+) is the marker of FAPs (1). Many factors can affect the fat-forming ability of FAPs, such as species, FAPs location, muscle damage, microenvironment, etc. More than 90% of FAPs in mice have the capability of adipogenic differentiation, and the frequency will decrease when muscles are damaged (10), however, only 30% of FAPs in humans have fat-forming ability (11). Although both IMF and SCF belong to white fat, compared with SCF, IMF has a weaker ability to form and store lipid droplets, develops later, and has lower expression levels of genes related to adipogenic differentiation and lipid metabolism (1).

In mammals, miRNAs act within Argonaute proteins to guide repression of mRNA (12). miRNAs are regarded as biomarkers and potential therapeutic targets in many diseases (13, 14). The study of miRNA key shear enzymes shows that miRNA plays a key role in the proliferation and differentiation of adipose tissue. Drosha and Dicer in mesenchymal stem cells inhibited adipocyte differentiation, and Drosha in 3T3-L1 cells inhibited adipogenesis (15, 16). There was a negative correlation between adipogenesis and miRNA production (17). Dicer enzyme is down-regulated in adipose tissue, resulting in reduced miRNA treatment (18). Adipocyte-specific knockout of Dicer in mice resulted in severe depletion of white adipose tissue and decreased expression of fat-related genes (19). MiRNAs that promote fat differentiation include miR-21, miR-133, mir-375, and miR-143. For example, miR-21 plays a key role in obesity treatment and lipid metabolism regulation by regulating thermogenesis, white adipocyte browning, VEGF signal transduction, apoptosis, and adipogenesis biological processes and related genes (20). MiR-133 promotes the differentiation of myogenic and white adipose precursor cells into brown adipocytes (21, 22). Overexpression of miR-375 enhanced 3T3-L1 adipocyte differentiation, increased mRNA levels of induced adipogenesis, and induction of adipocyte fatty acid-binding protein (aP2) and triglyceride (TG) accumulation (23). MiR-378 specifically increased C / EBP in adipocytes α and C / EBP β transcriptional activity of gene promoter (24). Overexpression of miR-23a-5p, miR-193a-5p and miR-193b-5p enhances precursor adipocyte differentiation and adipogenesis by targeting IGF2 (25). MiR-181d down regulated ADAMTS1 and promotes adipogenesis (26). Different from the above miRNAs, miRNAs such as let-7, miR-499, miR-206, and miR-30 inhibit the proliferation and differentiation of adipocytes. In the process of adipogenesis, the expression level of let-7 increased significantly. Constructed 3T3-L1 cells within overexpressing let-7 *in vitro*, and found that let-7 prevented cloning and prolonged cell cycle (27). MiR-499 inhibits adipogenic differentiation of skeletal muscle satellite cells by targeting the expression of adipogenesis marker gene PRDM16 (28). MiR-206 inhibits Runx1 translation and inhibits the adipogenic differentiation potential of FAPs (29). MiR-26 can inhibit FBXL19, adipogenesis and precursor adipocyte differentiation (30). These results show that miRNAs play an important regulatory role in adipogenesis and metabolism, and can be used as a therapeutic target for metabolic diseases.

At present, the living standard has been greatly improved, and people's requirements for the quality and quantity of meat have also increased. Therefore, studying the molecular mechanism of IMF and SCF deposition can improve the quality of meat products to meet the needs of consumers. This study aims to explore the mechanism of adipogenesis and lipid metabolism between different adipose tissues, by constructing miRNA tissue-specific expression patterns of IMF and SCF in Laiwu pigs using RNA-Seq. Therefore, studying the mechanism of fat deposition in different parts can not only improve meat quality and promote the development of animal husbandry but also prevent or treat a series of diseases affecting human health caused by excessive fat deposition. In this study, the intramuscular and subcutaneous adipose tissues of Laiwu pigs were selected as experimental materials. The gene expression profiles of different adipose tissues were analyzed by transcriptome sequencing and bioinformatics methods, and the key candidate genes related to lipid metabolism and adipogenic differentiation were screened and identified to explore the molecular mechanism related to optimizing fat deposition.

Abbreviations: RNA-seq, RNA sequencing technology; GO, Gene ontology; KEGG, the Kyoto Encyclopedia of Genes and Genomes; IMF, Intramuscular fat; SCF, subcutaneous fat; DEmiRNAs, differentially expressed miRNAs; DEgenes, differentially expressed target genes of miRNA; PCR, polymerase chain reaction; TPM, transcript per million reads; U, Uracil; A, Adenine.

Materials and methods

Total RNA extraction and quality control

The experimental animals used in this study are three Chinese breeds known as Laiwu pigs, which were 180 days, healthy, and of similar weight. All of the female

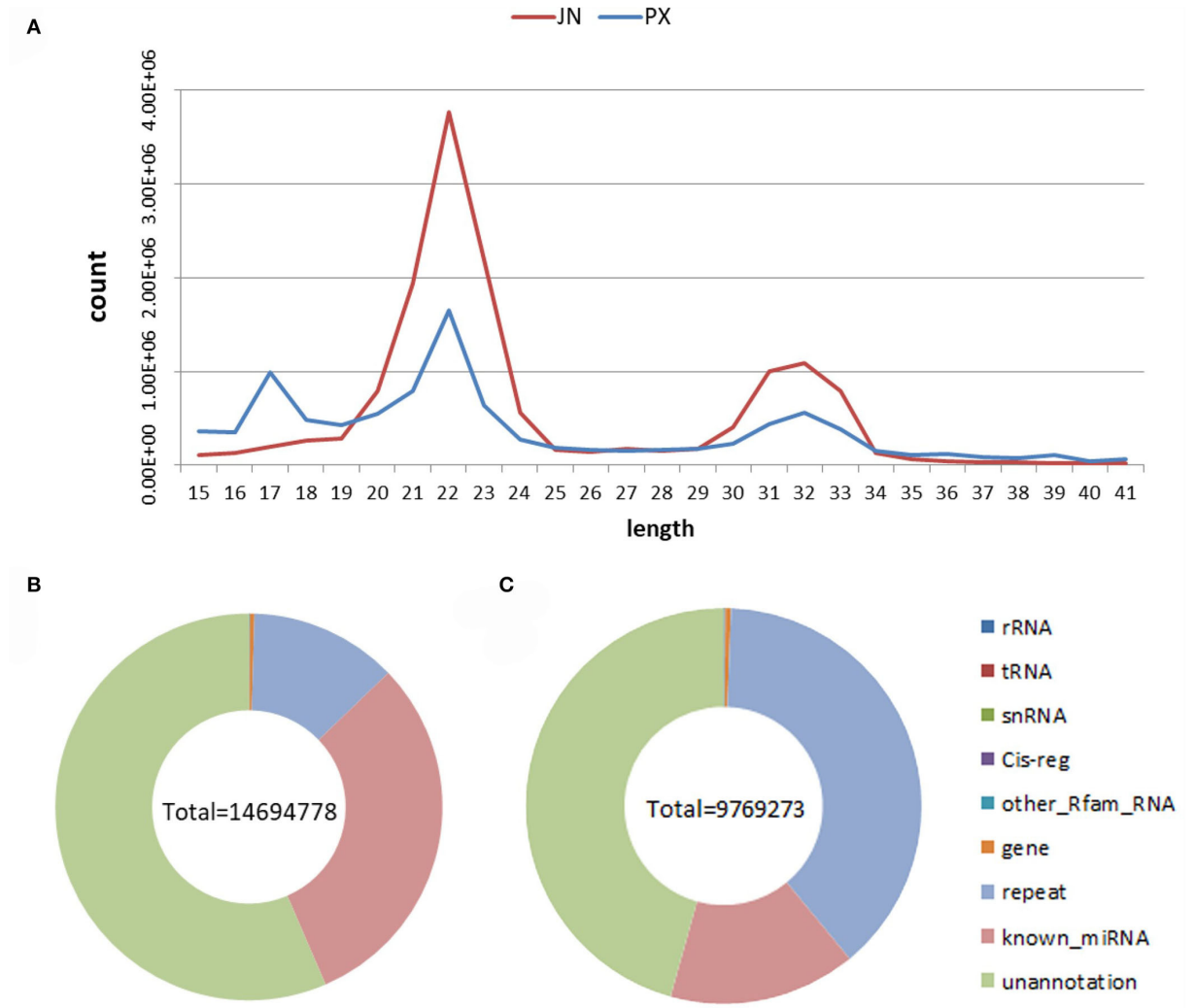


FIGURE 1
(A) Small RNA length distribution diagram. **(B)** L_JN; **(C)** L_PX. The type of total small RNA.

pigs were raised in the same environment with standard conditions for temperature, humidity, and ventilation in Daqian farm, Shandong province. After slaughtering pigs by bloodletting, SCF were rapidly stripped from longissimus dorsi and IMF were stripped from fat located between muscle fascia in the same tissue, respectively put them into the cryopreservation tubes marked L_ PX_ 1, 2, 3 and L_ JN_ 1, 2, 3 in advance. And frozen in liquid nitrogen, and then stored at −80°C. Total RNA was extracted by mirVana™ miRNA ISOlation Kit, according to the manufacturer's instructions. Using NanoDrop2000 spectrophotometer and Aglient Bioanalyzer 2100 to measure OD260 nm/OD280nm, RNA concentration and integrity, ensure total RNA can be used for subsequent analysis.

Small RNA library construction and sequencing

Two small RNA libraries were constructed using IMF (L_JN) and SCF (L_PX) tissues from Laiwu pigs, using Illumina TruSeq Small RNA reagent test kit. Equal amounts of total RNA from two adipose tissues of three pigs were mixed and the mixed RNA was used to construct a small RNA library. This process includes adding adapters to total RNA, reverse transcription of cDNA, PCR amplification, and electrophoresis to separate small RNA, and recovery and purification of cDNA. Using Agilent 2,100 assessed the quality of the cDNA library. Then, sequence the quality cDNA library by Illumina HiSeqTM 2,500 sequencing platform.

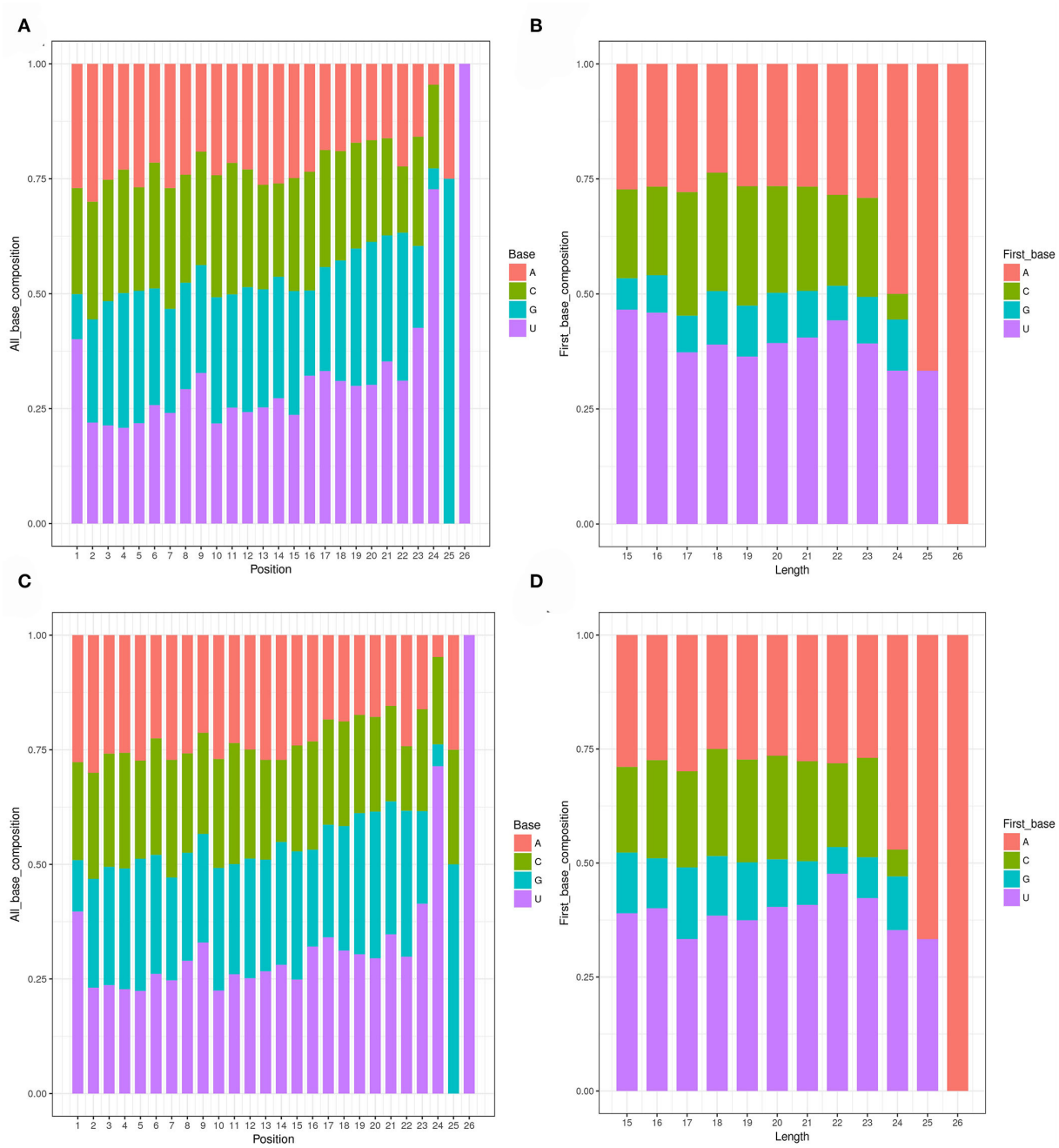


FIGURE 2
Structural characteristics of miRNAs. **(A)** Distribution map of base preference at different miRNA site in L_JN. **(B)** First base preference distribution map of miRNAs with lengths in L_JN. **(C)** Distribution map of base preference at different miRNA site in L_PX. **(D)** First base preference distribution map of miRNAs with lengths in L_PX.

Raw data quality control and small RNA annotation

It is necessary that remove joints and low-quality fragments in raw data to get clean reads for subsequent analysis. Reads with lengths of 15–41nt and not including N-base are retained. There are many kinds of small RNAs, including miRNA, tRNA (tiRNA, TRFs), rRNA, piRNA, snRNA, etc. In order to classify and annotate the small RNA in the sequencing results, clean reads were compared with the reference genome, Rfam database(31), cDNA sequence, species repeat database (32), and miRBase database (33) turn. Bowtie software was used to conduct an

error-free alignment with the miRNA mature sequence in the miRBase database, and the sequence on the alignment was considered to be a known miRNA.

Novel miRNA prediction

After the identification of known miRNAs, some sequences in clean reads that have not been annotated to miRBase may be novel miRNAs. Using miRDeep2 software to compare the FASTA file to the reference genome, the results retain the fully matched reads with a length of 18–25nt. By predicting the miRNA precursor and comparing it with the precursor information in the miRBase database, the reads on the comparison can be considered as a candidate new miRNA for quantitative analysis (33). Prediction of candidate new miRNA precursor structure by RNAfold software (34). The sequences which have a hairpin structure were considered to be novel miRNAs. Extract the mature and star sequences of predicted novel miRNAs, and count the number of novel miRNAs.

Differentially expressed miRNA identification and target prediction

Use the DESeq2 package (35), with |log2foldchange|≥1 and $P \leq 0.05$ as the screening conditions, to get differentially expressed miRNAs between IMF and SCF tissues. TPM (transcript per million reads) algorithm was used to calculate the expression level of the identified mature known miRNAs and novel miRNAs.

Use miRanda to predict target genes of DEmiRNAs according to score≥150 and energy≤-30 kcal/mol, and retain the miRNA-mRNA pairs which were differentially expressed in IMF and SCF. The predicted results of the target relationship between miRNA and mRNA were input into Cytoscape software to construct the miRNA-mRNA regulatory network.

Gene ontology and kyoto encyclopedia of genes and genomes pathway annotation of miRNA target genes, and miRNA-protein interaction analysis

Use clusterProfiler (36) package of R Studio to annotate target genes function, including GO enrichment and KEGG pathway analysis. GO terms and KEGG pathways were considered significantly enriched when $P \leq 0.05$. The pathway related to fat metabolism and the genes involved in this pathway were screened, and the protein-protein interaction network in this pathway was predicted by STRING database.

Q-PCR quantitative verification

The expression level of 6 miRNAs, ssc-miR-296-5p, ssc-miR-361-3p, ssc-miR-331-3p, ssc-miR-328, sscmiR-125b, and a predicted novel miRNA novel407_mature, and 2 mRNAs, CD180 and ELF4, were randomly selected and verified by qRT-PCR. Briefly, total RNA was extracted from the IMF and SCF tissue samples with total RNA Extraction Kit (DNase I) kit (genepool, cat# gpq1801). Take 3 μL RNA was electrophoretic with 1% agarose gel to detect the integrity of RNA. miRNA cDNA Synthesis Kit (GenePool,Cat# GPQ1804) was used for reverse transcription, and BIOER LineGene 9600Plus fluorescence quantitative instrument was used to perform the PCR analysis of the miRNAs. Repeated all of the reactions in triplicate and calculate the relative quantitative expression levels by $2^{-\Delta\Delta Ct}$ method. Incubated control miRNA gene U6 and mRNA gene GAPDH were used for the calculations of the relative expression level of the genes. Use the *T*-test to determine the significantly different expression levels.

Results

Evaluation of RNA-sequencing

We identified 15.79 and 12.42 million raw reads, respectively from IMF and SCF miRNA libraries and obtained 14.69 and 9.77 million clean reads after removing joints and low-quality fragments. According to the statistics of the clean reads sequence length, it is found that the peak of the length distribution of these small RNA sequences is 21–25 nt, and the other peak is about 31–33 nt, which may be the distribution of piRNA, tRNA or rRNA. Of clean reads, 85.45% (12.56/14.69 million reads, IMF) and 90.79% (8.87/9.77 million reads, SCF) were mapped to the reference genome, indicating sequencing quality can be used for subsequent analysis. Rfam comparative filtering, transcript filtering, repeat comparative filtering, and miRBase database comparative annotation were performed on clean reads to obtain small RNA annotation information. As shown in Figure 1, 30.70 and 15.29% of clean reads were annotated on miRBase, and 56.49 and 45.68% of clean reads were not annotated on any database. Reads not annotated to the database were used for novel miRNA prediction.

Identification and analysis of miRNA

After removing rRNA, tRNA, snoRNA, snRNA, and repeat sequences, 764 and 627 miRNAs were identified in IMF and SCF tissues, respectively by comparing with the miRBase database. The first base of mature miRNA

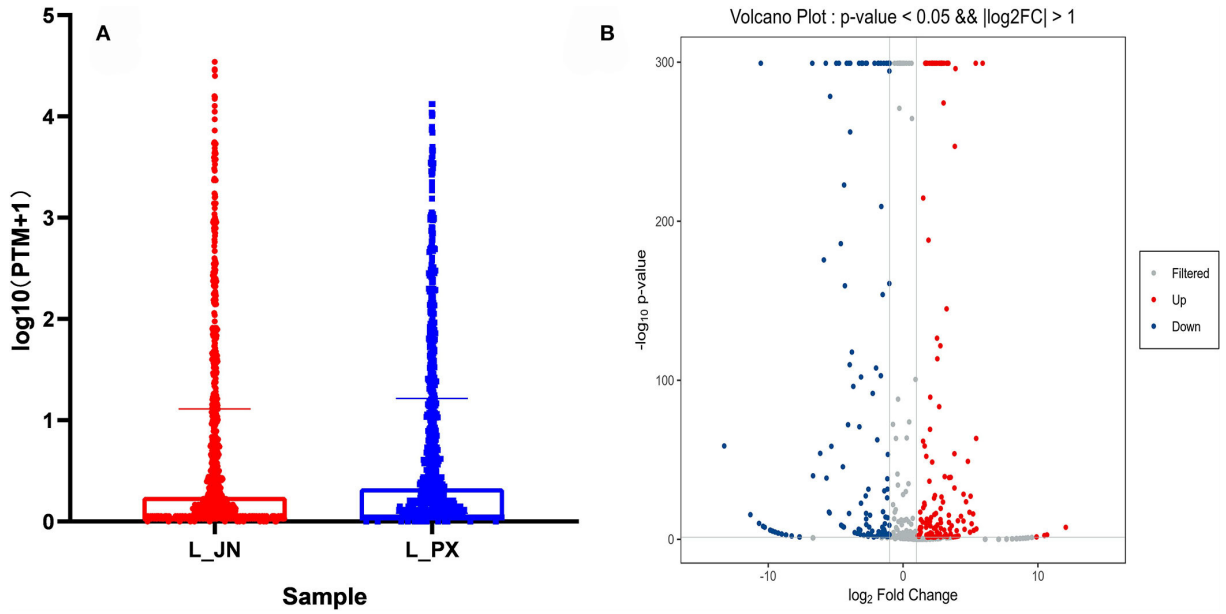


FIGURE 3
Differential expression of miRNAs. **(A)** Horizontal box diagram of miRNA expression. **(B)** Volcano map of differentially expressed miRNAs. The scatter in the figure represents the miRNAs, the black dots indicate the miRNAs with no significant differences, the red dots indicating a significant up-regulation of the miRNA, and the blue dots represent a significant down-regulation of the miRNA.

has a strong preference for base U due to the specificity of the enzyme digestion site (Figures 2A,C). The tenth base is generally miRNA, which is the active site of the target gene and has a strong preference for base A (Figures 2B,D). In this study, the results of the base preference of known miRNA indicated that the sequencing results are reliable.

On the basis of their shared sequence similarity, all miRNAs identified in our study were divided into 183 families. The let-7 family contains the most kinds of miRNAs, including known miRNA such as ssc-let-7a, ssc-let-7c, ssc-let-7d-3p, ssc-mir-98 et al., and novel miRNAs such as novel403_ mature. Let-7 family is a miRNA that has been proved to play a regulatory role in adipogenesis and lipid metabolism (37), indicating that there are differences in fat deposition between the IMF and SCF tissues. In addition, the miR-10 family has the highest expression level in two adipose tissues and can regulate apoptosis and cell differentiation.

Differential expression of miRNA

According to the diagram of miRNA expression level, it can be seen that the miRNA expression level in IMF is higher (Figure 3A). With |log2foldchange|≥1 and $P \leq 0.05$ as the screening conditions, a total of 286 differentially expressed miRNAs (DEmiRNAs) were identified in two adipose tissues, including 193 known miRNAs and 93 novel miRNAs. Among DEmiRNAs, 146 were up-regulated and 140 were down-regulated (Figure 3B). According to the order of expression, the top 10 miRNAs were taken from IMF and SCF, respectively (Supplementary Table 2). Among them, miR-26, let-7, miR-27, and miR-1 families play important roles in adipogenic differentiation and lipid metabolism (30, 37–39).

miRNA target genes prediction and functional enrichment analysis

The target genes of DEmiRNAs were predicted by miRanda software, and the differentially expressed genes (DEgenes) in IMF and SCF were screened. 93 DEmiRNAs and 616 DEgenes were involved in the regulatory network. Ssc-miR-370, ssc-miR-331-3p, ssc-miR-330, ssc-miR-361-3p, ssc-miR-744, ssc-miR-874, ssc-miR-339-5p, ssc-miR-328, and novel407_mature have more target genes (Figure 4). In order to explore the function of miRNAs, we followed the functional enrichment analysis of the target mRNAs of DEmiRNAs. The results of GO annotations show that target genes enriched in 294 GO terms, including 258 biological processes, 8 cell

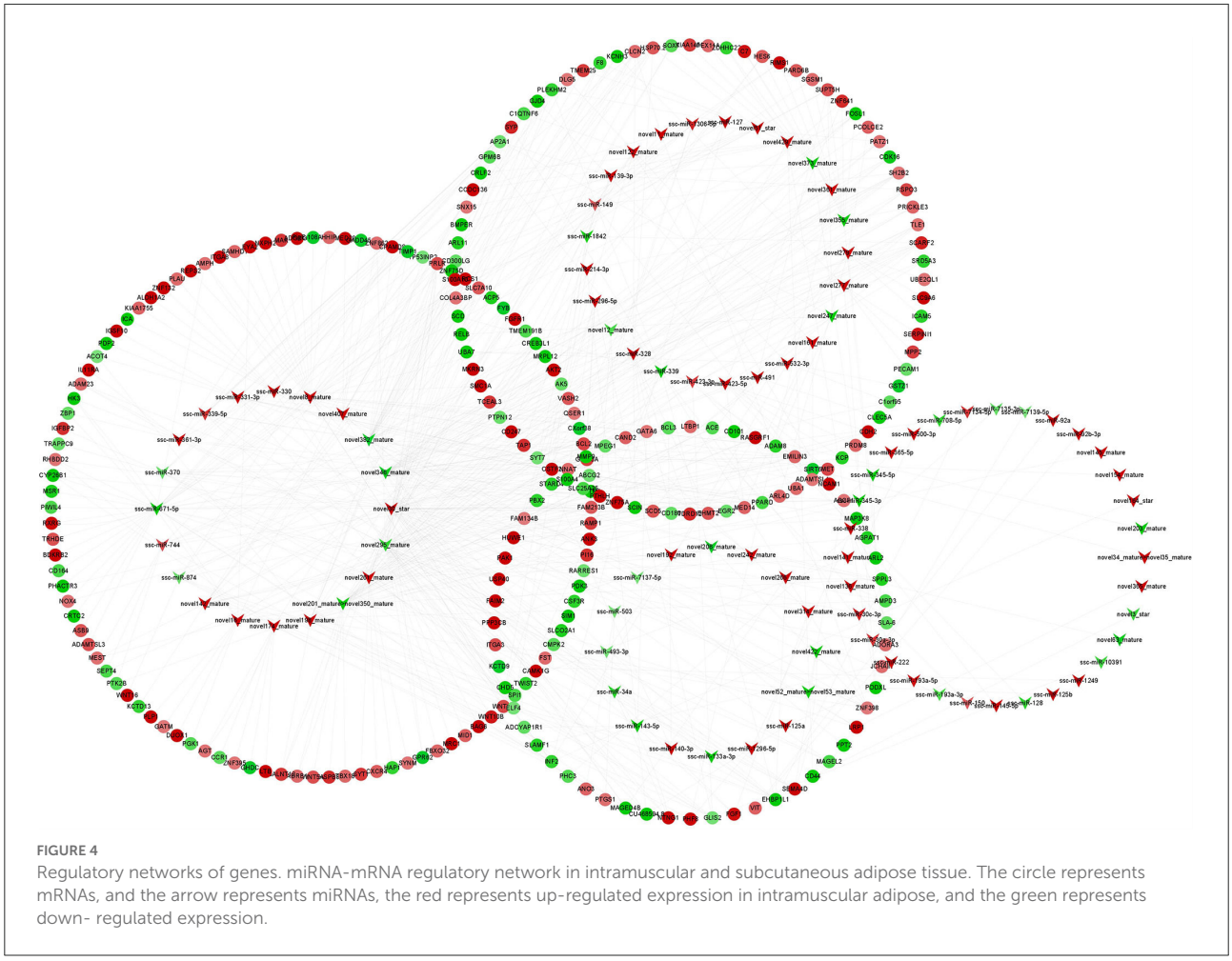


FIGURE 4
Regulatory networks of genes. miRNA-mRNA regulatory network in intramuscular and subcutaneous adipose tissue. The circle represents mRNAs, and the arrow represents miRNAs, the red represents up-regulated expression in intramuscular adipose, and the green represents down- regulated expression.

compositions, and 28 molecular functions. In the biological process category, target genes were involved in terms related to the regulation of cell growth and development, such as regulation of multicellular organismal development, regulation of multicellular organismal development, positive regulation of signal transduction, and regulation of fibroblast growth factor receptor signaling pathway. In the cell compositions category, target genes were involved in terms related to the structure of cell membranes, such as membrane raft, membrane microdomain, synaptic vesicle membrane, and exocytic vesicle membrane. In the molecular functions category, target genes were involved in terms related to the regulation of enzymes and transcription factor activity, such as DNA-binding transcription factor activity, RNA polymerase II-specific, cytokine receptor activity, transcription regulatory region nucleic acid binding, protein phosphatase binding, transition metal ion binding (Figure 5A). Furthermore, the predicted target genes were annotated in KEGG pathways to identify potential pathways that may be regulated by the differentially expressed miRNAs. The results indicated that target genes were enriched in 53 pathways like Wnt signaling pathway, MAPK signaling pathway, Melanoma, C-type lectin receptor signaling pathway, C-type lectin receptor signaling pathway, and PI3K-Akt signaling pathway (Figure 5B).

The relationship between miRNAs and target genes in the Wnt signaling pathway

Wnt signaling pathway plays an important role in the process of fat deposition. We found that miRNA target genes are partially enriched in this signaling pathway, including TLE1, PRICKLE3, RSPO3, FOSL1, PPARD, WNT16, WNT5A, PPP3CB, WNT10B, and WNT2B. The query results based on string database show that there is an interactive relationship between RSPO3, WNT16, WNT5A, WNT10B, and WNT2B (Figure 6A). These target genes are regulated by 17 miRNAs regulated, such as miR-331-3p, miR-339-5p, miR-874, novel346_mature, et al (Figure 6B). These DEmiRNAs and

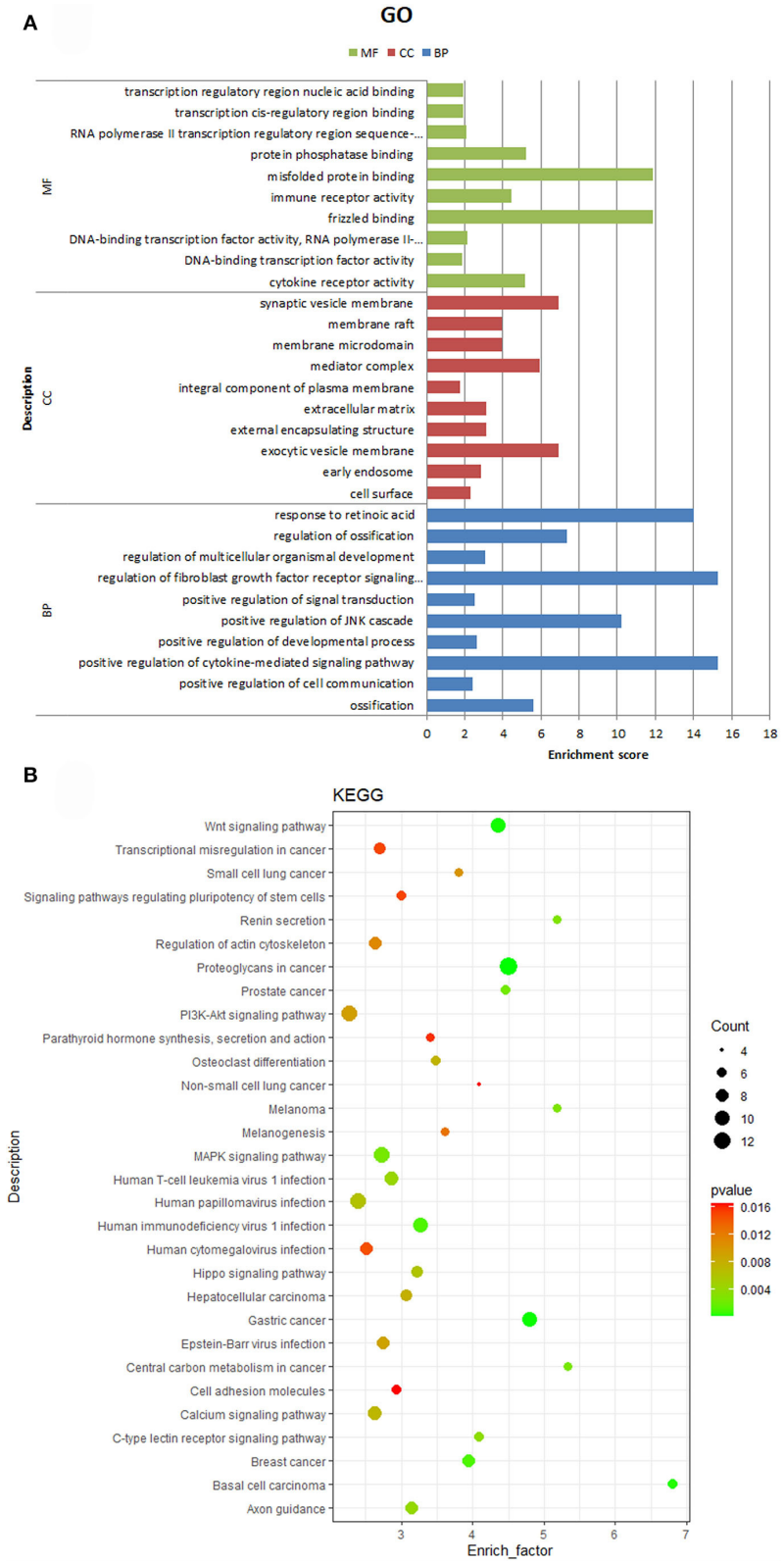


FIGURE 5
(A) GO enrichment analysis of target genes. The X-axis is the enrichment factor, and the Y-axis is GO term. **(B)** KEGG pathway analysis of the target genes. The X-axis is the enrichment factor, and the Y-axis is KEGG pathway.

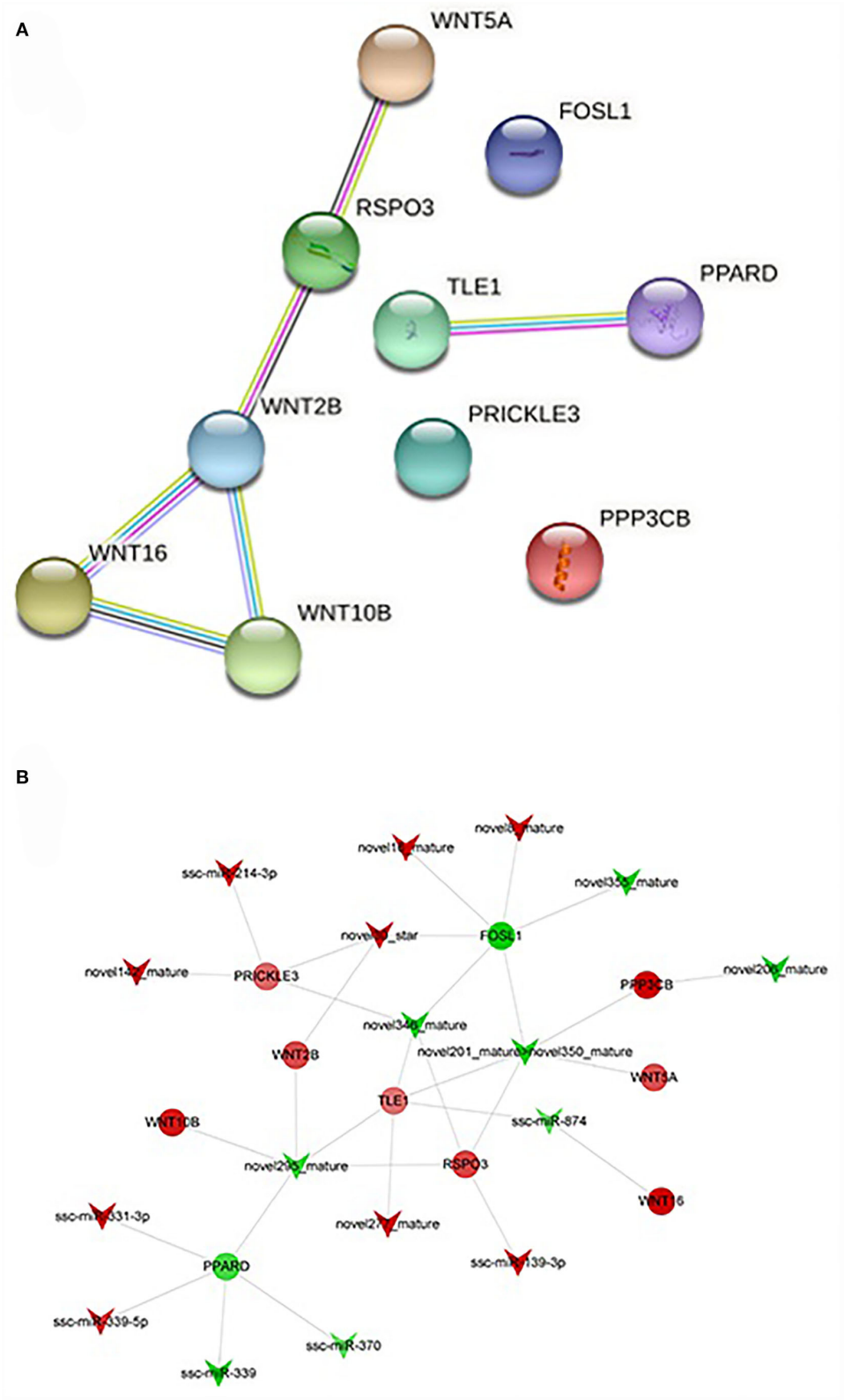


FIGURE 6
(A) Protein-protein interaction network of differentially expressed gengs related to Wnt signaling pathway. **(B)** The targeting relationship of genes in regulation of Wnt signaling pathway.

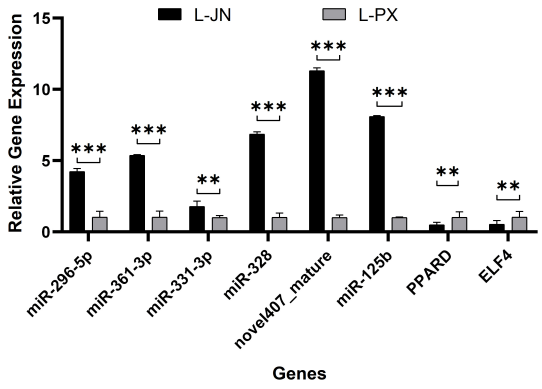


FIGURE 7
qRT-PCR verification of differentially expressed genes. The differential expression of genes between intramuscular and subcutaneous adipose tissue in Laiwu pig was verified qRT-PCR. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

DEgenes may cause the difference in fat deposition between IMF and SCF.

qRT-PCR results

The expression level of 6 differentially expressed miRNAs and 2 target genes were selected and verified by qRT-PCR (Figure 7). The results showed that 6 miRNAs (mir-296-5p, mir-361-3p, mir-331-3p, mir-328, novel407_nature, miR-125b) were downregulated, which was consistent with the results of RNA-Seq. In addition, we verified the expression of target genes (PPARD, ELF4) and proved that the sequencing results were reliable.

Discussion

Laiwu pig is a typical local pig breed in China. It has good IMF deposition and delicious meat. We previously identified the differential expression of miRNA in fat among different breeds of pigs, cattle, and sheep, and constructed miRNA-mRNA networks (40–42). In these results, not only miRNA regulatory networks related to fat regulation were found, but also many new miRNAs were identified. It shows that the lipid deposition difference in adipose tissue between different varieties is related to the expression of miRNA, so we regulate miRNA in different parts of adipose tissue in the same species. Exploring the differential mechanism of IMF and SCF deposition and identifying the miRNA expression profile in the two adipose tissues help to improve pork quality and understanding of diseases induced by ectopic fat deposition. In this study, 286 differentially expressed miRNAs (DEmiRNAs) were identified in IMF and SCF, including 193 known miRNAs and 93 novel miRNAs. The functional annotation analysis of DEmiRNAs showed that the target gene played an important role in the process of fat differentiation, proliferation, and lipid metabolism. Among these DEmiRNAs, miR-26a has the highest expression level in IMF, which may be the main miRNA leading to the difference in fat deposition between IMF and SCF.

MiR-26a plays different roles in different stages of adipocytes, especially in the early stages of adipogenic differentiation. Overexpression of miR-26a can reduce the expression of PTEN, STK11, and other genes, inhibit PI3K-Akt signaling pathway, and promote adipocyte differentiation and lipid accumulation in the early stage of 3T3-L1 cell differentiation (43). In the early stage of adipogenic differentiation, it can also induce the expression of UCP1 and mitochondrial metabolism-related genes, and promote the metabolism of brown fat (44). Insulin can promote the expression of miR-26a, inhibit CDK5 and FOXC2, and promote the differentiation of human adipose stem cells (45). Knockout of miR-26a in mice resulted in the rapid deposition of adipose tissue in normally fed adult animals; on the contrary, overexpression of miR-26a can inhibit FBXL19 expression and adipogenesis (30). The specific expression of miR-26a in fat can moderately reduce the levels of visceral fat and lipid. In this experiment, miR-26a was highly expressed in IMF and inhibited the accumulation of lipids in IMF cells.

In this study, we constructed miRNA-mRNA regulatory networks and found that miR-331-3p, miR-339-5p, miR-874, novel346_mature et al banded with key target genes and regarded them as key miRNA in the Wnt signaling pathway. At different stages of precursor adipocyte differentiation, the expression level of miR-331-3p will increase first and then decrease. At the same time, the expression of the PPAR family will also show the same trend, and the expression level of cell differentiation-related genes will decrease, indicating that miR-331-3p can promote precursor adipocyte differentiation and inhibit proliferation (46). Fat droplets in adipocytes overexpressing miR-331-3p are larger, inhibit the expression of fatty acid metabolism-related gene DLST, and promote fatty acid synthesis (46). Other studies showed that the expression count of miR-331-3p in skeletal muscle cells was 48 (47), which was significantly lower than that in intramuscular fat, indicating that the regulation mode of miR-331-3p was also different in different cell types at the same site. Regulating the expression of ssc-miR-331-3p at different stages of adipocyte development can promote IMF deposition without increasing SCF. In our previous study, miR-874 was downregulated in bovine intramuscular fat (compared with Holstein cattle), indicating that miR-874 is a potential key miRNA for intramuscular fat regulation (40). Similarly, in the experiment comparing subcutaneous fat deposition in Chinese Huainan pigs (HN,

the fat type) and Western commercial Duroc × (Landrace × Yorkshire) (DLY, the thin type) pigs, miR-874 is also at the center of regulating fat deposition network (48). These experimental results are consistent with the results of this study, indicating that miR-874 is a candidate regulator of fat deposition.

To explore the potential regulatory network of miR-331-3p in IMF, predict the target gene of miR-331-3p. Based on the differentially expressed transcriptome data of IMF and SCF in Laiwu pigs, the differential target genes of miR-331-3p in the two adipose tissues were screened for functional enrichment analysis. The results showed that miR-331-3p could significantly down-regulate SYT7, PBX2, PPARD, PDP2, ACOT4, EHBP1L1, ELF4, EGR2 et al (Figure 4). Among them, PPARD was confirmed to be a regulatory gene involved in adipocyte proliferation, differentiation, and lipid metabolism. Increasing PPARD expression in mice reduces lipids in adipose tissue and serum and increases the expression of lipid metabolism-related genes and adipogenesis and storage genes (49). PPARD agonists can promote cholesterol accumulation in macrophages, reduce obesity (50), increase fatty acid oxidation and oxidative phosphorylation in skeletal muscle, and increase insulin sensitivity (51). Therefore, PPARD agonists have been used as drug targets for the treatment of obesity and related diseases. Nucleotide polymorphism is the main factor affecting PPARD, and rs2016520 polymorphism has been proved to be closely related to human fat metabolism (52, 53). Overexpression of mutant PPARD (serine 112 mutated to alanine) and wild-type PPARD in mouse myocytes can promote adipogenesis and inhibit the expression of the myogenic marker gene MRF4, and mutant PPARD has a greater ability to promote adipogenesis (54). In the presence of PPARG ligand, porcine PPARD was overexpressed in myogenic cell C2C12, which promoted the expression of endogenous PPARG, adipogenesis, and adipogenic genes, and reduced the expression levels of myogenic protein and MRF4 (55). PPARD is involved in the biological process of adipogenesis and metabolism, which is a complex process. In the process of intramuscular adipogenesis, PPARD may not be directly involved and needs to be co-expressed with PPARG to promote the differentiation of myoblasts into adipocytes, but the specific mechanism is not clear and needs to be further discussed.

A number of novel miRNAs were identified in this study, and the target genes of novel miRNAs were significantly enriched in fat metabolism, immune response, fat metabolism-related diseases, and other signaling pathways. Among them, novel346_mature, novel295_mature, novel206_mature, novel355_mature et.al were involved in the Wnt signal pathway through banding with WNT2B, RSPO3, WNT5A, FOSL1 et.al. Wnt family genes are specifically expressed in bovine adipocytes, in which WNT2B regulates fat differentiation by activating FZD5 (56). RSPO3 promotes osteogenic differentiation of adipose stem cells through the LGR4-ERK signaling pathway (57). Subsequently, it was found that RSPO3 expression was increased in SCF, which inhibited adipogenesis and increased adipocyte apoptosis by regulating Wnt signal transduction (58). WNT5A is a fat factor that regulates the nonstandard Wnt signaling pathway and inhibits the differentiation of FAPs into adipocytes (59). FOSL1 plays an important role in suppressing adipogenesis by regulating C/EBPα, and FOSL1 transgenic mice have reduced fat deposition and the expression of fat marker genes (60). These results are consistent with the present study showing that these target genes play critical roles in intramuscular and subcutaneous fat deposition and lipid metabolism, resulting in significant differences in fat deposition between the two types of adipose tissue. In this study, we found that novel miRNAs regulate the expression levels of target genes through post-transcriptional regulation, and regulate the deposition of different fats by regulating the Wnt signaling pathway.

Conclusion

We employed RNA-seq technology and bioinformatics tools to identify miRNAs in intramuscular and subcutaneous adipose tissue of Laiwu pig. Totally 286 differentially expressed miRNAs (193 known miRNAs and 93 novel miRNAs) were identified in the intramuscular and subcutaneous adipose tissue, including 140 down-regulated miRNAs and 146 up-regulated miRNAs. The GO annotation and KEGG enrichment analysis of target genes revealed the miRNA target genes might be involved in the Wnt signaling pathway. In the Wnt signaling pathway, 17 miRNAs (including miR-331-3p, miR-339-5p, miR-874, and novel346_mature et al) target PPARD, WNT10B, RSPO3, WNT2B, and other genes, respectively, and participate in the deposition of intramuscular fat and subcutaneous adipose tissue. In conclusion, all the results provide a theoretical basis for further revealing the deposition mechanism of intramuscular fat and subcutaneous adipose tissue and cultivating lean pigs with high meat quality, and lay a foundation for promoting the development of animal husbandry.

Data availability statement

RNA-Seq data have been deposited in the NCBI Sequence Read Archive (SRA) database with accession numbers PRJNA870052 and PRJNA870173.

Ethics statement

The animal study was reviewed and approved by Experimental Animal Welfare and Ethical of Institute of Animal Science, Chinese Academy of Agricultural Sciences.

Author contributions

XM conceived, designed, performed the experiments, and wrote the paper. HF performed the experiment, data analysis, and wrote the paper. TL, SY, XZ, WH, AL, and LX performed the experiments. All authors have read and approved the final manuscript.

Funding

This work was supported by a grant from the National Natural Science Foundation of China (No. 31970541), the Major Science and Technology Project of New Variety Breeding of Genetically Modified Organisms (Nos. 2009ZX08008-004B and 2008ZX08008–003), the Agricultural Science and Technology Innovation Program (NO. ASTIP-IAS05), and the Basic Research Fund for Central Public Research Institutes of CAAS (Y2016JC22, Y2018PT68, 2013ywf-yb-5, and 2013ywf-zd-2).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.976603/full#supplementary-material

SUPPLEMENTARY TABLE 1
Gene and the corresponding primer sequences.

SUPPLEMENTARY TABLE 2
Highly expressed miRNA in IMF and SCF.

References

1. Li X, Fu X, Yang G, Du M. Review: enhancing intramuscular fat development via targeting fibro-adipogenic progenitor cells in meat animals. *Animal.* (2020) 14:312–21. doi: 10.1017/S175173111900209X

2. Smith U, Kahn BB. Adipose tissue regulates insulin sensitivity: role of adipogenesis, de novo lipogenesis and novel lipids. *J Intern Med.* (2016) 280:465–75. doi: 10.1111/joim.12540

3. Belzunce MA, Henckel J, Di Laura A, Hart AJ. Reference values for volume, fat content and shape of the hip abductor muscles in healthy individuals from Dixon MRI. *NMR Biomed.* (2021) 35:e4636. doi: 10.1002/nbm.4636

4. Jiang Z, Marriott K, Maly MR. Impact of inter- and intramuscular fat on muscle architecture and capacity. *Crit Rev Biomed Eng.* (2019) 47:515–33. doi: 10.1615/CritRevBiomedEng.2020031124

5. Akazawa N, Harada K, Okawa N, Tamura K, Moriyama H. Muscle mass and intramuscular fat of the quadriceps are related to muscle strength in non-ambulatory chronic stroke survivors: a cross-sectional study. *PLoS ONE.* (2018) 13:e0201789. doi: 10.1371/journal.pone.0201789

6. Kitessa SM, Abeywardena MY. Lipid-induced insulin resistance in skeletal muscle: the chase for the culprit goes from total intramuscular fat to lipid intermediates, and finally to species of lipid intermediates. *Nutrients.* (2016) 8:466. doi: 10.3390/nu8080466

7. Yang Y, Ding L, Zou X, Shen Y, Hu D, Hu X, et al. Visceral adiposity and high intramuscular fat deposition independently predict critical illness in patients with SARS-CoV-2. *Obesity.* (2020) 28:2040–8. doi: 10.1002/oby.22971

8. Baker JF, Mostoufi-Moab S, Long J, Zemel B, Ibrahim S, Tar atuta E, et al. Intramuscular fat accumulation and associations with body composition, strength, and physical functioning in patients with rheumatoid arthritis. *Arthritis Care Res.* (2018) 70:1727–34. doi: 10.1002/acr.23550

9. Almurdhi MM, Reeves ND, Bowling FL, Boulton AJ, Jeziorska M, Malik RA. Reduced lower-limb muscle strength and volume in patients with type 2 diabetes in relation to neuropathy, intramuscular fat, and vitamin D levels. *Diabetes Care.* (2016) 39:441–7. doi: 10.2337/dc15-0995

10. Joe AW, Yi L, Natarajan A, Le Grand F, So L, Wang J, et al. Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat Cell Biol.* (2010) 12:153–63. doi: 10.1038/ncb2015

11. Collao N, Farup J, De Lisio M. Role of metabolic stress and exercise in regulating fibro/adipogenic progenitors. *Front Cell Dev Biol.* (2020) 8:9. doi: 10.3389/fcell.2020.00009

12. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol.* (2014) 15:509–24. doi: 10.1038/nrm3838

13. Mishra S, Yadav T, Rani V. Exploring miRNA based approaches in cancer diagnostics and therapeutics. *Crit Rev Oncol Hematol.* (2016) 98:12–23. doi: 10.1016/j.critrevonc.2015.10.003

14. Yuntao G, Xiangyang M. [MicroRNAs in the regulation of brown adipocyte differentiation]. *Yi Chuan.* (2015) 37:240–9. doi: 10.16288/j.yczz.14-360

15. Oskowitz AZ, Lu J, Penfornis P, Ylostalo J, McBride J, Flemington EK, et al. Human multipotent stromal cells from bone marrow and microRNA: regulation of differentiation and leukemia inhibitory factor expression. *Proc Natl Acad Sci U S A.* (2008) 105:18372–7. doi: 10.1073/pnas.0809807105

16. Wang Q, Li YC, Wang J, Kong J, Qi Y, Quigg RJ, et al. miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. *Proc Natl Acad Sci U S A.* (2008) 105:2889–94. doi: 10.1073/pnas.0800178105

17. Xie H, Lim B, Lodish HF. MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. *Diabetes.* (2009) 58:1050–7. doi: 10.2337/db08-1299

18. Mori MA, Raghavan P, Thomou T, Boucher J, Robida-Stubbs S, Macotela Y, et al. Role of microRNA processing in adipose tissue in stress defense and longevity. *Cell Metab.* (2012) 16:336–47. doi: 10.1016/j.cmet.2012.07.017

19. Mudhasani R, Puri V, Hoover K, Czech MP, Imbalzano AN, Jones SN. Dicer is required for the formation of white but not brown adipose tissue. *J Cell Physiol.* (2011) 226:1399–406. doi: 10.1002/jcp.22475

20. Lhamyani S, Gentile AM, Giraldez-Perez RM, Feijoo-Cuaresma M, Romero-Zerbo SY, Clemente-Postigo M, et al. miR-21 mimic blocks obesity in mice: a novel therapeutic option. *Mol Ther Nucleic Acids.* (2021) 26:401–16. doi: 10.1016/j.omtn.2021.06.019

21. Trajkovski M, Ahmed K, Esau CC, Stoffel M. MyomiR-133 regulates brown fat differentiation through Prdm16. *Nat Cell Biol.* (2012) 14:1330–5. doi: 10.1038/ncb2612

22. Morozzi G, Beccafico S, Bianchi R, Riuzzi F, Bellezza I, Giambanco I, et al. Oxidative stress-induced S100B accumulation converts myoblasts into brown adipocytes via an NF-kappaB/YY1/miR-133 axis and NF-kappaB/YY1/BMP-7 axis. *Cell Death Differ.* (2017) 24:2077–88. doi: 10.1038/cdd.2017.132

23. Ling HY, Wen GB, Feng SD, Tuo QH, Ou HS, Yao CH, et al. MicroRNA-375 promotes 3T3-L1 adipocyte differentiation through modulation of extracellular signal-regulated kinase signalling. *Clin Exp Pharmacol Physiol.* (2011) 38:239–46. doi: 10.1111/j.1440-1681.2011.05493.x

24. Gerin I, Bommer GT, McCoin CS, Sousa KM, Krishnan V, MacDougald OA. Roles for miRNA-378/378* in adipocyte gene expression and lipogenesis. *Am J Physiol Endocrinol Metab.* (2010) 299:E198–206. doi: 10.1152/ajpendo.00179.2010

25. Mirra P, Desiderio A, Spinelli R, Nigro C, Longo M, Parrillo L, et al. Adipocyte precursor cells from first degree relatives of type 2 diabetic patients feature changes in hsa-mir-23a-5p,−193a-5p, and−193b-5p and insulin-like growth factor 2 expression. *FASEB J.* (2021) 35:e21357. doi: 10.1096/fj.202002156RRR

26. Chen SZ, Ning LF, Xu X, Jiang WY, Xing C, Jia WP, et al. The miR-181d-regulated metalloproteinase adamts1 enzymatically impairs adipogenesis via ECM remodeling. *Cell Death Differ.* (2016) 23:1778–91. doi: 10.1038/cdd.2016.66

27. Sun T, Fu M, Bookout AL, Kliewer SA, Mangelsdorf DJ. MicroRNA let-7 regulates 3T3-L1 adipogenesis. *Mol Endocrinol.* (2009) 23:925–31. doi: 10.1210/me.2008-0298

28. Jiang J, Li P, Ling H, Xu Z, Yi B, Zhu S. MiR-499/PRDM16 axis modulates the adipogenic differentiation of mouse skeletal muscle satellite cells. *Hum Cell.* (2018) 31:282–91. doi: 10.1007/s13577-018-0210-5

29. Wosczyna MN, Perez Carbajal EE, Wagner MW, Paredes S, Konishi CT, Liu L, et al. Targeting microRNA-mediated gene repression limits adipogenic conversion of skeletal muscle mesenchymal stromal cells. *Cell Stem Cell.* (2021) 28:1323–34.e8. doi: 10.1016/j.stem.2021.04.008

30. Acharya A, Berry DC, Zhang H, Jiang Y, Jones BT, Hammer RE, et al. miR-26 suppresses adipocyte progenitor differentiation and fat production by targeting Fbxl19. *Genes Dev.* (2019) 33:1367–80. doi: 10.1101/gad.328955.119

31. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, et al. Non-coding RNA analysis using the Rfam database. *Curr Protoc Bioinformatics.* (2018) 62:e51. doi: 10.1002/cpbi.51

32. Tarailo-Graovac M, Chen N. Using repeatmasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform.* (2009). doi: 10.1002/0471250953.bi0410s25

33. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* (2012) 40:37–52. doi: 10.1093/nar/gkr688

34. Denman RB. Using RNAFOLD to predict the activity of small catalytic RNAs. *Biotechniques.* (1993) 15:1090–5.

35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:550. doi: 10.1186/s13059-014-0550-8

36. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

37. Zhu Y, Zhang J, Hu X, Wang Z, Wu S, Yi Y. Extracellular vesicles derived from human adipose-derived stem cells promote the exogenous angiogenesis of fat grafts via the let-7/AGO1/VEGF signalling pathway. *Sci Rep.* (2020) 10:5313. doi: 10.1038/s41598-020-62140-6

38. Kang T, Lu W, Xu W, Anderson L, Bacanamwo M, Thompson W, et al. Correction: MicroRNA-27 (miR-27) targets prohibitin and impairs adipocyte differentiation and mitochondrial function in human adipose-derived stem cells. *J Biol Chem.* (2020) 295:16468. doi: 10.1074/jbc.AAC120.016601

39. Rodrigues AC, Spagnol AR, Frias FT, de Mendonca M, Araujo HN, Guimaraes D, et al. Intramuscular injection of miR-1 reduces insulin resistance in obese mice. *Front Physiol.* (2021) 12:676265. doi: 10.3389/fphys.2021.676265

40. Guo Y, Zhang X, Huang W, Miao X. Identification and characterization of differentially expressed miRNAs in subcutaneous adipose between Wagyu and Holstein cattle. *Sci Rep.* (2017) 7:44026. doi: 10.1038/srep44026

41. Miao X, Luo Q, Qin X, Guo Y. Genome-wide analysis of microRNAs identifies the lipid metabolism pathway to be a defining factor in adipose tissue from different sheep. *Sci Rep.* (2015) 5:18470. doi: 10.1038/srep18470

42. Li A, Huang W, Zhang X, Xie L, Miao X. Identification and characterization of CircRNAs of two pig breeds as a new biomarker in metabolism-related diseases. *Cell Physiol Biochem.* (2018) 47:2458–70. doi: 10.1159/000491619

43. Pomar CA, Serra F, Palou A, Sanchez J. Lower miR-26a levels in breastmilk affect gene expression in adipose tissue of offspring. *FASEB J.* (2021) 35:e21924. doi: 10.1096/fj.202100623R

44. Karbiener M, Pisani DF, Frontini A, Oberreiter LM, Lang E, Vegiopoulos A, et al. MicroRNA-26 family is required for human adipogenesis and drives characteristics of brown adipocytes. *Stem Cells.* (2014) 32:1578–90. doi: 10.1002/stem.1603

45. Zhang XX, Wang YM, Su YD, Zuo F, Wu B, Nian X. MiR-26a regulated adipogenic differentiation of ADSCs induced by insulin through CDK5/FOXC2 pathway. *Mol Cell Biochem.* (2021) 476:1705–16. doi: 10.1007/s11010-020-04033-w

46. Chen T, Cui J, Ma L, Zeng Y, Chen W. The Effect of MicroRNA-331-3p on Preadipocytes proliferation and differentiation and fatty acid accumulation in Laiwu pigs. *Biomed Res Int.* (2019) 2019:9287804. doi: 10.7287/peerj.preprints.27494v1

47. Nielsen M, Hansen JH, Hedegaard J, Nielsen RO, Panitz F, Bendixen C, et al. MicroRNA identity and abundance in porcine skeletal muscles determined by deep sequencing. *Anim Genet.* (2010) 41:159–68. doi: 10.1111/j.1365-2052.2009.01981.x

48. Wang J, Ren Q, Hua L, Chen J, Zhang J, Bai H, et al. Comprehensive analysis of differentially expressed mRNA, lncRNA and circRNA and their ceRNA networks in the longissimus dorsi muscle of two different pig breeds. *Int J Mol Sci.* (2019) 20:1107. doi: 10.3390/ijms20051107

49. Wang YX, Lee CH, Tiep S, Yu RT, Ham J, Kang H, et al. Peroxisome-proliferator-activated receptor delta activates fat metabolism to prevent obesity. *Cell.* (2003) 113:159–70. doi: 10.1016/S0092-8674(03)00269-1

50. Kang K, Reilly SM, Karabacak V, Gangl MR, Fitzgerald K, Hatano B, et al. Adipocyte-derived Th2 cytokines and myeloid PPARdelta regulate macrophage polarization and insulin sensitivity. *Cell Metab.* (2008) 7:485–95. doi: 10.1016/j.cmet.2008.04.002

51. Coll T, Alvarez-Guardia D, Barroso E, Gomez-Foix AM, Palomer X, Laguna JC, et al. Activation of peroxisome proliferator-activated receptor-{delta} by GW501516 prevents fatty acid-induced nuclear factor-{kappa}B activation and insulin resistance in skeletal muscle cells. *Endocrinology.* (2010) 151:1560–9. doi: 10.1210/en.2009-1211

52. Tang L, Lu Q, Cao H, Yang Q, Tong N. PPARD rs2016520 polymorphism is associated with metabolic traits in a large population of Chinese adults. *Gene.* (2016) 585:191–5. doi: 10.1016/j.gene.2016.02.035

53. Burch LR, Donnelly LA, Doney AS, Brady J, Tommasi AM, Whitley AL, et al. Peroxisome proliferator-activated receptor-delta genotype influences metabolic phenotype and may influence lipid response to statin therapy in humans: a genetics of diabetes audit and research Tayside study. *J Clin Endocrinol Metab.* (2010) 95:1830–7. doi: 10.1210/jc.2009-1201

54. Yu YH, Liu BH, Mersmann HJ, Ding ST. Porcine peroxisome proliferator-activated receptor gamma induces transdifferentiation of myocytes into adipocytes. *J Anim Sci.* (2006) 84:2655–65. doi: 10.2527/jas.2005-645

55. Yu YH, Wu SC, Cheng WT, Mersmann HJ, Ding ST. Ectopic expression of porcine peroxisome proliferator-activated receptor delta regulates adipogenesis in mouse myoblasts. *J Anim Sci.* (2008) 86:64–72. doi: 10.2527/jas.2007-0399

56. Pan C, Wang S, Yang C, Hu C, Sheng H, Xue X, et al. Genome-wide identification and expression profiling analysis of Wnt family genes affecting adipocyte differentiation in cattle. *Sci Rep.* (2022) 12:489. doi: 10.1038/s41598-021-04468-1

57. Zhang M, Zhang P, Liu Y, Lv L, Zhang X, Liu H, et al. RSPO3-LGR4 Regulates osteogenic differentiation of human adipose-derived stem cells via ERK/FGF signalling. *Sci Rep.* (2017) 7:42841. doi: 10.1038/srep42841

58. Loh NY, Minchin JEN, Pinnick KE, Verma M, Todorcevic M, Denton N, et al. RSPO3 impacts body fat distribution and regulates adipose cell biology *in vitro*. *Nat Commun.* (2020) 11:2797. doi: 10.1038/s41467-020-16592-z

59. Reggio A, Rosina M, Palma A, Cerquone Perpetuini A, Petrilli LL, Gargioli C, et al. Adipogenesis of skeletal muscle fibro/adipogenic progenitors is affected by the WNT5a/GSK3/beta-catenin axis. *Cell Death Differ.* (2020) 27:2921–41. doi: 10.1038/s41418-020-0551-y

60. Luther J, Driessler F, Megges M, Hess A, Herbort B, Mandic V, et al. Elevated Fra-1 expression causes severe lipodystrophy. *J Cell Sci.* (2011) 124:1465–76. doi: 10.1242/jcs.079855

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fgene.2022.910521

OPEN ACCESS

EDITED BY
Klaus Wimmers,
Leibniz Institute for Farm Animal Biology (FBN), Germany

REVIEWED BY
Xiangdong Ding,
China Agricultural University, China
Muhammed Walugembe,
Iowa State University, United States
Carina Visser,
University of Pretoria, South Africa

*CORRESPONDENCE
Bang Liu,
liubang@mail.hzau.edu.cn

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

RECEIVED 01 April 2022
ACCEPTED 29 July 2022
PUBLISHED 25 August 2022

CITATION
Yuan J, Zhou X, Xu G, Xu S and Liu B (2022), Genetic diversity and population structure of Tongcheng pigs in China using whole-genome SNP chip. *Front. Genet.* 13:910521. doi: 10.3389/fgene.2022.910521

Genetic diversity and population structure of Tongcheng pigs in China using whole-genome SNP chip

Jiao Yuan[1], Xiang Zhou[1,2,3], Guoqiang Xu[1], Sanping Xu[4] and Bang Liu[1,2,3]*

[1]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]The Cooperative Innovation Center for Sustainable Pig Production, Wuhan, China, [3]The Engineering Technology Research Center of Local Pig Breed Improvement of Hubei Province, Wuhan, China, [4]Department of Agricultural and Rural Bureau, Xianning, China

Tongcheng (TC) pigs, distinguished by their superior meat quality, are a Chinese indigenous pig breed. Recently, the genetic resources of TC pigs are under tremendous threat due to the introduction of cosmopolitan pig breeds and African swine fever disease. To promote their management and conservation, the present study assessed genetic diversity and population structure of TC pigs using single nucleotide polymorphism (SNP) markers. A total of 26, 999 SNPs were screened from 51, 315 SNPs in 68 TC pigs. The multi-dimensional scaling (MDS) analysis and neighbor-joining tree revealed that all 68 pigs were from a purebred population. The effective population size decreased over time, and it was 96 prior to generation 20. Both linkage disequilibrium (LD) and neutrality test indicated a low selection of TC pigs with average LD value of 0.15 ± 0.23. Genetic diversity results exhibited a minor allele frequency (MAF) of 0.23, observed heterozygosity (H_O) of 0.32, expected heterozygosity (He) of 0.31, and nucleotide diversity (Pi) of 0.31. All these parameters indicated a remarkably high genetic diversity of TC pigs. Additionally, 184 runs of homozygosity (ROH) segments were detected from the whole genome of TC pigs with an average ROH length of 23.71Mb, ranging from 11.26Mb to 69.02 Mb. The highest ROH coverage was found on chromosome 1 (10.12%), while the lowest was on chromosome 18 (1.49%). The average inbreeding coefficients based on ROH (F_{ROH}) was 0.04%. Fourteen ROH islands containing 240 genes were detected on 9 different autosomes. Some of these 240 genes were overlapped with the genes related to biological processes such as immune function, reproduction, muscular development, and fat deposition, including *FFAR2*, *FFAR4*, *MAPK8*, *NPY5R*, *KISS1*, and these genes might be associated with such traits as meat quality and disease resistance in TC pigs. Taken together, population structure and genetic diversity results suggested that the TC pig represented a valuable genetic resource. However, TC pig breed conservation program remains to be further optimized to ensure adequate genetic diversity and avoid inbreeding depression. Our findings provide theoretical basis for formulating management and conservation strategies for TC pigs.

KEYWORDS

tongcheng pig, population structure, genetic diversity, runs of homozygosity, SNP chip

Introduction

The Tongcheng (TC) pig, an important Chinese genetic resource, is mainly distributed in Tongcheng country (Hubei Province, China). They are distinguished by superior meat quality, moderate intramuscular fat, and intense flavor (Fan et al., 2006). Currently, TC pig attracts considerable attention due to its strong resistance to highly pathogenic Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) infection (Liang et al., 2017). Over past decades, commercial pig breeds have experienced intensive selection and have been imported into China. They dominate the Chinese pig industry, which poses a huge threat to indigenous pigs including TC pigs, thus resulting in the erosion to unique genetic resources (Wang et al., 2015). Additionally, the outbreak of African swine fever disease has placed TC pigs in danger. These severe challenges make the conservation of TC pigs an urgent and critical task.

An important prerequisite for the development and implementation of comprehensive conservation plans is to know the genetic structure of the existing livestock populations in a given region (FAO, 2015). The primary goal of this study is to investigate the population structure and genetic diversity so as to provide better conservation strategies for TC pigs. Numerous studies have shown that multi-dimensional scaling (MDS) and phylogenetic tree analyses are effective methods for the investigation of population structure, that the proportion of polymorphic loci (P_N), minor allele frequency (MAF), heterozygosity, and nucleotide diversity (Pi) in combination with linkage disequilibrium (LD) and effective population size (Ne) levels are good indicators reflecting the genetic diversity of the population, and that runs of homozygosity (ROH) and ROH island analyses can reveal the inbreeding of the genome (Cai et al., 2020; Xu et al., 2021). Owing to the rapid development of molecular biology in recent years, a variety of markers and DNA sequences of different species can be determined easily at low cost. Since single nucleotide polymorphism (SNP) markers are widely spread in the genome (Peripolli et al., 2017), they have been employed to study population structure and diversity of indigenous pig breeds in some countries such as China (Liu et al., 2020), Poland (Szmatola et al., 2020), Iberia (Saura et al., 2013), Russia, Kazakhstan, Ukraine (Traspov et al., 2016), Denmark (Cai et al., 2020), and South Africa (Hlongwane et al., 2020).

Therefore, to promote effective conservation and sustainable development of TC pigs, the present study characterized 1) population structure, 2) linkage disequilibrium and effective population size, 3) genetic diversity, and 4) runs of homozygosity. An insight into population structure and genetic diversity of TC pigs is crucial for the formulation of rational management strategies and the conservation of the unique traits of TC pigs.

Materials and methods

Samples, genomic data, and quality control

From 15 lineages with pedigree records from the Genetic Resource Preservation Farm (Tongcheng country, Hubei, China), we selected the representative 20 boars and 48 breeding sows as study object. Genomic DNA was extracted from ear tissues using a standard phenol/chloroform method (Moore, 1994). The extracted DNA samples were stored at −20 °C for further analysis.

All the individuals were genotyped with the "Zhongxin- I″ Porcine Chip (Beijing Compusen, China), using Infinium XT 96-Sample BeadChips which could provide the highest throughput array format based on the Infinium data, and a total of 51,315 SNPs were identified. The reference genomes of "Zhongxin- I″ Porcine Chip contained 9 populations of Chinese local pigs such as two-end-black pigs, Sutai, Laiwu, Erhualian, and Bamaxiang. Thus, this "Zhongxin- I″ Porcine Chip was selected for this study.

The quality control was conducted by the software Plink (v 1.90) (Purcell et al., 2007) to exclude unreliable genotypes. Briefly, the individuals with call rate (CR)≤95%, minor allele frequency (MAF) ≤0.01and Hardy-Weinberg equilibrium test (HWE) ($p \leq 10^{-6}$) were excluded (Zhao et al., 2018). Furthermore, those genotypes with genotype missing rate>5% were also excluded. After quality control, a total of 26, 999 SNPs (Supplementary Table 1) were screened from 51, 315 SNPs in 68 TC pigs for subsequent analysis.

Population structure analysis

Multi-dimensional scaling

The Plink (v 1.90) (Purcell et al., 2007) and Tassel (v 5.0) (Bradbury et al., 2007) were used for Multi-dimensional scaling (MDS) analysis, and the data were downscaled to three dimensions for population stratification.

Neighbor-joining (NJ) tree

TC pigs are currently a finite population, genetic drift and inbreeding are the two main reasons affecting genetic diversity. The degree of genetic drift and the increment in inbreeding is mainly determined by the smaller number of sexes, namely, the boar. So, the phylogenetic tree of 20 boars was constructed. The 20 boars cover all the pedigrees of TC pigs in the preservation farm.

To estimate the genetic distances among TC pigs, all 26, 999 SNPs were used to calculate the average proportion of alleles shared, Dst, using PLINK(Purcell et al., 2007). The definition of Dst is as follows:

$$D_{st} = \frac{IBS_2 + 0.5 + IBS_1}{N}$$

Where IBS_1 and IBS_2 are the numbers of loci that share one or two alleles at one locus. The genetic distance (D) between all pair-wise combinations of individuals was calculated as follows: $1 - D_{st}$.

Neighbor-joining (NJ) trees were conducted by using MEGA 11 (Tamura et al., 2021) based on the matrix of D.

The classification criteria are based on the requirements of the document "Administrative Measures for Livestock and Poultry Genetic Resources Conservation Farms and Gene Banks" issued by the Agriculture Ministry of China, namely, a single breed that meets the breeding standard is supposed to have no relationship within three generations. Assuming that individual X and parents D and E, there are no relationships between parents D and E within three generations, A is the common ancestor of individuals D and E, the inbreeding coefficients were calculated as follows:

$$F_X = \sum\left[\left(\frac{1}{2}\right)^{n_1+n_2+1}(1+F_A)\right]$$

And the relationships between parents is calculated as follows:

$$r_{DE} = \frac{\sum\left[\left(\frac{1}{2}\right)^{n_1+n_2}(1+F_A)\right]}{\sqrt{(1+F_D)(1+F_E)}}$$

Where D and E are non-inbred individuals, and thus the value of F_D and F_E is 0. According to the formula, F_X is 0.03125 and r_{DE} is 0.0625. No kinship within 3 generations and the kinship coefficient of less than 0.0625 are the two criteria for boar pedigree division. And the genetic distance (D) between individuals which is also calculated as D = 1-r.

Analyses of linkage disequilibrium and effective population size

Linkage disequilibrium

r^2 was used to evaluate Linkage disequilibrium (LD) extent of whole genome of TC pigs, and r^2 referred to the LD extent of each pair of SNPs per chromosome (Hill and Robertson, 1968). The r^2 of pairwise SNPs was calculated by using the parameters "plink--file--r2 --ld-window 99,999 --ld-window-r2 0 --out" in Plink (v 1.90) (Purcell et al., 2007). The physical distances between SNPs were divided into 100-kb intervals to visualize the decline of LD, and then the average of r^2 in each interval was estimated. The distribution of LD was plotted by R package ggPlot 2 (https://ggplot2.tidyverse.org/).

Effective population size (Ne)

Ne was estimated based on LD value (Corbin et al., 2012), the formula was as follows:

$$Ne_{(t)} = \frac{1}{(4f(C_t))}\left(\frac{1}{E\left[r^2_{adj}|C_t\right]} - \alpha\right)$$

Where $Ne_{(t)}$ is the effective population size prior to t generations; C_t is the recombination rate prior to t generations; r_{adj}^2 indicates the estimated linkage disequilibrium after sampling bias was adjusted; and α is a constant.

SNP-marker distances between 0 and 1,000 Mb were divided with 30 distance bins (50 kb each). The r^2 values corresponding to various distances were used to estimate Ne at different time points using SNeP version 1.1 (Barbato et al., 2015). The time span ranged from prior to generation 1,000 to 1 generation.

Genetic diversity indices

Genetic diversity is the variety of alleles and genotypes present in the group under study (population, species, or group of species). It is easy to understand that genetic diversity is the extent of heritable variation in a population, or Species. In this study, Seven indices were used to estimate TC-genetic diversity, of which P_N, MAF, Ho, and He were analyzed by software vcftools and Plink (v 1.90) (Purcell et al., 2007). The sequences were aligned by DnaSP v5 (Librado and Rozas, 2009), and Pi (Nei, 1987) were used to assess nucleotide polymorphisms. Neutrality test was conducted to assess the selective neutrality of mutations based on Tajima's D (Tajima, 1989), Fu and Li's F*, and Fu and Li's D* (Fu and Li, 1993).

Runs of homozygosity (ROH)

Measurement of ROH

To determine ROH, the Plink (v 1.90) parameters and thresholds (Purcell et al., 2007) were set as follows: 1) Sliding window was 20 SNPs across the entire genome; 2) Proportion of homozygous overlapping windows was 5%; 3) Minimum

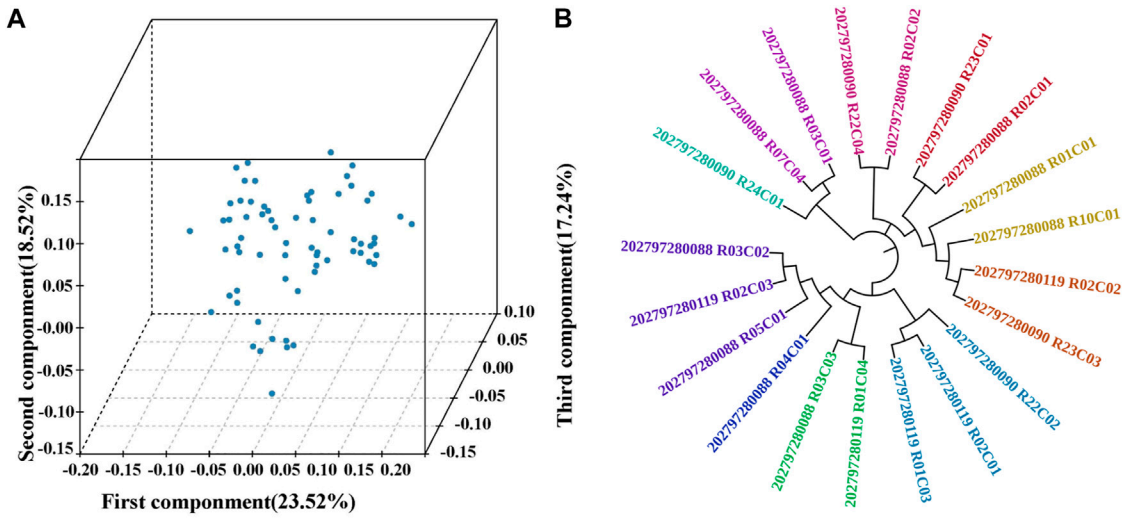


FIGURE 1
Population structure of Tongcheng (TC) pigs. **(A)** Multi-dimensional scaling analysis of TC pigs. The first three components (first component, second component and third component) were analyzed. **(B)** Phylogenetic relationships (unrooted NJ-tree) between 20 TC boars. 20 boars were divided into 10 families, and individuals with the same color represent one family, while different colors are different families.

length of a ROH was 1Mb; 4) Maximum gap between consecutive homozygous SNPs was 1,000 kb; and 5) Maximal number of SNPs with missing genotypes was 5 within a ROH, and the maximal number of heterozygous genotype was 1 within a ROH. Since the minimum ROH length was set as 1 Mb, no linkage disequilibrium (LD) -based pruning was needed to exclude short and common ROH induced by LD (Purfield et al., 2012).

The genomic ROH distribution was further investigated. Specifically, the mean number of ROHs per individual, the average length of ROH of population, the total number and average length of ROH per chromosome, and the percentage of chromosomes covered by ROH were calculated.

Genomic inbreeding coefficients

Genomic inbreeding coefficients per individual were calculated based on runs of homozygosity (F_{ROH}) according to the genome autozygotic proportion, as previously reported (McQuillan et al., 2008):

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

Where $\sum L_{ROH}$ is the sum of the length of all ROHs detected from an individual, and L_{AUTO} is the total length of the autosomal genome covered by SNPs included in the array. we excluded sibling effects on F_{ROH} and calculated F_{ROH} for 41 individuals without full-sib families, according to the pedigrees of 68 pigs provided by the preservation farm.

ROH islands and candidate genes

The R software was used to statistically analyze the occurrence frequency for which each SNP per individual fell inside a sliding window. Further, the top 1% of all SNPs across all ROH with the highest occurrences were defined as candidate SNPs under directional selection. All the adjacent SNPs within the top 1% were merged to form ROH islands. Our ROH islands were aligned against porcine Sscrofa11.1. The Ensemble Biomart (http://www.biomart. org) online tool was used to search annotated genes within aligned ROH islands. DAVID (http://david.abcc.ncifcrf.gov/) was employed to determine the function of annotated genes.

Results

Population structure

MDS analysis was conducted to cluster 68 TC pigs. The eigenvalues of the first three components were 23.52%, 18.52%, and 17.24%, respectively (Figure 1A). All the individuals were closely clustered with no visible separate subgroups. Further, the family structure of 20 boars were reconstructed based on the pairwise genetic distance. And the relationship between families was controlled within 0.0625. The clustering pattern showed that all individuals fell into ten families and the numbers per family ranged from 1 to 3 (Figure 1B), indicating that TC pigs were purebred from various families.

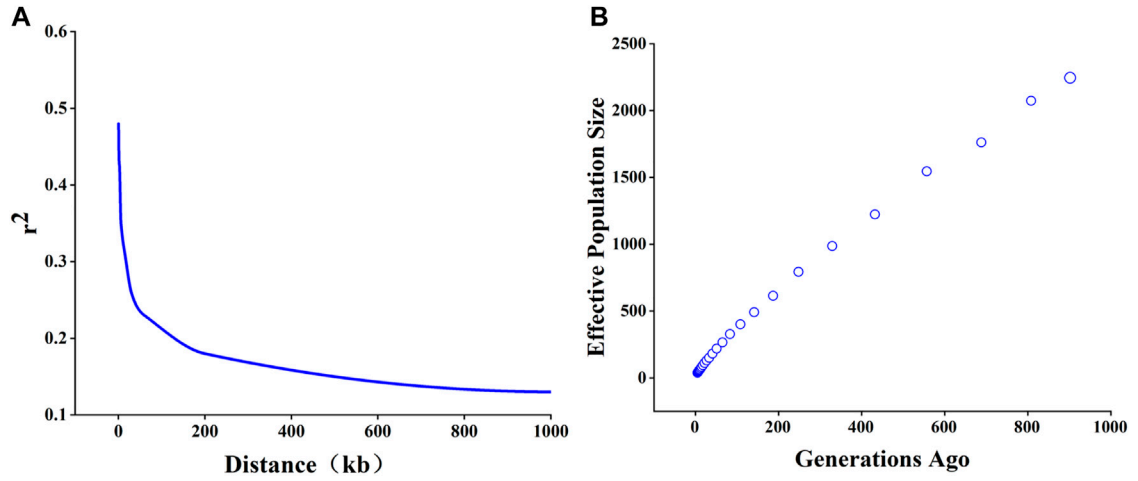


FIGURE 2
Linkage disequilibrium (LD) decay and effective population size (Ne) of Tongcheng (TC) pigs. **(A)** LD decay of TC pigs. **(B)** Ne of 68 TC pigs. The time span ranged from prior to generation 1,000 to 1 generation.

TABLE 1 Genetic diversity of Tongcheng pigs.

Number of SNPs	Genetic diversity			Nucleotide diversity		
	P_N	MAF	Ho	He	Pi	
	84.85%	0.23 ± 0.15	0.32 ± 0.16	0.31 ± 0.15	0.31	
	Tajima's D test		Fu and Li's D* test		Fu and Li's F* test	
26, 999	Statistic	Statistical significance	Statistic	Statistical significance	Statistic	Statistical significance
	2.38	*, $p < 0.05$	3.00	**, $p < 0.02$	3.23	**, $p < 0.02$

P_N, proportion of polymorphic loci; MAF minor allele frequency, Ho observed heterozygosity, He expected heterozygosity, Pi Nucleotide diversity, The following is the same.

Linkage disequilibrium and effective population size

The mean ± SD of estimated r^2 were 0.15 ± 0.23. The LD value was decreased with the increase of the distance between SNPs (Figure 2A). When physical distance between adjacent SNPs was 0.5, 1, and 3Mb, the corresponding average r^2 of TC pigs was 0.48, 0.45, and 0.42. The corresponding physical distance between pair-wise SNPs when r^2 decayed below 0.3 was chosen as the threshold for the assessment of the extent of LD patterns. Our results indicated that TC pigs had a short LD extent ($r_{0.3}^2$ = 15 kb). As shown in Figure 2B, Ne exhibited in a sharp decline within the time span from prior to the 1000th generation to the 5th generation. The Ne was approximately 401, 219, 96, 64 prior to generation 100, 50, 20, 12, respectively.

Genetic diversity

Seven indices were used to estimate genetic diversity of TC pigs, and the results were presented in Table 1. Most TC pigs exhibited a high P_N, ranging from 0.668 to 0.905. The minor allele frequency (MAF) was 0.23 ± 0.15, varying from 0.01 to 0.50. MAF of more than 23.15% SNPs was higher than 0.40, and that of 22.09% SNPs was lower than 0.10 (Supplementary Figure S1).

In TC pigs, He and Ho ranged from 0.01 to 0.5 and from 0.01 to 0.701, respectively, and Ho (0.32) was slightly higher than He (0.31) (Supplementary Figure S2). The Pi was 0.31. The results of Tajima's D, and Fu and Li's F*, and Fu and Li's D* were mostly positive at the corresponding loci. The means of Tajima's D test and Fu and Li's D* test for the entire population was 2.38 ($p < 0.05$) and 3.00 ($p < 0.02$), respectively.

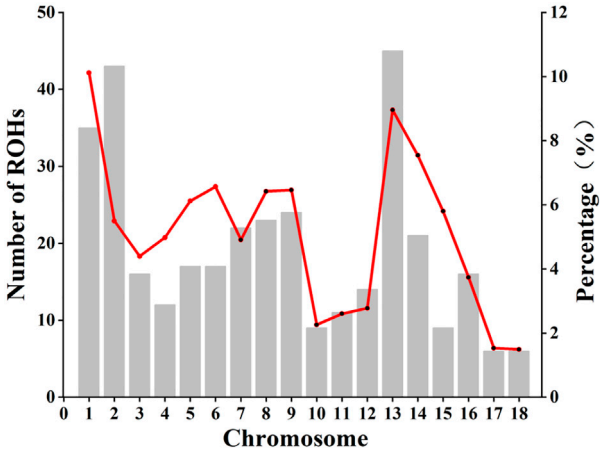


FIGURE 3
The number of runs of homozygosity (ROH) on each chromosome and the proportion covered by ROH. Bars and the red line represent the number of ROH and the ROH coverage, respectively, on each chromosome.

1 to 9, and the total length of ROH per individual was mainly distributed between 50 Mb and 75 Mb. The ROH number and the coverage ROH on each chromosome were illustrated in Figure 3. The length of ROH per chromosome was decrease with the decreased chromosome length (Supplementary Figure S3). The number of ROHs was large on chromosome 1, 2, 3, and 13, while the smallest number of ROHs was observed on chromosome 17 and 18 (Supplementary Table S2). The length of ROH was largest on chromosome 1 with an average ROH coverage of 10.12%, whereas the lowest ROH number was observed on chromosome 18 (ROH coverage of 1.49%). We excluded sibling effects on F_{ROH} and calculated F_{ROH} for 41 individuals without full-sib families. Based on ROH data, individual F_{ROH} was evaluated as 0.04%, indicating a low genomic inbreeding level.

As shown in Figure 4A, the percentage of SNPs in ROH region varied with chromosome. The most frequent SNP in ROH region (61 occurrences, 89.71%) was mapped at ~54 Mb on *sus scrofa* chromosome (SSC) 8. We identified a total of 14 ROH island regions (Supplementary Table S3) with minimum region length of 0.78 Mb on SSC 14 and maximum of 5.43 Mb on SSC 11, and a total

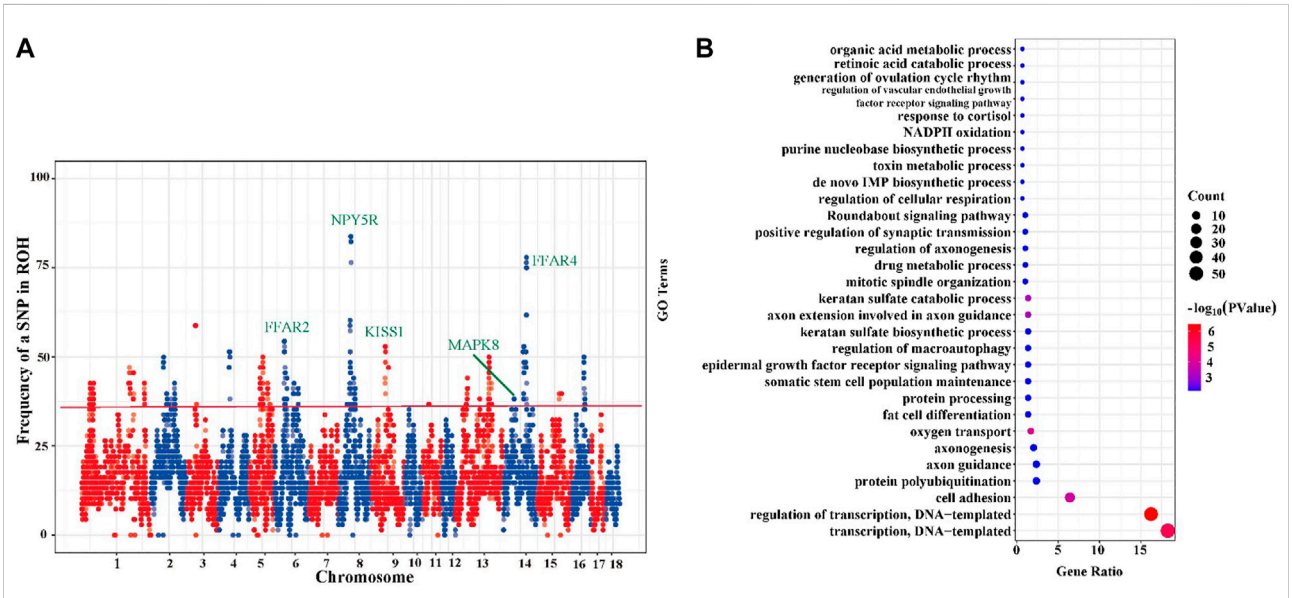


FIGURE 4
Incidence of each SNP in all runs of homozygosity (ROH) and the GO enrichment analysis of candidate genes. **(A)** Incidence of each SNP in all ROH segment. Particular genomic positions can be seen, on chromosomes 3,4,6,8,9,13,14,15,16 featured in ROH which were shared in over 50% of the sampled animals. **(B)** Bubble chart of GO terms enriched ($p < 0.05$) based on ROH islands. The larger the bubble, the more genes are enriched in this term, and the darker the red, the smaller the p value.

Runs of homozygosity

A total of 184 ROHs was detected from the entire genome with an average ROH length of 23.71 Mb, ranging from 11.26 Mb to 69.02 Mb. The average number of ROH per individual ranged from

of 305 SNPs were detected from these 14 ROH islands. Subsequently, the detected 305 SNPs were genetically annotated, and a total of 240 candidate genes were obtained. All these 240 candidate genes located in ROH islands were subjected to GO enrichment analysis and these genes were found to be

significantly enriched in 30 GO terms ($p < 0.05$) (Figure 4B, Supplementary Table S4). Among these terms, we focused on three important terms related to the germplasm characteristics of TC pigs, namely "fat cell differentiation (GO:0045444)," "regulation of macroautophagy (GO:0016241)," "generation of ovulation cycle rhythm (GO:0060112)". TC pigs as germplasm resources are characterized by excellent meat quality, moderate intramuscular fat, and strong flavor. We identified several important candidate genes that may be involved in the formation of these genetic characteristics, such as *FFAR2*, *FFAR4*, *MAPK8*, *NPY5R*, and *KISS1*.

Discussion

The alleles of TC pigs were expected to be diverse due to the absence of intense directional selection of production traits. Maintaining genetic diversity within and across breeds is crucial for sustainable development of livestock production (Ollivier and Foulley, 2005). To reveal genetic structure and diversity of TC pigs, we investigated population structure, linkage disequilibrium, effective population size, genetic diversity and runs of homozygosity based on genome-wide SNP data.

Population structure of TC pigs

To understand the population structure, we conducted MDS analysis and constructed Neighbor-joining (NJ) tree. Figure 1A showed a weak differentiation between 68 TC pigs, which indicated that TC breeding pigs conservation population is an unstratified purebred population. Furthermore, NJ tree showed a sufficient pedigree number (10), and the number of boars in each pedigree varied from 1 to 3, suggesting that there was an uneven distribution of families in this population. The uneven family distribution will increase the possibility of decreasing genetic diversity of TC pigs. To avoid the reduction in genetic diversity and inbreeding depression, it is recommended to maintain equivalent numbers of boar in each family.

Genetic diversity of TC pigs

Genetic diversity is affected by various factors. This study comprehensively assessed the genetic diversity of TC pigs by investigating LD, Ne, genetic diversity parameters (P_N, MAF, heterozygosity, and Pi) and neutrality tests. All these parameters indicated a significant level of genetic diversity.

LD describes the non-random association of alleles at different loci, and this association results from processes such as migration, selection, and genetic drift in finite populations (Wang, 2005). The pattern of LD decay can provide information on the evolution of a population. Therefore, comparing the extent of LD between populations can reveal the overall diversity level of various populations. As expected, the most tightly-linked SNP pairs in TC pig genome exhibited the highest r^2, and average r^2 was rapidly decreased with the increasing distance between pairwise SNPs, which was consistent with the pattern observed in previous studies of pigs and in other species (Ai et al., 2013; Chen J. et al., 2018). The distance between pairwise SNPs in TC pigs was 15 kb when $r^2 = 0.3$ (Jungerius et al., 2005). Previous studies of the genetic diversity have shown that commercial pigs generally have greater LD than Chinese indigenous pigs, with physical distance between pairwise SNPs ranging from 334 kb (Landrace) to 757 kb (White Duroc) for commercial pigs and from 4.5 kb (Wuzhishan) to 744 kb (Min) for Chinese pigs, and 1.5 kb for wild pigs (Ai et al., 2013; Wang et al., 2018; Xu et al., 2019; Liu et al., 2020; Wang et al., 2021). In this study, the LD extension of TC pigs was much smaller than commercial pigs, and it was at low level among Chinese indigenous pigs. LD can reflect the genomic diversity and haplotype diversity of the population. Genome diversity has been reported to be decreased with the LD extension increasing (McVicker et al., 2009). Based on this, it could be speculated that the genetic diversity of the TC pigs is more likely to remain stable. In addition, the sharp decline of LD within short distance (15 kb) indicated high haplotype diversity in the studied population (Figure 2A), and the similar trend has been reported in cattle (Shin et al., 2013; Kim et al., 2018). When a new mutation occurs in a finite population, LD is affected, and the influence degree is dependent on the frequency of the allele induced by the new mutation. With the accumulation of copies of the mutant allele, the LD between this mutated allele and other alleles depends on recombinant rate, random drift, population admixture, and selection. The LD between mutated alleles and their neighboring alleles is generally lower than the average LD. In this study, more new alleles were generated in recent generations of TC pigs with no artificial selection, which might be one of reasons for the rapid decay of LD. This result indicated the abundant genetic variation of TC pig genome.

Linkage disequilibrium (LD) can provide an insight into the evolution of one population. The degree of LD between SNPs can be used to infer ancestral Ne, and Ne is the number of individuals of an ideal population with the same inbreeding coefficients as the actual breeding population (Frankham, 1996). Balanced polymorphisms will be lost when population size becomes too small (Kimura, 1964), thus reducing the genetic diversity of populations. A large effective population size contributes to the conservation of numerous genetic variations. Our data indicated that the Ne of TC before generation 12 (Ne_{12}) was 64, which was larger than the minimum threshold (50) set by the Food and Agricultural Organization (FAO, 2000), and that the average Ne before generation 20 (Ne_{20}) was estimated to be 96. The Ne_{20} of TC pigs was generally smaller than that of global pig breeds such as the Large White (214.4), Duroc (207.2) and Landrace (207.5) (Traspov et al., 2016), and it was comparable to Chinese indigenous pig breeds, which ranged from 85.0 for Shaziling pigs to 165.0 for Erhualian pigs (Wang et al., 2018).

This result might be explained by the fact that the population of the Chinese indigenous breeds were smaller than global pig breeds due to the extensive importation of commercial breeds into China during past decades. A comprehensive population structure analysis showed that TC pig families with fewer boars needed to increase their numbers. Further, it is possible to properly introduce purebred TC pigs from other protected areas and increase the pedigree of TC pigs, which can effectively increase the population size.

It has been reported that Chinese indigenous pigs are richer in genetic variability than commercial pigs (Ai et al., 2013; Wang et al., 2015). Hence, Chinese indigenous pigs are expected to have larger Ho. Our estimated heterozygosity values (average Ho 0.32, and He 0.31) were consistent with those in most previous studies of different pig populations (Ai et al., 2013; Traspov et al., 2016; Wang et al., 2018). Our Ho values were considerably higher than those reported in previous studies of cosmopolitan pig breeds, which ranged from 0.25 (Hampshire) to 0.34 (Large White) (Traspov et al., 2016), and our Ho values of TC breed were similar to those of some Chinese local breeds ranging from 0.15 of Meishan breed to 0.349 of Laiwu breed (Traspov et al., 2016; Wang et al., 2018; Xu et al., 2019; Liu et al., 2020; Wang et al., 2021). Our He values of TC breed were similar to those of cosmopolitan pig breeds, which ranged from 0.27 (Berkshire) to 0.395 (Landrace) (Wang et al., 2018), but they were higher than those of some Chinese local breeds ranging from 0.14 (Jinhua) to 0.4 (Kele) (Ai et al., 2013; Xu et al., 2019). In addition to the heterozygosity, MAF is also extensively used for genetic diversity studies since it can distinguish between common and rare variant in the population, and A higher proportion of low MAF values is associated with high genetic diversity is associated with high genetic diversity (Murray et al., 2010; O'Brien et al., 2014). In this study, the mean MAF was 0.23, and the proportion of MAF<0.1 was high, which indicated the high genetic diversity of the TC pig population. Pi are the basic parameters to assess genetic diversity. Pi is to measure the degree of intrapopulation haplotype mutation. Our results showed that the TC pig population had higher nucleotide diversity (Pi, 0.31) than Chinese indigenous pig breed Bamei (Pi, 0.01) and commercial breed (Pi, 0.01–0.02) (Zhang et al., 2018). The high Pi of the genome provided support for the fast-decaying LD. Additionally, the positive mean values of Tajima's D, Fu and Li's F*, and Fu and Li's D* TC indicated that pigs might have experienced recent size reductions, but they might not indicate the deviations from the neutral expectation. The studies of molecular and phenotypic evolution have revealed that positive selection of advantageous mutations drives divergence between populations (species) and reduce polymorphism within populations (species) (Suzuki, 2010). All our parameters showed a high level of genetic diversity of TC pigs. Since genetic diversity is affected by multiple factors, genetic variation should be continuously monitored in breed conservation to prevent an irreversible erosion of genetic diversity of livestock populations, thus maximizing breed adaptability (Biscarini et al., 2015).

ROH distribution and candidate genes of tongcheng pigs

ROH level of TC pigs

Short ROH segments are known as indicators of distant consanguinity, whereas long ROH segments are more likely to result from recent inbreeding (Curik et al., 2014). Average ROH length of Chinese pigs varied considerably, ranging from the lowest value of 20.6 Mb in wild boar to the largest value of 168 Mb in Diannanxiaoer (Wang et al., 2018).

Broman and Weber (1999) first identified long homozygous segments in the human genome, and Gibson et al. (Broman and Weber, 1999) described the potential of long homozygous segments for inbreeding assessment. F_{ROH} was first defined by McQuillan et al. (2008). Since 2010, F_{ROH} is an accurate estimator of inbreeding coefficient, and thus it is used for studying inbreeding and selection in animal populations. F_{ROH} was the first reported in 2012 in pigs (Bosse et al., 2012), and recently, it has been used for calculating inbreeding coefficient of Jinhua pigs (F_{ROH} 0.17) (Xu et al., 2019). For example, the Laiwu pig breed has shrunk with the introduction of a large number of commercial breeds, and its F_{ROH} was 0.06 in 2021 (Fang et al., 2021). In this study, we excluded sibling effects on F_{ROH} and calculated F_{ROH} for 41 individuals without full-sib families. TC pigs exhibited a low F_{ROH} (0.04%), indicating a low inbreeding level and preservation of genetic diversity.

ROH islands and candidate genes

ROH contributes to the investigation of inbred genomic regions within a population, which were first defined as ROH islands by Nothnagel et al. (2010). Selection leaves certain peaks across the genome, and these peaks are called hotspots and considered to be the signal of selective sweeps (Curik et al., 2014). These hotspots (stretches of homozygous sequences) are shared by a large proportion of individuals in a population, and they are designated as ROH islands.

For instance, some candidate genes such as *FFAR2*, *FFAR4*, *MAPK8*, *NPY5R*, and *KISS1* located in ROH islands were found to be related to fat cell differentiation, fat deposition, and reproduction. Of them, *FFAR2* and *FFAR4* (free fatty acid receptors 2 and 4) were mammalian receptor of short-chain fatty acids (SCFAs). Free fatty acid receptors (*FFAR*) were found on the apical or basolateral membranes of enteroendocrine and immune cells at the gut mucosal lining (Rohrl et al., 2011; Mielenz, 2017; Psichas et al., 2017; Kimura et al., 2020). Previous study has reported that the expression of *FFAR2* is potentially related to the abundance of short chain fatty acids (Vigors et al., 2019). Recent research has indicated that SCFAs might affect immunity, metabolism, and the pathogenesis of obesity, diabetes, and inflammatory bowel disease in humans through gut microbiota (Meijer et al., 2010). One study of knockout mice has indicated that *FFAR2* plays a major role in mediating the effect of gut microbiota on host obesity and immunity (Bjursell et al., 2011). However, studies in pigs have found

that the adipogenic effect of SCFAs in porcine is unlikely to be mediated by *FFAR*2 (Li et al., 2014). TC pigs have excellent meat quality and flavor, indicating that gene *FFAR* may play a key role in fatty acid composition in TC pig population, but more evidence is needed. Candidate gene *MAPK8*, harboured on ROH islands, has been reported to control an important trait in TC pigs, and this gene is significantly correlated with backfat thickness in carcasses (Otieno et al., 2005). In addition, the mitogen-activated protein kinase (*MAPK*) signaling pathway participates in multiple biological processes including innate immunity, cell growth, stress response, apoptosis, and differentiation (Purves et al., 2001). The *MAPK* signaling pathway is activated by cells after infection and intoxication (Widmann et al., 1999). Previous study has shown that the TC pig strong resistance to highly pathogenic PRRSV infection (Liang et al., 2017), indicating that gene *MAPK8* might play a important role in PRRSV infection in TC pig population, but more evidence is needed. Neuropeptide Y5 receptor (*NPY5R*) is expressed in hypothalamic areas that control feeding (Larsen et al., 1993). The *NPY* (including *NPY1R* and *NPY5R*) is mainly responsible for promoting food intake and reducing energy consumption as well as fat deposition in mice (Nguyen et al., 2012). The selection signals of *NPY5R* genes have also been detected in multiple Chinese local pig breeds, such as Tongcheng (Wang et al., 2015), Rongchang, and Jinhua (Elbers et al., 2009), Laiwu (Chen M. H. et al., 2018), and Anqing six-end-white pigs (Zhang et al., 2020). Kisspeptin (first named metastin), a peptide product of *KISS1* gene, is a neuropeptide responsible for regulating reproduction (Ohtaki et al., 2001). Kisspeptin effectively stimulates LH and FSH release in prepubertal gilts (Lents et al., 2018). TC pigs have early sexual maturity, and the sows exhibit sexual maturity at the age of 4–5 months (Data unpublished). Therefore, *KISS1* gene may be an essential gene affecting the reproduction performance of sows, which deserves more attention.

Conclusion

Population structure and genetic diversity are of great importance for genetic improvement programs of indigenous pig and the conservation and effective use of genetic resources. The present study provides insights into the population structure and genetic diversity of TC pigs in China. The results showed that the studied TC pig populations were free from admixture of other breeds. Our investigated parameters such as P_N, MAF, He and Ho, and nucleotide diversity showed a remarkably high level of genetic diversity of TC pigs. The population structure and genetic diversity analyses indicated that TC pigs represented a valuable genetic resource, but their breed conservation program remains to be further optimized so as to ensure adequate genetic diversity and avoid inbreeding depression. Our findings lay a theoretical basis for conservation of TC pigs and provide genome data for subsequent studies of meat traits such as fatty acids and meat quality.

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 All the animal experiments were approved and conducted ethically by the Institutional Animal Care and Use Committee of Huazhong Agricultural University. (Approval No. 2017005).

Author contributions

BL and XZ designed the project, JY and GX collected the datasets, SX provided samples, JY and GX carried out the coding of the computational analysis, JY drafted the manuscript, and BL and XZ revised and polished the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China (31930104).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.910521/full#supplementary-material

References

Ai, H. S., Huang, L. S., and Ren, J. (2013). Genetic diversity, linkage disequilibrium and selection signatures in Chinese and western pigs revealed by genome-wide SNP markers. *Plos One* 8 (2), e56001. doi:10.1371/journal.pone.0056001

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

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, 33. doi:10.3389/fgene.2015.00033

Bjursell, M., Admyre, T., Goransson, M., Marley, A. E., Smith, D. M., Oscarsson, J., et al. (2011). Improved glucose control and reduced body fat mass in free fatty acid receptor 2-deficient mice fed a high-fat diet. *Am. J. Physiol. Endocrinol. Metab.* 300 (1), E211–E220. doi:10.1152/ajpendo.00229.2010

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: Consequence of demography and the recombination landscape. *PLoS Genet.* 8 (11), e1003100. doi:10.1371/journal.pgen.1003100

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 (19), 2633–2635. doi:10.1093/bioinformatics/btm308

Broman, K. W., and Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* 65 (6), 1493–1500. doi:10.1086/302661

Cai, Z. X., Sarup, P., Ostersen, T., Nielsen, B., Fredholm, M., Karlskov-Mortensen, P., et al. (2020). Genomic diversity revealed by whole-genome sequencing in three Danish commercial pig breeds.ARTN skaa229. *J. Anim. Sci.* 98 (7), skaa229. doi:10.1093/jas/skaa229

Chen, J., Peng, J., Xiao, Q., Pan, Y., Zhang, X., Lo, L. J., et al. (2018). The genetic diversity and population structures of indigenous pig breeds in Zhejiang Province revealed by GGRS sequencing. *Anim. Genet.* 49 (1), 36–42. doi:10.1111/age.12625

Chen, M. H., Wang, J. Y., Wang, Y. P., Wu, Y., Fu, J. L., and Liu, J. F. (2018). Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. *BMC Genet.* 19, 31. doi:10.1186/s12863-018-0622-y

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Animal Breed. Genet.* 129 (4), 257–270. doi:10.1111/j.1439-0388.2012.01003.x

Curik, I., Ferencakovic, M., and Solkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi:10.1016/j.livsci.2014.05.034

Elbers, C. C., de Kovel, C. G. F., van der Schouw, Y. T., Meijboom, J. R., Bauer, F., Grobbee, D. E., et al. (2009). Variants in neuropeptide Y receptor 1 and 5 are associated with nutrient-specific food intake and are under recent selection in Europeans. *Plos One* 4 (9), e7070. doi:10.1371/journal.pone.0007070

Fan, B., Tang, Z. L., Xu, S. P., Liu, B., Peng, Z. Z., and Li, K. (2006). Germplasm characteristics and conservation of Tongcheng pig: A case study for preservation and utilization of Chinese indigenous pig breeds. *Anim. Genet. Resour. Inf.* 39, 51–63. doi:10.1017/s1014233900002133

Fang, Y., Hao, X., Xu, Z., Sun, H., Zhao, Q., Cao, R., et al. (2021). Genome-wide detection of runs of homozygosity in Laiwu pigs revealed by sequencing data. *Front. Genet.* 12, 629966. doi:10.3389/fgene.2021.629966

FAO (2000). *Secondary guidelines of farm animal genetic resources management plans. Management of small populations at risk*. Rome: Food and Agriculture Organization of the United Nations.

FAO (2015). *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*. Editors B. D. Scherf and D. Pilling (Rome: FAO Commission on Genetic Resources for Food and Agriculture Assessments). Available at: http://www.fao.org/3/a-i4787e/index.html.

Frankham, R. (1996). Introduction to quantitative genetics (4th edn). *Am. J. Hum. Genet.* 46 (7), 1231.

Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133 (3), 693–709. doi:10.1093/genetics/133.3.693

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38 (6), 226–231. doi:10.1007/BF01245622

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

Jungerius, B. J., Gu, J. J., Crooijmans, R. P. M. A., van der Poel, J. J., Groenen, M. A. M., van Oost, B. A., et al. (2005). Estimation of the extent of linkage disequilibrium in seven regions of the porcine genome. *Anim. Biotechnol.* 16 (1), 41–54. doi:10.1081/Abio-200053402

Kim, S., Cheong, H. S., Shin, H. D., Lee, S. S., Roh, H. J., Jeon, D. Y., et al. (2018). Genetic diversity and divergence among Korean cattle breeds assessed using a BovineHD single-nucleotide polymorphism chip. *Asian-Australas. J. Anim. Sci.* 31 (11), 1691–1699. doi:10.5713/ajas.17.0419

Kimura, I., Ichimura, A., Ohue-Kitano, R., and Igarashi, M. (2020). Free fatty acid receptors in health and disease. *Physiol. Rev.* 100 (1), 171–210. doi:10.1152/physrev.00041.2018

Kimura, M. C. J., and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49, 725–738. doi:10.1093/genetics/49.4.725

Larsen, P. J., Sheikh, S. P., Jakobsen, C. R., Schwartz, T. W., and Mikkelsen, J. D. (1993). Regional distribution of putative NPY Y1 receptors and neurons expressing Y1 mRNA in forebrain areas of the rat central nervous system. *Eur. J. Neurosci.* 5 (12), 1622–1637. doi:10.1111/j.1460-9568.1993.tb00231.x

Lents, C. A., Heidorn, N. L., Barb, R., and Ford, J. J. (2018). Reproduction research central and peripheral administration of kisspeptin activates gonadotropin but not somatotropin secretion in prepubertal gilts. *Soc. reproduction Fertil.* 136 (6), 879–887. doi:10.1530/rep-07-0502

Li, G. L., Yao, W., and Jiang, H. L. (2014). Short-chain fatty acids enhance adipocyte differentiation in the stromal vascular fraction of porcine adipose tissue. *J. Nutr.* 144 (12), 1887–1895. doi:10.3945/jn.114.198531

Liang, W., Ji, L. K., Zhang, Y., Zhen, Y. R., Zhang, Q. D., Xu, X. W., et al. (2017). Transcriptome differences in porcine alveolar macrophages from Tongcheng and large white pigs in response to highly pathogenic porcine reproductive and respiratory Syndrome Virus (PRRSV) infection. *Int. J. Mol. Sci.* 18 (7), E1475. doi:10.3390/ijms18071475

Librado, P., and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25 (11), 1451–1452. doi:10.1093/bioinformatics/btp187

Liu, C., Li, P., Zhou, W., Ma, X., Wang, X., Xu, Y., et al. (2020). Genome data uncover conservation status, historical relatedness and candidate genes under selection in Chinese indigenous pigs in the taihu lake region. *Front. Genet.* 11, 591. doi:10.3389/fgene.2020.00591

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 (3), 359–372. doi:10.1016/j.ajhg.2008.08.007

McVicker, G., Gordon, D., Davis, C., and Green, P. (2009). Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* 5 (5), e1000471. doi:10.1371/journal.pgen.1000471

Meijer, K., de Vos, P., and Priebe, M. G. (2010). Butyrate and other short-chain fatty acids as modulators of immunity: What relevance for health? *Curr. Opin. Clin. Nutr. Metab. Care* 13 (6), 715–721. doi:10.1097/MCO.0b013e32833eebe5

Mielenz, M. (2017). Invited review: Nutrient-sensing receptors for free fatty acids and hydroxycarboxylic acids in farm animals. *Animal* 11 (6), 1008–1016. doi:10.1017/S175173111600238X

Moore, D. (1994). Purification and concentration of DNA from aqueous solutions: Preparation and analysis of DNA. *Curr. Protoc. Mol. Biol.* 25 (1), 2.1.1–2.1.9. doi:10.1002/j.1934-3647.1994.tb00220.x

Murray, C., Huerta-Sanchez, E., Casey, F., and Bradley, D. G. (2010). Cattle demographic history modelled from autosomal sequence variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365 (1552), 2531–2539. doi:10.1098/rstb.2010.0103

Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

Nguyen, A. D., Mitchell, N. F., Lin, S., Macia, L., Yulyaningsih, E., Baldock, P. A., et al. (2012). Y1 and Y5 receptors are both required for the regulation of food intake and energy homeostasis in mice. *Plos One* 7 (6), e40191. doi:10.1371/journal.pone.0040191

Nothnagel, M., Lu, T. T., Kayser, M., and Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol. Genet.* 19 (15), 2927–2935. doi:10.1093/hmg/ddq198

O'Brien, A. M. P., Meszaros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, J. F., Van Tassell, C. P., et al. (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. *Livest. Sci.* 166, 121–132. doi:10.1016/j.livsci.2014.05.007

Ohtaki, T., Shintani, Y., Honda, S., Matsumoto, H., Hori, A., Kanehashi, K., et al. (2001). Metastasis suppressor gene KiSS-1 encodes peptide ligand of a G-protein-coupled receptor. *Nature* 411(6837), 613–617. doi:10.1038/35079135

Ollivier, L., and Foulley, J. L. (2005). Aggregate diversity: New approach combining within- and between-breed genetic diversity. *Livest. Prod. Sci.* 95 (3), 247–254. doi:10.1016/j.livprodsci.2005.01.005

Otieno, C. J., Bastiaansen, J., Ramos, A. M., and Rothschild, M. F. (2005). Mapping and association studies of diabetes related genes in the pig. *Anim. Genet.* 36 (1), 36–42. doi:10.1111/j.1365-2052.2004.01217.x

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., and Baldi, F. (2017). Runs of homozygosity: Current knowledge and applications in livestock. *Anim. Genet.* 48 (3), 255–271. doi:10.1111/age.12526

Psichas, A., Larraufie, P. F., Goldspink, D. A., Gribble, F. M., and Reimann, F. (2017). Chylomicrons stimulate incretin secretion in mouse and human cells. *Diabetologia* 60 (12), 2475–2485. doi:10.1007/s00125-017-4420-2

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 (3), 559–575. doi:10.1086/519795

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13, 70. doi:10.1186/1471-2156-13-70

Purves, T., Middlemas, A., Agthong, S., Jude, E. B., Boulton, A. J., Fernyhough, P., et al. (2001). A role for mitogen-activated protein kinases in the etiology of diabetic neuropathy. *FASEB J.* 15 (13), 2508–2514. doi:10.1096/fj.01-0253hyp

Rohrl, C., Kaindl, U., Koneczny, I., Hudec, X., Baron, D. M., Konig, J. S., et al. (2011). Peroxisome-proliferator-activated receptors γ and β/δ mediate vascular endothelial growth factor production in colorectal tumor cells. *J. Cancer Res. Clin. Oncol.* 137 (1), 29–39. doi:10.1007/s00432-010-0856-1

Saura, M., Fernandez, A., Rodriguez, M. C., Toro, M. A., Barragan, C., Fernandez, A. I., et al. (2013). Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. *PLoS One* 8 (10), e78314. doi:10.1371/journal.pone.0078314

Shin, D. H., Cho, K. H., Park, K. D., Lee, H. J., and Kim, H. (2013). Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. *Asian-Australas. J. Anim. Sci.* 26 (12), 1672–1679. doi:10.5713/ajas.2013.13320

Suzuki, Y. (2010). Statistical methods for detecting natural selection from genomic data. *Genes Genet. Syst.* 85(6), 359–376. doi:10.1266/ggs.85.359

Szmatola, T., Jasielczuk, I., Semik-Gurgul, E., Szyndler-Nedza, M., Blicharski, T., Szulc, K., et al. (2020). Detection of runs of homozygosity in conserved and commercial pig breeds in Poland. *J. Anim. Breed. Genet.* 137 (6), 571–580. doi:10.1111/jbg.12482

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123 (3), 585–595. doi:10.1093/genetics/123.3.585

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi:10.1093/molbev/msab120

Traspov, A., Deng, W., Kostyunina, O., Ji, J., Shatokhin, K., Lugovoy, S., et al. (2016). Population structure and genome characterization of local pig breeds in Russia, Belorussia, Kazakhstan and Ukraine. *Genet. Sel. Evol.* 48, 16. doi:10.1186/s12711-016-0196-y

Vigors, S., O'Doherty, J. V., Ryan, M., and Sweeney, T. (2019). Analysis of the basal colonic innate immune response of pigs divergent in feed efficiency and following an *ex vivo* lipopolysaccharide challenge. *Physiol. Genomics* 51 (9), 443–448. doi:10.1152/physiolgenomics.00013.2019

Wang, C., Wang, H. Y., Zhang, Y., Tang, Z. L., Li, K., and Liu, B. (2015). Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol. Ecol. Resour.* 15 (2), 414–424. doi:10.1111/1755-0998.12311

Wang, J. L. (2005). Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1459), 1395–1409. doi:10.1098/rstb.2005.1682

Wang, X., Wang, C., Huang, M., Tang, J., Fan, Y., Li, Y., et al. (2018). Genetic diversity, population structure and phylogenetic relationships of three indigenous pig breeds from Jiangxi Province, China, in a worldwide panel of pigs. *Anim. Genet.* 49 (4), 275–283. doi:10.1111/age.12687

Wang, X., Zhang, H., Huang, M., Tang, J., Yang, L., Yu, Z., et al. (2021). Whole-genome SNP markers reveal conservation status, signatures of selection, and introgression in Chinese Laiwu pigs. *Evol. Appl.* 14 (2), 383–398. doi:10.1111/eva.13124

Widmann, C., Gibson, S., Jarpe, M. B., and Johnson, G. L. (1999). Mitogen-activated protein kinase: Conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79 (1), 143–180. doi:10.1152/physrev.1999.79.1.143

Xu, P., Wang, X. P., Ni, L. G., Zhang, W., Lu, C. L., Zhao, X., et al. (2019). Genome-wide genotyping uncovers genetic diversity, phylogeny, signatures of selection, and population structure of Chinese Jiangquhai pigs in a global perspective. *J. Anim. Sci.* 97 (4), 1491–1500. doi:10.1093/jas/skz028

Xu, Z., Mei, S., Zhou, J., Zhang, Y., Qiao, M., Sun, H., et al. (2021). Genome-wide assessment of runs of homozygosity and estimates of genomic inbreeding in a Chinese composite pig breed. *Front. Genet.* 12, 720081. doi:10.3389/fgene.2021.720081

Zhang, J. X., Yang, B. C., Wen, X. C., and Sun, G. Q. (2018). Genetic variation and relationships in the mitochondrial DNA D-loop region of Qinghai indigenous and commercial pig breeds. *Cell. Mol. Biol. Lett.* 23, 31. doi:10.1186/s11658-018-0097-x

Zhang, W., Yang, M., Zhou, M., Wang, Y. L., Wu, X. D., Zhang, X. D., et al. (2020). Identification of signatures of selection by whole-genome resequencing of a Chinese native pig. *Front. Genet.* 11, 566255. doi:10.3389/fgene.2020.566255

Zhao, S., Jing, W., Samuels, D. C., Sheng, Q., Shyr, Y., and Guo, Y. (2018). Strategies for processing and quality control of Illumina genotyping arrays. *Brief. Bioinform.* 19 (5), 765–775. doi:10.1093/bib/bbx012

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fgene.2022.910998

OPEN ACCESS

EDITED BY
Natalia A Zinovieva,
L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia

REVIEWED BY
Mario Calus,
Wageningen University and Research, Netherland
Gábor Mészáros,
University of Natural Resources and Life Sciences Vienna, Austria

*CORRESPONDENCE
Heather J. Huson,
hjh3@cornell.edu

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

RECEIVED 01 April 2022
ACCEPTED 18 August 2022
PUBLISHED 26 September 2022

CITATION
Jaafar MA, Heins BJ, Dechow C and Huson HJ (2022), The impact of using different ancestral reference populations in assessing crossbred population admixture and influence on performance.
Front. Genet. 13:910998.
doi: 10.3389/fgene.2022.910998

The impact of using different ancestral reference populations in assessing crossbred population admixture and influence on performance

Mohd A. Jaafar[1], Bradley J. Heins[2], Chad Dechow[3] and Heather J. Huson[1]*

[1]Department of Animal Science, Cornell University, Ithaca, NY, United States, [2]West Central Research and Outreach Centre, University of Minnesota, Morris, MN, United States, [3]Department of Animal Science, Penn State University, State College, University Park, PA, United States

Crossbreeding is a process in which animals from different breeds are mated together. The animals produced will exhibit a combination of both additive and non-additive genetic improvement from parental breeds that increase heterozygosity and negate inbreeding depression. However, crossbreeding may also break up the unique and often beneficial gene combinations in parental breeds, possibly reducing performance potential as the benefits of heterosis depends on the type of crossbreeding systems used and heritability of the traits. This effect of crossbreeding, especially on the genome architecture, is still poorly understood with respect to 3-breed crossbreeding systems. Thus, this study examined variation in genomic ancestry estimations relative to pedigree-based estimations and correlated breed composition to key production and health traits. Two rotational crossbred populations, referenced as ProCROSS and Grazecross were assessed and totaled 607 crossbred cattle. ProCROSS is a product of rotational crossbreeding of Viking Red (VKR), Holstein (HOL), and Montbeliarde (MON). In contrast, Grazecross consists of Viking Red (VKR), Normande (NOR), and Jersey (JER). Both breeding programs were aimed at capitalizing on the positive effect of heterosis. The VKR is a marketing term for Swedish Red, Danish Red, and Finnish Ayrshire breed which complicated breed determination. Therefore, genomic breed composition estimates were compared using two different representations of VKR, one of which was based on parents used in the crossing system and a second based on genotypes from the ancestral breeds that comprise VKR. Variation of breed composition estimates were assessed between pedigree and genome-based predictions. Lastly, Genomic estimations were correlated with production and health traits by comparing extreme performance groups to identify the relationship between breed ancestry and performance. With the exception of the JER breed composition in Grazecross, all other estimates of the purebred contribution to the ProCROSS and Grazecross showed a significant difference in their genomic breed estimation when using the VKR ancestral versus the VKR parental reference populations for admixture analysis. These observations

were expected given the different relationship of each VKR representation to the crossbred cattle. Further analysis showed that regardless of which VKR reference population was used, the degree of MON and HOL breed composition plays a significant role in milk and fat production in ProCROSS, while the degree of VKR and NOR ancestry were related to improved health performance in Grazecross. In all, identifying the most appropriate and informative animals to use as reference animals in admixture analysis is an important factor when interpreting results of relationship and population structure, but some degree of uncertainty exists when assessing the relationship of breed composition to phenotypic performance.

KEYWORDS

crossbreed, dairy cattle, admixture, breed ancestry, ProCROSS, grazecross, viking red

1 Introduction

Within domesticated animals, "pure" breeds have been developed and are a recognized population of individuals displaying specific attributes. A purebred animal is bred from parents of the same breed and is expected to inherit characteristics attributed to the breed that are likely under extreme selective pressure or possibly fixed within the breed, ensuring the propagation of these breed-specific traits in future generations. In contrast, crossbred animals, or animals bred from parents of differing breeds, will exhibit a combination of characteristics from both parental breeds. Intentional breeding of purebred or crossbred animals has inherent challenges and advantages with respect to offspring expressing the desired traits of mated individuals and managing inbreeding depression. To this end, much research has been invested in studying the development of pure breeds and crossbreds.

Globally, dairy cattle have undergone natural and artificial selection with varying degrees of selective pressure resulting in many well recognized pure breeds of cattle and less characterized, but often environmentally adapted, indigenous or admixed cattle populations (Vanraden and Sanders, 2003; Stella et al., 2010; Zhao et al., 2015; Gautason et al., 2021). While the intense selection of purebred cattle has created breeds exceptionally known for milk production like the Holstein breed, milk solids in the Jersey, or superior health traits in the Norwegian Red breed, it has also led to inbreeding depression (Pryce et al., 2014; Gurgul et al., 2016; Dechow and Hansen, 2017: Gautason et al., 2021). Crossbreeding is commonly seen as a method of producing high production animals well-adapted to local environments and a tool to mitigate inbreeding depression (Mbole-Kariuki et al., 2014; Leroy et al., 2016).

In particular, the Holstein breed plays an important role in the dairy industry as a pure breed, noted for its exceptional production and its contribution to many crossbreeding programs similarly aiming to increase production (Vanraden and Sanders, 2003). Despite the breed's notoriety for superior production, it was also noted for its declining reproductive performance up until approximately 2010 (VanRaden, 2017; Berry, 2018). Prior to 2010, fertility was inadvertently selected against due to its negative correlation to production traits under extreme selection, with poor fertility likely exacerbated by inbreeding depression (Berry et al., 2014). Inbreeding depression is the consequence of accumulating deleterious mutations inherited from common ancestors in the lineage and is often expressed when in a homozygous state. As intensive selection propagates homozygosity across the genome to stabilize trait selection within the breed, crossbreeding increases genomic diversity with resulting heterosis or hybrid vigor in offspring. Extensive experimentation and research have been and continue to be conducted in the dairy industry to explore the benefits and drawbacks of crossbreeding, including optimization of breeds to cross and the identification of breed influence on performance (Touchberry, 1992; McAllister et al., 1994; Heins et al., 2006a).

Multiple crossbreeding schemes have been explored with differing use of breeds as well as the number of breeds used in a program and how many generations crossbreeding continoues. Another common consideration is whether to continue breeding among the crossbred animals generated or to breed future generations to purebred animals. One such crossbreeding system found to monopolize on heterosis is a 3-breed rotational crossbreeding system using complementary breeds (Dechow and Hansen, 2017). The 3-breed crossbreeding rotation of complementary breeds was found to maintain the highest level of heterosis (86%) due to the relationship of bulls and cows mated being more remote. One such example of a 3-breed rotational cross known as ProCROSS, was developed and is marketed by ProCROSS International, owned by parent companies, VikingGenetics and Coopex Montbeliarde (Randers, Denmark and Roulans, France). This system starts by crossbreeding a pure Holstein (HOL) cow to a Viking Red (VKR) bull. The first generation (F1) progeny heifers are then mated to pure Montbeliarde (MON) bulls. Each successive generation of crossbred progeny is subsequently bred to bulls from HOL, VKR, and MON in a rotational pattern. The HOL, VKR, and MON were selected due to their historical trait

improvements, with the HOL breed contributing milk volume and solid content to the 3-breed rotation and the MON and VKR breeds contributing strength, health, and fertility performance. Extensive research investigating the rotation of the 3-breeds and performance comparison to pure Holstein has been conducted (Heins et al., 2006a, Heins et al., 2006b, Heins et al., 2008a, Heins et al., 2008b, Heins et al., 2010, Heins et al., 2011, Heins et al., 2012, Heins, 2019). In addition, another crossbred population, known as Grazecross, was designed for low-input grazing systems. Grazecross consists of a 3-breed rotation of Jersey (JER), crossed to Normande (NOR), crossed to VKR. The JER is well known for its milk solid content but more diminutive stature than HOL, while NOR breeds are reported to produce high milk protein content with outstanding grazing ability (Heins et al., 2012).

Of particular note in the ProCROSS and Grazecross systems is the use of VKR cattle as developed and marketed by Viking Genetics (Randers, Denmark). The VKR cattle developed by Viking Genetics in the 1980s combined the genetic improvement programs of the previously separate Swedish Red and White, Finnish Ayrshire, and Danish Red populations. The idea of including the VKR in the crossbreeding program may increase the level of heterosis in ProCROSS and Grazecross by introducing new variant that are not routinely used in the United States. However, the effect of this mixed breed on crossbred genome architecture is poorly understood. Often the unique haplotypes or gene combinations created within a purebred population are signature to the purebred breed themselves and give rise to the crossbred performance. However, crossbreeding may also break up the unique and often beneficial gene combinations in purebred breeds, possibly reducing performance potential as the benefits of heterosis depend on the type of crossbreeding systems used and traits heritability.

This study focused on understanding the impact of using different reference populations in determining ancestry in admixed populations and how this influences global genome ancestry estimates. ProCROSS and Grazecross cattle and respective ancestral reference populations were the focus of our study and provided a unique opportunity to explore the impact of reference populations by assessing two different datasets to represent VKR ancestry, one being a combination of Swedish Red (SWD), Danish Red (DNR), and Finnish Ayrshire (FAY), and the other being commercially marketed VKR bulls used as sires in the breeding program. The first objective of this study was to characterize the genomic ancestry and assess variation in results dependent upon the reference population used. We expected the commercial VKR population to give a more exact representation of breed composition for the ProCROSS and Grazecross than the VKR ancestral reference panel due to the close relationship between commercial VKR parental animals and the crossbred populations. This study's second aim was to compare breed estimation variation between pedigree records and genomic breed estimations using the two different VKR reference populations. Lastly, we explored the effect of global breed estimation and the influence of ancestry on production traits of milk volume, protein, fat, and the health trait of somatic cell score in the ProCROSS and Grazecross. The effect of reference population on the correlation between breed composition and performance was similarly assessed. Our results demonstrate how the use of specific animals representing the reference populations in an admixture study can alter results of breed composition in admixed animals. Therefore methology, including reference populations used and how this may affect interpretation, should be reviewed in research publications before considering applications to crossbreeding programs.

2 Materials and methods

2.1 Datasets

The ProCROSS and Grazecross data were from the University of Minnesota West Central Research and Outreach Center (UMN) dairy herd in Morris, Minnesota, United States. Data included 378 ProCROSS and 229 Grazecross cattle genotypes, three major production traits (Milk, Fat, and Protein Yield), and one health trait (Somatic Cell Score). Other relevant information such as sire and dam information, birth date, and recent sire breed were also included in these datasets. Both groups of admixed cattle have birthdates spanning from 2003 to 2018.

As described in the introduction, both cattle herds are produced through rotational crossbreeding, where ProCROSS cattle are generated through the rotational mating of cattle representing VKR, HOL, and MON breeds to hybrid cows. In contrast, Grazecross cattle utilize VKR, NOR, and JER. A more detailed description of the crossbreeding rotations are in Pereira and Heins (2019). To investigate the admixture of ProCROSS and Grazecross, cattle representing the purebred and admixed breeds used in the two rotational systems were needed for comparison. Table 1 summarizes the datasets used in this study with their respective numbers and source. The SNP data of all animals studied were obtained through bovine genotyping array kits, BovineHD DNA Analysis Kit (HD150K), BovineSNP50 DNA Analysis BeadChip (50K) and CLARIFIDE® 50k (ZL5) (Illumina- Neogen, Lansing, MI, United States Clarifide–Zoetis, San Diego, CA, United States). In this study, two different datasets were used to represent the VKR population which has its own admixed origin as well. The first VKR representation included the 13 bulls marketed as VKR which were directly used in the breeding program for the ProCROSS and Grazecross in the study. These bulls were predominantly SWD and DNR crosses. The genotypes of the 13 bulls were provided by Viking Genetics (Randers, Denmark).

TABLE 1 Summary of cattle samples by population, respective breeds, number of animals and genotype source.

Population	Breed	Sample (n)[a]	SNP[b] chip	Origin of the data
Admixed	ProCROSS	378	HD150K/50K	UMN[c]
	Grazecross	229	HD150K/50K	UMN[c]
Ancestral	Holstein	92	HD150K/50K/ZL5[d]	UMN[c]
	Jersey	73	HD150K/50K	Gautier et al., 2010, Illumina®
	Montbéliarde	34	HD150K/50K	Gautier et al., 2010, Illumina®
	Normande	35	HD150K/50K	Gautier et al., 2010, Illumina®
	Swedish red	23	HD150K	Upadhyay et al. (2019)
	Finnish ayrshire	27	50K	Iso-Touru et al. (2016)
	Danish red	4	50K	Viking genetics
	Viking red	13	50K	Viking genetics

[a]Samples call rate >90%.
[b]SNP, Single Nucleotide Polymorphism.
[c]UMN, University of Minnesota.
[d]Zoetis low-density chip, version 5.

The second VKR representation was the VKR's ancestral breeds consisting of four DNR, 23 SWD and 27 FAY cattle, totaling 54 animals. While the DNR was from parental individuals, the SWD and FAY genotypes were purebred populations from Sweden and Finland, respectively (Iso-Touru et al., 2016; Upadhyay et al., 2019). The HOL genotypes were similarly directly related to the two admixed populations and represented animals used for breeding in the admixed populations or purebred counterparts within the same herds. The JER, MON, and NOR genotypes were obtained from Gautier et al. (2010) and represented purebred animals from the United States (JER) and Europe (MON and NOR), respectively.

2.2 Filtering and quality control of genomic data

The total number of SNP markers available in the merged datasets for ProCROSS and Grazecross admixed populations with their ancestral populations were 156,731 and 149,030, respectively. However, the number of SNPs genotyped per animal varied considerably due to the differing densities of the genotyping platforms and potential laboratory batch effects in genotyping. In the ProCROSS dataset, 36% (n = 137) were genotyped with Bovine150K, while the rest (n = 241) were genotyped with BovineSNP50 platform. Grazecross animals were genotyped using the BovineSNP50 DNA Analysis BeadChip (50K) (n = 59) or BovineHD DNA Analysis Kit (HD150K) (n = 170). The SNP used for analysis were identified based on those common across genotyping platforms and passing quality control thresholds. Given the crossbred nature of much of the population and limited genotypes available for many of the purebred samples, imputation was not pursued. Quality control (QC) analyses for both sets of autosomal SNPs were calculated using SNP and Variation Suite (SVS) v8.x (Golden Helix, Inc., Bozeman, MT). Various thresholds for quality control measures were examined in an effort to maximize the number of SNP available for analysis while narrowing the SNPs used to ones informative for differentiating admixed and ancestry populations. This involved the evaluation of Minor Allele Frequencies (MAF) between 0.05 and 0.10 and SNP call rate between 0.6 and 0.95. The effects of each different threshold combination were assessed through Principal Component Analysis (PCA) diagrams. The selected thresholds (MAF: >0.05, SNPs call rate: >95%) resulted in tighter clustering of individuals within a population and relatively clear separation between ancestral and admixed populations (Supplementary Figure S1). In addition, the SNPs left were further pruned for linkage disequilibrium (LD) using a threshold of $r^2 > 0.75$. The SNPs were also excluded if they were unmapped to the UMD 3.1 bovine genome assembly (Zimin et al., 2009) or mapped to sex chromosomes.

2.3 Principal component analysis and estimation of fixation index

Genomic data were analyzed through PCA in SVS (Golden Helix, Inc., Bozeman, MT) with an additive model identifying the first 10 principal components. Besides identifying the SNP quality thresholds to maximize the number of SNPs used in the analysis, PCA was also used to assess population structure within admixed populations and compare to ancestral breeds. Analyses were conducted to confirm the relationship of the ancestral populations to the admixed population and identify if any substructure existed within the crossbred population. In total, six different datasets were analyzed by PCA, including two PCA for ProCROSS, two PCA for Grazecross, one PCA comparing VKR to their ancestral breeds, and one using all individuals representing

VKR (parental and ancestral) with admixed populations and other purebreds. Within the two PCA each for ProCROSS and Grazecross, the reference population used to represent VKR cattle was alternated. One PCA for each admixed population used founding breeds of Viking Red, including SWD, FAY, and DNR, whereas the other PCA for each admixed population used cattle marketed as VKR, which were sires of animals within the admixed populations. The fifth PCA compared the marketed VKR with their ancestral breeds as noted above, to characterize the sub-structure related to VKR cattle. The other PCA incorporated all animals representing VKR, including the parental sires marketed as VKR and the individuals from the ancestral breeds. This PCA was run to identify any substructure related to VKR or how VKR clustered compared to the other purebred breeds or admixed population when combined. Estimation of fixation index (F_{ST}) was based on Wright's F statistic using SVS v8.8.5 (Golden Helix, Inc., Bozeman, MT) to investigate the genetic difference between the populations.

2.4 Genomic breed composition estimation

The genomic-based breed composition was estimated from genomic data using a maximum likelihood model implemented in ADMIXTURE 1.23 software (Alexander et al., 2009). PLINK software version 1.9 (Chang et al., 2015) was used to generate data input files for ADMIXTURE. This analysis identified genomic breed composition in the admixed populations that was used to compare with pedigree estimations of breed composition. The same datasets used in PCA analysis were analyzed using unsupervised clustering analysis with different K values where K represented the expected number of genetic clusters or ancestral populations. Due to nontypical breeds being detected in the pedigree information, such as JER in ProCROSS, an additional admixture analysis was run with all animals from the PCA datasets, including both admixture populations and all potential ancestral breed populations using K = 5 and 7. Different K values were examined to see the effect of VKR's ancestral breeds in the population. In addition, admixture was run using all available individuals for each ancestral population versus a more balanced number of individuals to represent each population, thereby examining the influence of the number of individuals representing a breed on population structure. Reduced numbers of HOL and JER cattle were selected based on their breed purity through analysis against 9 other purebreds: Montbéliarde, Normande, Norwegian Red, Guernsey, Brown Swiss, Ayrshire, Braunvieh, Short Horn and Finnish Ayrshire (Upadhyay et al., 2019; Iso-Touru et al., 2016; Gautier et al., 2010) to provide a more comparable number of animals representing each ancestral breed. This analysis yielded 49 HOL and 44 JER animals with minimum 95% breed purity which were then included for admixture analysis with the admixed populations.

2.5 Comparison of genomic breed composition to performance traits

Phenotypic records were provided by Dairy Records Management Systems (Raleigh, NC) that included estimates of 305-days lactation records for Milk Yield (MY), Fat Yield (FY), Protein Yield (PY), and Somatic Cell Score (SCS). This dataset included more than 5,000 animals, including both admixed and parental populations used in the ProCROSS and Grazecross populations. However, only admixed animals with genotype data were considered and used for downstream analysis. The mean value of a trait was used for all animals with multiple recorded measurements. First, a predictive model was fit based on linear regression using linear model function in R version 4.1.1 (R Core Team, 2021). Models for each quantitative trait (Model ProCROSS and Model Grazecross) were analyzed using a base model, with covariate of genomic breed composition (GBC), breed generation (Gen) and sire breed (Sire), where the threshold for significance was considered at p-value < 0.05. Genomic breed composition (GBC) was obtained from the breed estimation produced from the Admixture analysis. Breed generation (Gen) was the generation number of the produced admix animals due to the rotational crossbreeding systems. Using this model, the effect of different explanatory variables on admixed animals' performance was investigated.

$$\boldsymbol{Model\ ProCROSS = lm\ (Trait\ (s)\ \sim\ GBC + Gen + Sire)}$$

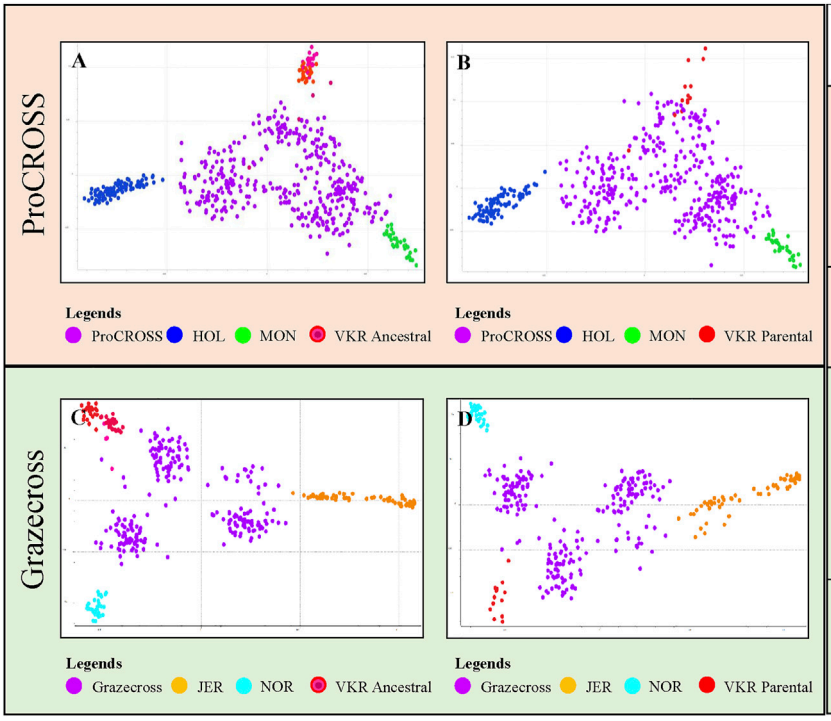
$$\boldsymbol{Model\ Grazecross = lm\ (Trait\ (s)\ \sim\ GBC + Gen + Sire)}$$

Next, we evaluated the genomic breed composition of the animals between the two extreme tails in terms of performance to determine if there was a significant difference in the breed composition of animals related to performance. To achieve this, first the phenotypic data were processed in R version 4.1.1 (R Core Team, 2021) to determine normality and opposing extreme tails. Extreme tails in this study were defined by the top 20% as high-performance animals, while the lowest 20% were low-performance animals. The normality of each group was then determined using the Shapiro Test. The association of genomic breed composition from admixture analysis with extreme performance was conducted using a Student's t-test function available in R (Chambers and Hastie, 1992). All associations between variables were assessed through the p-value produced by each test.

3 Result

3.1 Filtering and quality control of genomic data

The optimum number of informative SNP was identified by weighing various QC on the SNP common across genotyping platforms within the merged datasets. Since



PCA Diagram	Breeds Population	Total SNPs # after quality control	PC1 (%)	PC2 (%)
A	ProCROSS + HOL + MON + VKR Ancestral Breeds (SWD, FAY, DNR)	9,792	14.25	6.80
B	ProCROSS + HOL + MON + VKR Parental	9,742	14.15	5.13
C	Grazecross + JER + NOR + VKR Ancestral Breeds (SWD, FAY, DNR)	10,696	12.91	5.43
D	Grazecross + JER + NOR + VKR Parental	10,795	11.98	4.62

FIGURE 1
Principal component analysis comparing admixed populations of ProCROSS and Grazecross with their ancestral breeds. The top diagrams **(A,B)** depict ProCROSS with ancestral populations of HOL, MON, and VKR. The bottom diagrams **(C,D)** depict Grazecross with ancestral population of JER, NOR, and VKR. Left-hand diagrams **(A,C)** use the ancestral breeds of SWD, FAY and DNR to represent VKR. The right-hand diagrams **(B,D)** use commercial VKR sires used in the ProCROSS and Grazecross populations to represent VKR. The right-hand table shows the number of SNPs used for the PCA with the percentage of variation described by principal components 1 (*X*-axis) and 2 (*Y*-Axis) for each diagram.

three different types of commercial bovine SNP chips were used for genotyping crossbred and reference animals in each dataset, we expected to see a reduced number of common SNP shared across platforms and passing QC (Wang et al., 2020). Indeed, only 15,708 SNP were common across the five commercial bovine SNP chips. Stringent quality parameters (MAF: >0.05, SNPs call rate: >95%, $r^2 > 0.75$) were selected to ensure no biasness in terms of the source of the SNPs; nevertheless, the number of SNPs should be sufficient to cover and represent the majority of regions across the genome and were successful in distinguishing population structure. Comparing genotyping call rates across all the SNP in both merged datasets implied that most of the samples had less than 0.3 call rate, which agreed with the total number of samples genotyped on 50K SNP (Illumina and Clarifide©) instead of HD150K (Illumina). Thus, SNPs with less than 95% call rate were excluded, resulting in the removal of 146,897 SNP for ProCROSS and 138,183 SNP for Grazecross. Next, 38 SNPs and 45 SNPs were excluded because minor alleles were less than 0.05, while 4 SNPs and 7 SNPs were dropped due to LD more than 0.75 with other SNP for ProCROSS and Grazecross datasets, respectively. In total, 9,792 SNPs and 10,795 SNPs were retained for ProCROSS and Grazecross datasets, respectively, for downstream analysis.

3.2 Principal component analysis

3.2.1 Viking red representation

Two VKR representations were assessed to investigate the effect of differing reference populations on breed composition and ancestry relationship to performance traits in the admixed populations. The PCA analysis of the VKR with its ancestral breeds (Supplementary Figure S2) showed a high level of similarity between populations with little substructure. In all, the VKR parental animals were spread along the *x*-axis, reflecting PC1 (1.94%) and among all of the ancestral breeds. Principal component 2 (0.97%) differentiated the DNR and one VKR parental animal from the others. Nonetheless, the different representations of VKR in the datasets provided different admixture measures that were used to correlate with the performance traits. The PCA plots in Figure 1 demonstrated the effect of two different VKR datasets on ProCROSS and Grazecross. The additional PCA analysis combining both VKR representations mirrored the same general clustering of the previous analyses (Supplementary Figure S3). All

TABLE 2 Pairwise genetic differentiation (F_{ST}) value for all populations used in the ProCROSS and Grazecross breeding programs.

Population	ProCROSS	NOR	MON	JER	HOL	Ancestral VKR	Parental VKR
ProCROSS							
NOR	0.0706						
MON	0.0442	0.1158					
JER	0.0956	0.1350	0.1469				
HOL	0.0511	0.1272	0.1467	0.1630[a]			
Ancestral VKR	0.0375	0.0977	0.1143	0.1336	0.1131		
Parental VKR	0.0488	0.0938	0.1116	0.1267	0.1124	0.0216[b]	
Population	**Grazecross**	**NOR**	**MON**	**JER**	**HOL**	**Ancestral VKR**	**Parental VKR**
Grazecross							
NOR	0.0501						
MON	0.0826	0.1163					
JER	0.0605	0.1495	0.1617				
HOL	0.0881	0.1279	0.1474	0.1761[a]			
Ancestral VKR	0.0378	0.0981	0.1156	0.1510	0.1137		
Parental VKR	0.0463	0.0943	0.1123	0.1422	0.1135	0.0218[b]	

[a]The highest F_{ST}, value within each admixed population.
[b]The lowest F_{ST}, value within each admixed population.

PCA were completed using different sets of autosomal SNPs identified after QC optimization, as listed in Figure 1.

Limited genetic sub-structure was revealed between populations in both admixed population PCA, further validated by the pairwise F_{ST} value (Table 2). In the ProCROSS population, the lowest genetic differentiation was observed between the two VKR representations, which is 0.0216 while the highest genetic differentiation was between HOL and JER with 0.1630. The same observation with the same population pair was shown in the Grazecross population with the lowest of 0.0218 and the highest of 0.1761, respectively. Comparing both VKR representations showed that parental VKR has a closer relationship with ProCROSS and Grazecross populations than ancestral VKR as expected.

3.2.2 Population structure of ProCROSS and Grazecross populations

The PCA (Figure 1) for both populations revealed the expected distribution for each population cluster where the admixed population was at the center of the PCA while parental sources surrounded the admixed population in a triangle-like distribution for a three-crossbreed rotation. In the ProCROSS PCA, the highest component yielded 14.25% total variation, separating HOL and MON; the second component segregated the VKR population and yielded 6.58% total variation when using the ancestral VKR (Figure 1A) and 5.13% total variation when using the parental VKR (Figure 1B). Less variation was expected when comparing the parental VKR given their direct relationship to the ProCROSS. Some VKR individuals, regardless of which group representation, clustered with the ProCROSS, contributing to lower variation in the second component and demonstrating the higher degree of similarity between the VKR and the ProCROSS in general. In the Grazecross PCA, the highest component accounted for 12.91% of the total variation, separating JER and NOR; the second component accounted for 5.43% variation using the ancestral VKR (Figure 1C) and 4.62% variation using the parental VKR (Figure 1D), with a clear separation between VKR and Grazecross. Overall, similar PCA clustering can be observed in the ProCROSS and Grazecross PCA with clear sub-clustering of the admixed populations, especially in Grazecross. Since both admixed populations were developed from continuous rotational crossbreeding, the admixed individuals sub-cluster based on the most recent breed of sire (Supplementary Figure S4). The dominant ancestry or breed is determined by the most recent purebred sire from the 3 major breeds. This is because, without considering the generation number to produce the admixed animals, the most recent purebred sire will contribute at least 50% to the offspring's breed composition.

3.3 Breed composition

3.3.1 Pedigree based

In general, both ProCROSS and Grazecross admixed animals have proportions of ancestral breeds between 23.4% and 34%, which was expected with the 3 breeds crossbreeding system

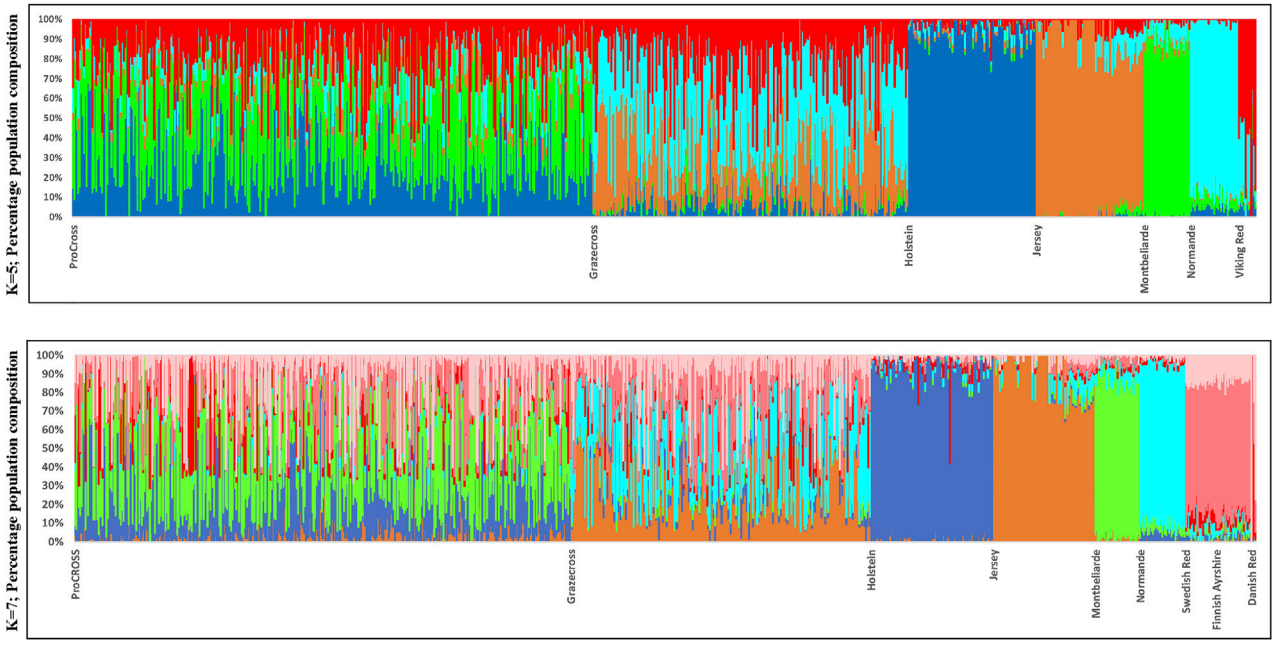


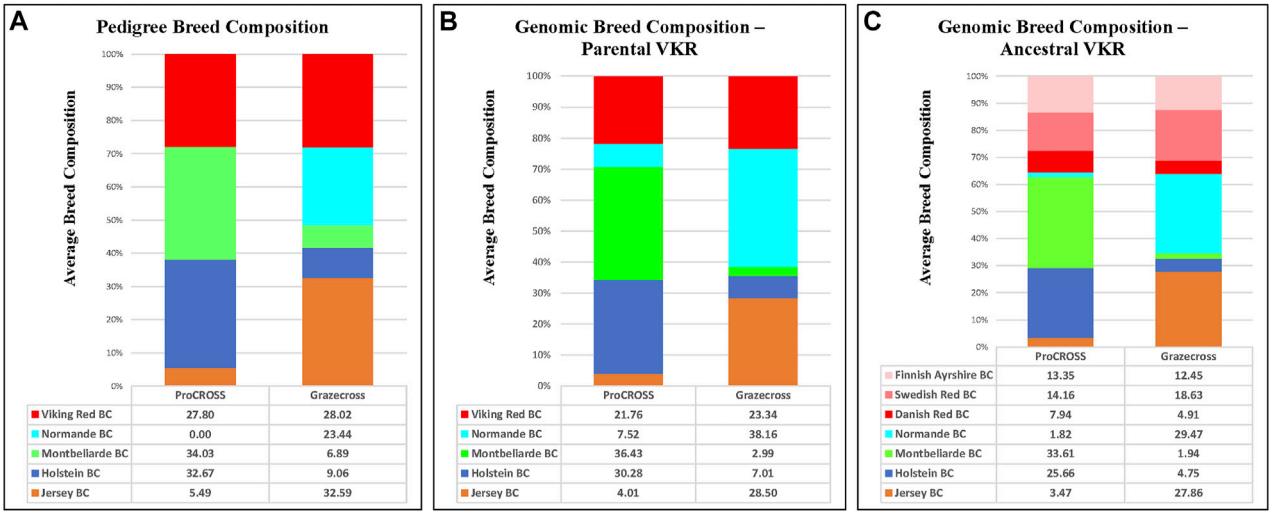
FIGURE 2
Admixture analysis reflecting genetic clustering of breeds for both ProCROSS and Grazecross admixed and all parental populations for $K = 5$ and $K = 7$. Individual vertical bars along the *X*-axis represent individual cattle grouped by breed; *Y*-axis provides a measure of the composition of each genetic population found within individuals.

(Figures 2, 3). Ideally, three subgroup signatures were considered to associate with the typical ancestry sources in each admixed population. However, nontypical breeds were identified in some of the admixed individuals based on pedigree information. For example, JER is not a typical breed component in ProCROSS populations, while HOL and MON are atypical in Grazecross. Due to that reason, a slightly lower estimation of the major breeds in the Grazecross population was due to two nontypical breeds detected at marginally higher levels in the pedigree information. A higher HOL breed composition of 9.0% was estimated because the foundation animal of the Grazecross population is purebred HOL (Pereira and Heins, 2019). Recent sire breed plays a major role in determining admixed individual's breed composition. Due to the same effect of foundation animal, ProCROSS admixed animals with HOL as the recent sire breed were estimated to have the HOL breed proportion between 57.8% and 62.5%, while the other two recent sire sources of MON and VKR were estimated between 50.0% and 57.0%.

3.3.2 Genomic based breed composition with admixture

Admixture results were used to estimate genomic breed composition for ProCROSS and Grazecross individuals. Admixture analysis was run with $K = 5$ and $K = 7$ in an unsupervised clustering to assess overall population structure, comparing ProCROSS or Grazecross and their ancestral populations, again, interchanging representation of the VKR group as shown in Figure 2 (Top) was the result from admixture analysis $K = 5$ where VKR were the commercial parental source animals, while Figure 2 (Bottom) used VKR's ancestral breeds. A clear distinction could be seen for most purebred populations, which shows greater homogeneity and less admixture within the HOL, JER, MON, and NOR. In contrast, the SWD and FAY shared the same genetic signature, excluding the DNR breed signature representation.

In slight contrast, in the admixture analysis (Supplementary Figure S3) combining both VKR and its ancestral populations along with all other breeds and both admixed populations, a higher degree of admixture was observed within individuals of the commercial VKR, SWD, and FAY, whereas DNR still produced its own unique genetic signature. Using this same combined dataset of all individuals, admixture results were used to generate average genomic-based breed composition (Figure 3). All nontypical breeds in both admixed populations were detected at low levels within the admixed individuals. Depending on which VKR datasets were used, JER composition was estimated between 3.5 and 4.0% in ProCROSS, while MON ranged from 2.0% to 3.0% in Grazecross. However, a slightly higher nontypical HOL composition was detected in Grazecross, ranging between

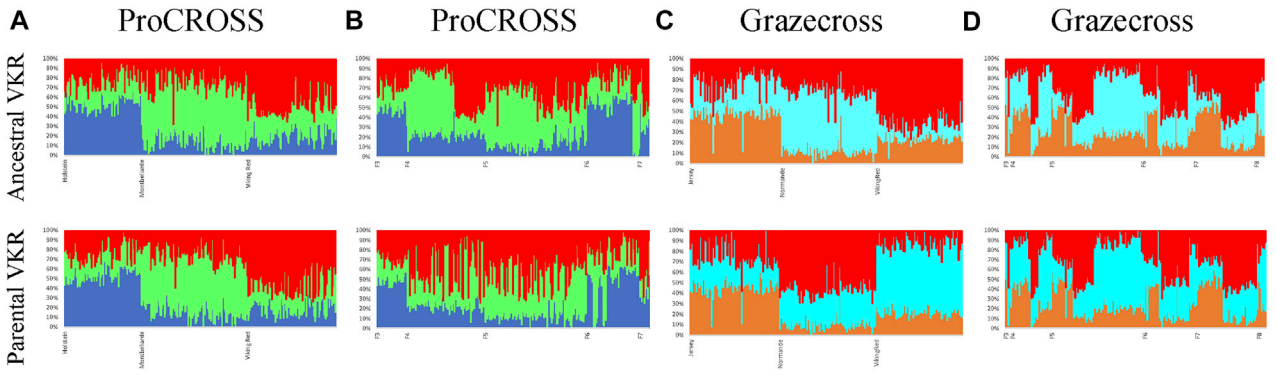


	ProCROSS	Grazecross
Viking Red BC	27.80	28.02
Normande BC	0.00	23.44
Montbellarde BC	34.03	6.89
Holstein BC	32.67	9.06
Jersey BC	5.49	32.59

	ProCROSS	Grazecross
Viking Red BC	21.76	23.34
Normande BC	7.52	38.16
Montbellarde BC	36.43	2.99
Holstein BC	30.28	7.01
Jersey BC	4.01	28.50

	ProCROSS	Grazecross
Finnish Ayrshire BC	13.35	12.45
Swedish Red BC	14.16	18.63
Danish Red BC	7.94	4.91
Normande BC	1.82	29.47
Montbellarde BC	33.61	1.94
Holstein BC	25.66	4.75
Jersey BC	3.47	27.86

FIGURE 3
Average breed composition using pedigree or genomic data; parental or ancestral VKR representation. **(A)** Pedigree based breed composition **(B)** Genomic based breed composition with commercial VKR as reference panel **(C)** Genomic based breed composition with ancestral VKR as reference panel.



Sire Breed	Animal Number
Holstein*	73
Montbeliarde*	104
Viking Red*	87

Generation Number	Animal Number
F3	34
F4*	84
F5*	97
F6	42
F7	7

Sire Breed	Animal Number
Jersey	51
Normande*	56
Viking Red*	51

Generation Number	Animal Number
F3	5
F4	24
F5	56
F6	31
F7	39
F8	3

*Significantly different (*p-value* <0.05) between ancestral VKR and parental VKR

FIGURE 4
Admixture analysis for ProCROSS and Grazecross individuals using all major parental populations at $K = 3$. The top row used ancestral VKR (FAY, DNR, and SWR) as the reference population whereas the bottom row used parental VKR as the reference population. Columns **(A,C)** ordered admixed individuals across the *X*-axis based on their recent sire while columns **(B,D)** ordered animals based on the generation of birth to better view the impact of these variables on the admixed populations. Tables within each column describe the distribution of admixed individuals based on assigned variables.

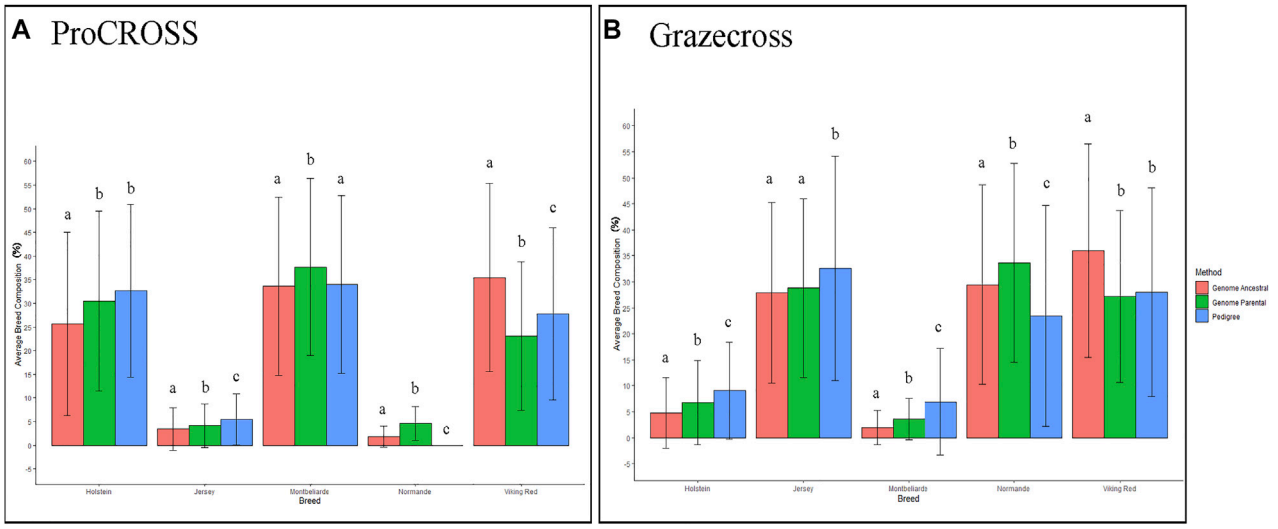


FIGURE 5
Comparison of average breed composition based on pedigree and genomic breed estimations comparing VKR population influence. Barplot A is for ProCROSS datasets and Barplot B is for Grazecross dataset. *X*-axis represent the average breed composition; *Y*-axis provides the percentage of breed composition. A *t*-test is used to compare between the methods/datasets the change in breed composition average. Error bars use one standard deviation and letters indicate significant difference between estimations with p-value < 0.05 within breed composition.

4.8 and 7.0%. These results coincide with expectations of all the admixed individuals used in both programs originating from HOL (Pereira and Heins, 2019) and pedigree breed composition estimates. Figure 3 also suggests that the usage of ancestral VKR as the parental panel may overestimate the VKR composition in the admixed population. Collectively, ancestral VKR representation is estimated to contribute 35.5 and 36% in ProCROSS and Grazecross, respectively. This estimation is significantly different (p-value < 0.05) from the estimation produced from using the commercial VKR as the parental panel, which is 21.8 and 23.3% for ProCROSS and Grazecross populations. This difference in VKR composition influenced the estimation of other breed compositions. Except for Jersey composition in Grazecross, the *t*-test analysis revealed significant differences (p-value < 0.05) in all breed composition estimates between the two VKR representation datasets (Figure 3).

The same observation could also be seen in genomic breed composition, in which the recent sire breed has the most extensive composition in the admixed population. The magnitude of the effect of the sire breed was analyzed in ADMIXTURE software. Figure 4 was the admixture plot produced based on $K = 3$, representing the three main breeds in each admixed population ordered based on recent sire breeds (Figures 4A,C) and breed generation information (Figures 4B,D). The admixture plots arranged based on sire breeds showed three genetic patterns across both the admixed populations as followed in Supplementary Figure S4. Factoring in the VKR representation showed a significant difference in the major breed composition estimation except for JER. The rotational systems used for these research herds produced a mixed individual developed from multiple generations of crossbreeding. The ProCROSS population consisted of offspring from the F3 generation up to the F7 generation while the Grazecross population consisted of offspring from the F3 generation up to the F8 generation. Despite this, very little sub-clustering was observed based on order by generation as opposed to order by recent sire. This result may have been confounded by some of the admixed individuals in this dataset undergoing different rotations of sire breed (for example, HOL-MON-VKR versus HOL-VKR-MON) for the purpose of other research objectives. Still, a comparison of the breeds estimation produced between the two VKR representations showed a small effect on the breed estimation when considering generation where F4 and F5 individuals in ProCROSS showed a significant difference. No significant difference in breed composition was observed in the Grazecross individuals in the aspect of generation number.

The reliability of pedigree-estimated breed composition can be compromised by missing, inaccurate, or incomplete records (Akanno et al., 2018). The pedigree composition was compared with estimates of genomic breed composition produced from the Q matrix in unsupervised mode; genomic breed composition offers more robust result estimations (Chiang et al., 2010). Factors such as a study population composed of only unrelated animals, adequate representation of all ancestral breeds, and low levels of linkage equilibrium between markers improve genomic breed-composition accuracy produced in the Q matrix (Alexander et al., 2009). Therefore, pedigree and

TABLE 3 p-value of comparison between 20% high performance and 20% low-Performance group in both admixed populations with different VKR's datasets.

	Ancestral VKR					Parental VKR				
ProCROSS	**Trait**	**Mean performance difference**	**P(T ≤ t) two-tail**			**Trait**	**Mean performance difference**	**P(T ≤ t) two-tail**		
			HOL BC	**MON BC**	**VKR BC**			**HOL BC**	**MON BC**	**VKR BC**
	Milk yield	7428.96 (lb)*	0.00049*	0.00311*	0.73512	Milk yield	7428.96 (lb)*	0.00026*	0.01237*	0.41987
	Fat yield	266.86 (lb)*	0.00011*	0.00012*	0.87084	Fat yield	266.86 (lb)*	0.00008*	0.00025*	0.96101
	Protein yield	239.76 (lb)*	0.00155*	0.00263*	0.96903	Protein yield	239.76 (lb)*	0.00100*	0.00785*	0.74218
	Somatic cell score	2.93*	0.52640	0.35307	0.71494	Somatic cell score	2.93*	0.29760	0.78010	0.44794
Grazecross	**Trait**	**Mean performance difference**	**P(T ≤ t) two-tail**			**Trait**	**Mean performance difference**	**P(T ≤ t) two-tail**		
			JER BC	**Nor BC**	**VKR BC**			**JER BC**	**Nor BC**	**VKR BC**
	Milk yield	6609.60 (lb)*	0.31386	0.70962	0.58421	Milk yield	6609.60 (lb)*	0.35468	0.50522	0.90795
	Fat yield	253.75 (lb)*	0.22590	0.64534	0.10836	Fat yield	253.75 (lb)*	0.21537	0.73295	0.13973
	Protein yield	217.50 (lb)*	0.42124	0.85183	0.57467	Protein yield	217.50 (lb)*	0.45279	0.67317	0.82319
	Somatic cell score	2.96*	0.52466	0.00060*	0.00325*	Somatic cell score	2.96*	0.61365	0.00090*	0.00231*

Significant difference with p-value < 0.05.

10.3389/fgene.2022.910998

genomic breed estimations were compared across admixed individuals using the two different VKR representations.

Figure 5 categorizes the variation in breed estimations for each pure breed used in the analysis, distinguishing pedigree, VKR parental, and VKR ancestral populations. For every breed found in ProCROSS or Grazecross, except for JER in Grazecross, there was a significant difference (p-value < 0.05) in average breed estimation dependent upon the use of either the ancestral VKR or the parental VKR. However, results were not as straightforward when comparing the two genomic estimates to pedigree estimates, with sometimes one or the other VKR representation causing changes in estimates or being more similar to pedigree estimates. In ProCROSS, the use of ancestral VKR resulted in a comparable MON composition but reduced the HOL composition and increased the VKR composition as compared to the pedigree estimation. In Grazecross, ancestral VKR also caused an increase in both VKR and NOR composition while decreasing the JER composition based on the p-value produced from the t-test. On the other hand, the commercial parental VKR estimated a closer HOL and VKR composition with the pedigree estimation in ProCROSS and Grazecross populations. The smaller datasets consisting of select HOL and JER animals that demonstrated >95% breed purity did not yield significant differences (p-value < 0.05) in breed estimations for the admixed individuals.

3.4 Phenotypic association with breed composition

The effect of recent sire breed and breed generation was assessed with the genomic breed composition. Following the PCA plot in Supplementary Figure S4, a clear sub-cluster was shown in both admixed populations. This sub-cluster can also be seen in the admixture plot of Figure 4 when the admixed individuals are ordered based on their recent sire breed. The overall plot for the four (4) major traits MY, FY, PY, and SCS versus these two variables (Supplementary Figure S5) suggested a linear relationship between them, particularly for breed recent sire. Higher MY, FY, and PY ProCROSS cattle tend to have HOL as the recent sire, whereas Grazecross cattle with lower SCS have VKR as their recent sire. In assessing breed generation, most cattle in both admixed populations with the smallest number of generations (F3) tend to produce animals with higher performance in MY, FY, and PY traits than other generations.

Both linear regression models supported the observed trends that recent breed sire has a significant effect on performance traits (p-value < 0.05) including all three major breeds for MY, FY, and PY performance in the ProCROSS population and SCS performance in Grazecross (Supplementary Figure S4). However, the generation of the admixed animal (F3, F4, F5, F6, etc) typically did not significantly affect trait performance. Only the F6 generation of ProCROSS individuals showed an association to higher performance in contrast to the F3 generation that seemed to carry this trend in the plot. Both models successfully explained a high number of variance with a R-squared value between 93.1 and 97.4%.

The t-test analysis investigated whether breed composition is significantly different in high versus low-performance groups, confirming that breed composition is indeed important to performance. Fifty-two admixed individuals in each of the two extreme tails were selected in the ProCROSS population, while 32 individuals represented each extreme group in the Grazecross population. Then, the t-test was used to evaluate the difference in average genomic breed composition using the two VKR representations. The t-test results (Table 3) suggested that the percentage of MON and HOL composition plays a significant role in MY, FY, and PY in ProCROSS, whereas VKR and NOR composition plays a significant role in Grazecross SCS. Despite the degree of certain breeds being significantly different within elite and poor performing ProCROSS and Grazecross for specific traits, there was a substantial degree of variation in breed composition within each performance group. For instance, in the high-performance ProCROSS population, the HOL composition ranged between 9.0 and 68.0%, whereas MON composition ranged between 10.0 and 75.0% in individuals. On the other hand, low-performance animals were estimated to have 3.0–68.0% HOL composition and 4.0–75.0% MON composition. The same amount of variation could be seen in the Grazecross population. High-performance Grazecross individuals were estimated to have between 3.0 and 56.0% NOR or 26.0%–78.0% VKR composition. While low-performance individuals had a similar degree of variation of 5.0%–68.0% NOR or 10.3%–80.0% VKR composition. Despite many of the previous analyses showing that VKR representation had a significant effect on breed estimation, the usage of different VKR representations did not show an effect on determining performance association with the breed composition. Both VKR representations yielded the same breeds corresponding to different traits' performance in both admixed populations, with the same ancestries being significantly associated with performance.

4 Discussion

The PCA plots in Figure 1 provided the first insight into the genomic population structure of the admixed ProCROSS and Grazecross and their three major parental breeds. Distinct separation was seen between the breeds and the admixed populations except for an overlap between VKR representation and the ProCROSS. In general, there was a greater distinction between the ancestral VKR and the ProCROSS and more overlap between the parental VKR and ProCROSS, as expected given the difference in relationship to the ProCROSS progeny. Even though VKR is the marketing term for the breed that share the same ancestry, such as SWD, FAY, DNR, and even Norwegian Red. Our selected SNPs successfully showed

to differentiate these two VKR representations. The observation was supported by F_{ST} calculated in Table 2; both ProCROSS and Grazecross showed a comparable genetic differentiation from parental VKR at 0.038, which is lower than the F_{ST} estimated using ancestral VKR (0.049 and 0.046 respectively). Although the number of individuals used to represent both VKR datasets was imbalanced, estimated F_{ST} provide intrinsic evidence of the sufficient power of the number of informative SNPs used in the analysis to discriminate between all the populations, including both VKR representations. The 229 Grazecross animals showed a higher degree of separation within their population, with 3 sub-groups reflecting recent sires. While the 378 ProCROSS had the same basic sub-structure reflecting recent sire used, there was less separation between the 3 sub-groups. This slight variation in the degree of sub-structure may result from 17 unique sires used in developing the larger population of ProCROSS animals compared to 12 sires used in the development of the smaller Grazecross population. Other considerations may be due to the ProCROSS breeds being more related than the Grazecross breeds due to the effect of their parental source. VKR is not purebred with historical crossbreeding, which may include HOL (Dechow and Hansen, 2017). In addition, Red and white HOL have been used historically in MON improvement (Heins et al., 2006a; Heins et al., 2006b). On the contrary, Grazecross parental NOR is unlikely to have contributed to the development of VKR, and JER was mostly pure until recently (Dechow and Hansen, 2017).

Due to the genomic similarities between breeds or signatures of breed mixture within individuals, the estimated genomic breed composition for purebred animals is complicated by small admixture components. Thus, not all animals designated as purebred have 100% genomic breed composition for their respective breed categories, as shown in Figure 2. This increases complexity in estimating accurate breed composition in an admixed population. The purebred populations used in this analysis included HOL, JER, MON, and NOR. The animals representing each breed had an average breed composition of 92, 90, 86, and 87%, respectively. In comparing different representative purebred animals having differing purebred breed averages, no significant effect on the admixed population breed composition estimation was seen. These findings support our assumption that any changes in the breed composition estimation are due to the different sources of the VKR population used in the analysis.

With the exception of the JER breed composition in Grazecross, all other estimates of the purebred contribution to the ProCROSS and Grazecross showed a significant difference in their estimation when using the VKR ancestral versus the VKR parental (Figure 5) groups. Subsequently, both estimations were compared to the estimation produced from pedigree information. It is noted that in Figure 3, the use of VKR parental animals in ProCROSS increased the MON origin to 36.4%, compared to the pedigree estimation of 34.0% which was a significant difference between the estimations. This could be explained by VKR and MON being selected for the same fertility and health traits for over 30 years (Heringstad et al., 2007; Dezetter et al., 2015). This observation may prove that both breeds shared comparable allele frequencies due to similar selection pressure on the same haplotype, causing increased similarity between these two breeds. Whereas the use of VKR parental in Grazecross also significantly increased NOR composition. Hence, including VKR ancestral reference breeds introduced a higher noise level in estimating purebred composition (Figure 2), offsetting the estimated composition for the ancestral breeds. The rest of the estimated genomic breed composition using the VKR ancestral datasets showed similar patterns. Except for VKR estimations, using the ancestral VKR representation decreased other breed composition within admixed individuals compared to both pedigree and genomic estimates using the parental VKR representation.

Next, the correlation between genomic breed composition and four traits was investigated. Both linear regression models suggested that all three major breeds significantly affect the MY, FY, and PY production traits in ProCROSS and health trait SCS trait in Grazecross as denoted in Supplementary Figure S5. This observation was further explored by comparing the extreme performance groups of each trait. Surprisingly, the results shown in Table 3 indicate no significant difference in ancestry importance for performance when comparing the two sources of VKR (p-value < 0.05). Both VKR representations identified the same breeds as being significantly different between performance groups and at similar levels within the elite or poor performance groups. As expected, HOL and MON composition were significantly higher in the animals with higher performance in MY, FY, and PY. HOL is known to have superior performance in milk yield, whereas MON was improved for milk solids to produce speciality cheese, resulting in a higher emphasis on the fat and protein yield traits. Thus, both of these breeds contribute to the high production performance of the admixed population. In contrast, in Grazecross, only NOR and VKR showed a significant difference in composition related to SCS. The Grazecross admixed population was developed to cater to demand for efficient high performing animals in low input grazing environments. Thus, different traits and breeds were used for this admixed population. NOR was developed in Northwestern France, a region known to have pasture. This environment produced a breed with exceptional feed conversion rates, making it a great genetic source for developing low grazing crossbred cattle (Dillon et al., 2003). The same study showed that NOR has a substantially higher survival rate than three dairy cattle breeds, including Dutch Holstein, Irish Holstein, and Montbeliarde. In addition, VKR was developed to have higher fertility and health performance (Heringstad et al., 2007). This complements our finding that high-performance Grazecross animals in SCS have a higher NOR and VKR composition. However, we did not see the same complementary effect on SCS performance in ProCROSS with regards to their VKR ancestry. One thought potentially explaining this variation in VKR correlation to SCS is the

influence of JER ancestry within Grazecross that is not present in ProCROSS. The JER breed is known to have higher SCS compared to other dairy breeds in the USA (Dechow and Hansen, 2017). Thus, as NOR and VKR breed proportions increase in Grazecross, the JER breed proportion is decreasing which may contribute to the reduction in SCS even though JER were not statistically associated with this trait. It may be the unique combination of the different breeds used in each of these 3-breed rotations that alters the degree of impact of any breed on a performance trait. Furthermore, individually assessing the animals within the two extreme groups showed a dispersed distribution of the identified breeds. For instance, higher HOL composition cattle have the highest chance of producing high-performance production animals, yet some elite production animals only had 4% HOL composition and vice versa in the low-performance animals. These observations may suggest that some haplotypes at specific locations have a large effect on determining the performance of these traits. For instance, the elite 4% HOL ProCROSS suggests that they possess HOL ancestry at key genomic regions most influential on production and additional HOL ancestry is not actually required to be elite. Therefore, the association between increased performance and increased HOL breed ancestry could be based on the fact that increased HOL ancestry increases the chance that an animal possesses HOL ancestry at the most important regions of the genome as seen in the 4% HOL ProCROSS and the remainder of the HOL ancestry is incidental to their performance. Key haplotypes may be directly linked and preserved in particular ancestral breeds used in the crossbreeding program. Identifying the existence and influence of breed-specific haplotypes in future studies may leverage breed complementary and accelerate selection through more precise and accurate genomic prediction. In all, the correlation analysis was also able to capture the effect of heterosis on the performance of both admixed populations. Interestingly, the linear regression model suggested that individuals developed from generation six (two round of 3 breed rotational crossbreeding systems) significantly capitalize on heterosis similar with what we expected from generation three (One round of 3 breed rotational crossbreeding systems). This observation may serve as additional evidence of the advantages provided by a 3-breed rotational crossbreeding system to retain the highest heterosis level in the crossbreeding population.

5 Conclusion

The main inquiry of this project was to identify the impact of different animal groups serving as the reference population when assessing admixture in crossbred cattle. The results showed differences in the estimated breed composition of both ProCROSS and Grazecross individuals when two different VKR reference panels were used. Interestingly, the breed composition estimates from admixture were significantly different from pedigree estimates but the variation sometimes diverged, with one panel having an increased estimate of a particular breed when the other reference panel would show a lower estimate of that same breed as compared to the pedigree estimate. Surprisingly, despite the two different VKR references panels providing significantly different breed compositions of ProCROSS and Grazecross, we found no significant difference in the relationship of breed composition to performance traits as HOL, MON, NOR, and VKR were all similarly important to ProCROSS and Grazecross performance. It is important to identify the most appropriate and informative animals to use as reference animals in admixture analysis to correctly interpret relationship and population structure results. Any application of admixture output in directing dairy cattle crossbreeding strategies should proceed with caution depending upon the reference populations used to prevent over- or under-interpretation of the contribution and impact of ancestry breeds.

Data availability statement

1) The data presented in the study are deposited in the University of Minnesota Digital Conservancy repository, https://hdl.handle.net/11299/227134; 2) The data is uploaded and restricted for 2 years, which means if people want to download, they have to send the data owner's (Bradly J. Heins) an email to request access. The data will open access after the 2 years per UM.

Ethics statement

Ethical review and approval was not required for the animal study because the genomic data used in this study was generated for other projects in which the original IACUC approval was obtained for research. This study only used the genomic data and did not include any sampling of animals. Approval documentation for the original sampling can be supplied if necessary.

Author contributions

Project conception and development was a collaborative effort among all authors. HH initiated the original project. MJ compiled data, conducted all genetic analysis. MJ and HH drafted the manuscript. BH and CD supplied majority of the datasets used and background information on the ProCROSS and Grazecross. All authors contributed to manuscript review.

Funding

This work is supported by Organic Agriculture Research and Extension Initiative [grant no. 2016-51300-25862/project

accession no. 1010366] from the USDA National Institute of Food and Agriculture.

Acknowledgments

The authors thank Viking Genetics (Randers, Denmark) for graciously providing genotypes of Viking Red bulls used in this research herd.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.910998/full#supplementary-material

References

Akanno, E. C., Abo-Ismail, M. K., Chen, L., Crowley, J. J., Wang, Z., Li, C., et al. (2018). Modeling heterotic effects in beef cattle using genome-wide SNP-marker genotypes1. *J. Anim. Sci.* 96 (3), 830–845. doi:10.1093/jas/skx002

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109

Berry, D. P. (2018). Symposium review: Breeding a better cow-Will she be adaptable? *J. Dairy Sci.* 101 (4), 3665–3685. doi:10.3168/jds.2017-13309

Berry, D. P., Wall, E., and Pryce, J. E. (2014). Genetics and genomics of reproductive performance in dairy and beef cattle. *animal* 8 (1), 105–121. doi:10.1017/S1751731114000743

Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S. M., and Lee, J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaSci* 4, 7–015. doi:10.1186/s13742-015-0047-8

Chiang, C. W., Gajdos, Z. K., Korn, J. M., Kuruvilla, F. G., Butler, J. L., Hackett, R., et al. (2010). Rapid assessment of genetic ancestry in populations of unknown origin by genome-wide genotyping of pooled samples. *PLoS Genet.* 6, e1000866. doi:10.1371/journal.pgen.1000866

Dechow, C. D., and Hansen, L. B. (2017). "Capitalizing on breed differences and heterosis," in *Large dairy herd management. 3rd ed* (Champaign: American Dairy Science Association), 369–378. doi:10.3168/ldhm.0526

Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D., and Ducrocq, V. (2015). Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. *J. Dairy Sci.* 98 (7), 4904–4913. doi:10.3168/jds.2014-8386

Dillon, P., Buckley, F., O'Connor, P., Hegarty, D., and Rath, M. (2003). A comparison of different dairy cow breeds on a seasonal grass-based system of milk production. *Livest. Prod. Sci.* 83 (1), 21–33. doi:10.1016/s0301-6226(03)00041-1

Gautason, E., Schönherz, A. A., Sahana, G., and Guldbrandtsen, B. (2021). Genomic inbreeding and selection signatures in the local dairy breed Icelandic Cattle. *Anim. Genet.* 52 (3), 251–262. doi:10.1111/age.13058

Gautier, M., Laloë, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PloS one* 5, e13038. doi:10.1371/journal.pone.0013038

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K., and Bugno-Poniewierska, M. (2016). The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *J. Appl. Genet.* 57 (4), 527–530. doi:10.1007/s13353-016-0337-6

Heins, B. J. (2019). "Opportunities and challenges in crossbreeding dairy cattle in temperate regions" in *Pryce, J., Advances in breeding of dairy cattle* (Cambridge, UK: Burleigh Dodds Science Publishing). doi:10.19103/as.2019.0058.06

Heins, B. J., Hansen, L. B., De Vries, A., and De Vries, A. (2012). Survival, lifetime production, and profitability of Normande × Holstein, Montbéliarde × Holstein, and scandinavian red × Holstein crossbreds versus pure holsteins. *J. Dairy Sci.* 95 (2), 1011–1021. doi:10.3168/jds.2011-4525

Heins, B. J., Hansen, L. B., Hazel, A. R., Seykora, A. J., Johnson, D. G., and Linn, J. G. (2010). Birth traits of pure Holstein calves versus Montbeliarde-sired crossbred calves. *J. Dairy Sci.* 93 (5), 2293–2299. doi:10.3168/jds.2009-2911

Heins, B. J., Hansen, L. B., and Seykora, A. J. (2006b). Calving difficulty and stillbirths of pure holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and scandinavian red. *J. Dairy Sci.* 89 (7), 2805–2810. doi:10.3168/jds.S0022-0302(06)72357-8

Heins, B. J., Hansen, L. B., and Seykora, A. J. (2006a). Production of pure holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and scandinavian red. *J. Dairy Sci.* 89 (7), 2799–2804. doi:10.3168/jds.S0022-0302(06)72356-6

Heins, B. J., Hansen, L. B., Seykora, A. J., Hazel, A. R., Johnson, D. G., and Linn, J. G. (2008b). Crossbreds of Jersey × Holstein compared with pure holsteins for body weight, body condition score, dry matter intake, and feed efficiency during the first one hundred fifty days of first lactation. *J. Dairy Sci.* 91 (9), 3716–3722. doi:10.3168/jds.2008-1094

Heins, B. J., Hansen, L. B., Seykora, A. J., Hazel, A. R., Johnson, D. G., and Linn, J. G. (2011). Short communication: Jersey × Holstein crossbreds compared with pure holsteins for production, mastitis, and body measurements during the first 3 lactations. *J. Dairy Sci.* 94 (1), 501–506. doi:10.3168/jds.2010-3232

Heins, B. J., Hansen, L. B., Seykora, A. J., Johnson, D. G., Linn, J. G., Romano, J. E., et al. (2008a). Crossbreds of Jersey × Holstein compared with pure holsteins for production, fertility, and body and udder measurements during first lactation. *J. Dairy Sci.* 91 (3), 1270–1278. doi:10.3168/jds.2007-0564

Heringstad, B., Klemetsdal, G., and Steine, T. (2007). Selection responses for disease resistance in two selection experiments with Norwegian red cows. *J. Dairy Sci.* 90 (5), 2419–2426. doi:10.3168/jds.2006-805

Iso-Touru, T., Tapio, M., Vilkki, J., Kiseleva, T., Ammosov, I., Ivanova, Z., et al. (2016). Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim. Genet.* 47 (6), 647–657. doi:10.1111/age.12473

Leroy, G., Baumung, R., Boettcher, P., Scherf, B., and Hoffmann, I. (2016). Review: Sustainability of crossbreeding in developing countries; definitely not like crossing a meadow. *Animal* 10 (2), 262–273. doi:10.1017/s175173111500213x

Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. D. C., Kiara, H., et al. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* 113 (4), 297–305. doi:10.1038/hdy.2014.31

McAllister, A. J., Lee, A. J., Batra, T. R., Lin, C. Y., Roy, G. L., Vesely, J. A., et al. (1994). The influence of additive and nonadditive gene action on lifetime yields and profitability of dairy cattle. *J. Dairy Sci.* 77 (8), 2400–2414. doi:10.3168/jds.S0022-0302(94)77183-6

Pereira, G. M., and Heins, B. J. (2019). Activity and rumination of Holstein and crossbred cows in an organic grazing and low-input conventional dairy herd. *Transl. Animal Sci.* 3 (4), 1435–1445. doi:10.1093/tas/txz106

Pryce, J. E., Haile-Mariam, M., Goddard, M. E., and Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet. Sel. Evol.* 46 (1), 71–14. doi:10.1186/s12711-014-0071-7

R Core Team (2021). R: A language and environment for statistical computing. *R Found. Stat. Comput*. Vienna.Austria. *URL*.

Stella, A., Ajmone-Marsan, P., Lazzari, B., and Boettcher, P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185 (4), 1451–1461. doi:10.1534/genetics.110.116111

Touchberry, R. W. (1992). Crossbreeding effects in dairy cattle: The Illinois experiment, 1949 to 1969. *J. Dairy Sci.* 75 (2), 640–667. doi:10.3168/jds.s0022-0302(92)77801-1

Upadhyay, M., Eriksson, S., Mikko, S., Strandberg, E., Stålhammar, H., Groenen, M. A., et al. (2019). Genomic relatedness and diversity of Swedish native cattle breeds. *Genet. Sel. Evol.* 51 (1), 56–11. doi:10.1186/s12711-019-0496-0

VanRaden, P. M. (2017). "Genomic tools to improve progress and preserve variation for future generations," in *Book of abstracts of the 68th annual meeting of the European federation of animal science*. (Vol. 79).

VanRaden, P. M., and Sanders, A. H. (2003). Economic merit of crossbred and purebred US dairy cattle. *J. Dairy Sci.* 86 (3), 1036–1044. doi:10.3168/jds.S0022-0302(03)73687-X

Wang, Y., Wu, X. L., Li, Z., Bao, Z., Tait, R. G., Jr, Bauck, S., et al. (2020). Estimation of genomic breed composition for purebred and crossbred animals using sparsely regularized admixture models. *Front. Genet.* 11, 576. doi:10.3389/fgene.2020.00576

Zhao, F., McParland, S., Kearney, F., Du, L., and Berry, D. P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* 47 (1), 49–12. doi:10.1186/s12711-015-0127-3

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* 10 (4), R42–R10. doi:10.1186/gb-2009-10-4-r42

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fgene.2022.909986

OPEN ACCESS

EDITED BY
Klaus Wimmers,
Leibniz Institute for Farm Animal Biology (FBN), Germany

REVIEWED BY
Este van Marle-Köster,
University of Pretoria, South Africa
Ottmar Distl,
University of Veterinary Medicine Hannover, Germany

*CORRESPONDENCE
Sowah Addo,
sowah.addo@uni-kassel.de

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

RECEIVED 31 March 2022
ACCEPTED 30 September 2022
PUBLISHED 18 October 2022

CITATION
Addo S and Jung L (2022), An insight into the runs of homozygosity distribution and breed differentiation in Mangalitsa pigs.
Front. Genet. 13:909986.
doi: 10.3389/fgene.2022.909986

An insight into the runs of homozygosity distribution and breed differentiation in Mangalitsa pigs

Sowah Addo* and Lisa Jung

Animal Breeding Section, University of Kassel, Witzenhausen, Germany

Mangalitsa pigs exhibit three distinct coat color patterns based on which they are described as Red, Blond, and Swallow-bellied. The current study investigated genome-wide diversity and selection signatures in the three breeds using fixation index, runs of homozygosity and population structure analyses. The analyses were originally based on quality-controlled data on 77 Mangalitsa animals from Germany, including 23 Blond, 30 Swallow-bellied and 24 Red Mangalitsa genotyped with a customized version of the ProcineSNP60 v2 Genotyping Bead Chip. Also, 20 Hungarian Mangalitsa genotypes were included as outgroup data for comparison. Estimates of observed heterozygosity were 0.27, 0.28, and 0.29, and inbreeding coefficients estimated based on runs of homozygosity were 24.11%, 20.82%, and 16.34% for Blond, Swallow-bellied and Red Mangalitsa, respectively. ROH islands were detected in all breeds, however, none of these were shared amongst them. The KIF16B gene previously reported to play a role in synaptic signaling was found in a ROH island (SSC17: 16–26) in Swallow-bellied Mangalitsa. The same gene was found to harbor a significantly differentiated SNP (MARC0032380) while contrasting either Blond or Red to Swallow-belied Mangalitsa. In the Red Mangalitsa, some ROH islands were associated with genes that play a role in meat quality traits, i.e., ABCA12, VIL1, PLSCR5, and USP37. Our population structure analysis highlighted a separation of the three breeds, but also showed the closest relatedness between Red and Blond Mangalitsa pigs. Findings of this study improve our understanding of the diversity in the three breeds of Mangalitsa pigs.

KEYWORDS

Mangalitsa pig, ROH islands, genetic diversity, selection signatures, runs of homozygosity

Introduction

Domestication and selection events can lead to both favorable and unfavorable allele reconstitution in animal species. Following its creation by domestication of wild pigs (*Sus scrofa ferus*) in the 19th century, Sumadija pigs of Serbia were reared under favorable conditions that made them lose their original form to become one of the progenitors of

Mangalitsa pigs (Jukes 2017). Although originally from Serbia, Mangalitsa was systematically developed in Hungary at a time when market demand for good quality fat, bacon and less fibrous meat necessitated the crossing of small Hungarian sows such as Alfoldi, Bakony and Szalonta with Serbia's improved Sumadija (Egerszegi et al., 2003; Jukes 2017). The product of such crosses was the Blond Mangalitsa, which was subsequently crossed either with Black Syrmian to develop Swallow-bellied Mangalitsa or with Szalonta to develop Red Mangalitsa (Zsolnai et al., 2013; Jukes 2017). Thus, three types of Mangalitsa pigs exist and these exhibit varying phenotypes, particularly coat color variation. In spite of the differences, Mangalitsa pigs are broadly described as fat-type, curly-haired pigs with strong motherliness and adaptability, but low reproductive performance (Egerszegi et al., 2003). They have geographical predominance in Hungary but are also distributed across Serbia, Romania, Austria, Croatia, Germany and Switzerland.

The three breeds of Mangalitsa are usually managed with restricted gene flow amongst them. Zsolnai et al. (2006) investigated the genetic relationships between the breeds using 10 microsatellite markers, and proposed a rejection of the hypothesis that Mangalitsa individual form just one unpartitioned population. They also found Blond and Swallow-bellied Mangalitsa to be genetically closer to each other than to Red Mangalitsa as did Marincs et al. (2013) who based their analysis on mitochondrial DNA D-loop sequences. Meanwhile, in a separate study, a conclusion has been drawn that mitochondrial DNA D-loop polymorphism could not distinguish between the three breeds (Molnár et al., 2013). Marincs et al. (2013) also unraveled the presence of both common European and Mangalitsa-specific mitochondrial DNA D-loop haplotypes in a Hungarian population and concluded that these pigs may have originated either by introgression of common European bloodline into the Mangalitsa breed or by isolation of some Mangalitsa ancestor species. Furthermore, Frank et al. (2017), for the first time found the mitogenomes of some Mangalitsa animals to be highly related to the Croatian Turopolje breed, which they attributed to either common origin of maternal lineages or admixture events. A comparative study of whole genome sequences of the three breeds of Mangalitsa and a Duroc pig highlighted on one hand, 52 Mangalitsa-specific genes involved in lipid metabolic processes, and on the other hand, several exonic polymorphisms unique to each of the breeds (Molnár et al., 2014). More recently, an investigation into the genetic basis of different colorations in Mangalitsa pigs revealed a display of signature of divergent directional selection in the solute carrier family 45-member 2 (SLC45A2) gene for the comparison of Red and Blond Mangalitsa breeds (Bâlteanu et al., 2021).

In spite of their genetic differentiation, there was no mentioning of breed type in a number of studies involving Mangalitsa pigs, and it was not immediately clear if animals of different coat colors where used (García et al., 2006; Wilkinson et al., 2013; Herrero-Medrano et al., 2014; Manunza et al., 2016; Schachler et al., 2020). In one of such studies, the mean genomic inbreeding coefficient estimated for 20 Hungarian Mangalitsa animals based on runs of homozygosity was high (0.22) for which reason the authors advocated for special conservation interventions to be put in place (Yang et al., 2017). Gorssen et al. (2021) recently presented a repository of ROH island for several breeds of eight animal species among which the Hungarian Mangalitsa pigs used in Yang et al. (2017) were featured. The repository shows high incidence of ROH occurring on chromosomes 11, 13, 14, and 17 in the Mangalitsa pigs. Bâlteanu et al. (2019) found runs of homozygosity based inbreeding coefficient ranging from 0.09 to 0.14 in different populations of Mangalitsa pigs and argued that the pattern of homozygosity in these local breeds is comparable to those of the majority of cosmopolitan breeds. Albeit, they found a clear indication of strong and recent inbreeding in Romanian and Hungarian Red Mangalitsa pigs which was attributed to mating of related individuals and a reduction in population size. In Germany, Managalitsa pigs are referred to as "Wollschweine", and an attempt to preserve their genetics is evidenced by the naming of Mangalitsa as breed of the year 1999 (Flegler 1999). Nowadays, there is growing interest of German breeders in breeding all three types of Mangalitsa, while the Society for the Conservation of Old and Endangered Livestock Breeds (GEH e.V.) is rapidly promoting the establishment of a herd book for the Mangalitsa pigs in Germany.

The availability of different Mangalitsa genotypes that exhibit phenotypic variability offers new possibilities of studying signatures of selection that may have played a role in the development of the three breeds. Besides, the ROH landscape of these breeds can be conveniently compared to those of other breeds published in a novel ROH repository to improve our understanding of selective sweeps in pigs. Therefore, the present study aimed at 1) investigating genome-wide relatedness among Blond, Red and Swallow-belied Mangalitsa pigs, 2) identifying genomic regions with high level of inbreeding termed ROH islands, and 3) finding candidate genes that may be associated with significantly differentiated genomic regions in the breeds.

Materials and methods

Animal description

In this study, 109 animals belonging to the three main breeds of Mangalitsa pigs were initially considered. These include Blond, Red and Swallow-bellied Mangalitsa, hereafter referred to as BM, RM, and SM, respectively. The animals were from individual breeders and animal parks, and each had a registration number provided by the GEH e.V. in Germany. Animals were distinguishable, predominantly by their coat-color variation (Figure 1). BM individuals have a general grey to yellow to yellowish red coat-color with the head and leg regions often

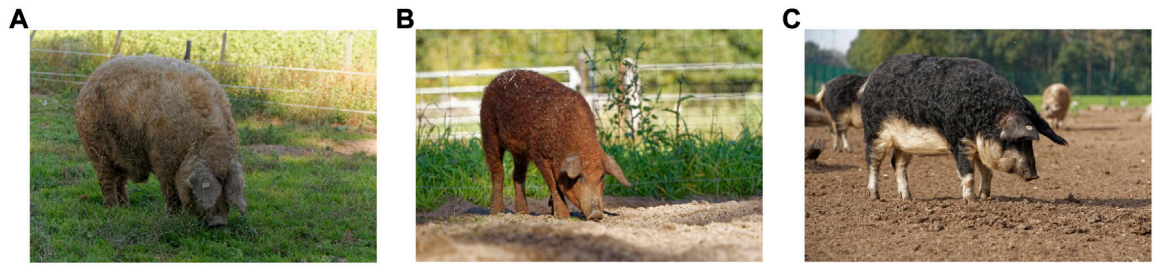


FIGURE 1
The three different breeds of Mangalitsa pigs including Blond **(A)**, Red **(B)** and Swallow-bellied **(C)**. Pictures were provided by Rudi Gosmann.

almost black (Figure 1A). RM tend to have darker or lighter shades of reddish-brown (Figure 1B) while SM have a blackish-brown coloration at the back and flanks and yellow, white or silvery grey at the underside, belly and cheek areas (Figure 1C). No extensive pedigree data were available at the time of sample collection, and a number of animals were said to have been produced by crossing two of the three breeds. Due to issues of misidentification and uncertainty, these animals, totaling 17, were labeled as mixed breed (MM) and removed from the analyses. The remaining animals included 34 SM, 29 RM, and 29 BM pigs. Additionally, we included data on 20 Hungarian Mangalitsa pigs (HUMA) of unspecified breed type from previous studies (Yang et al., 2017) for comparison.

Genotyping and quality control

The collection of hair samples, DNA extraction and genotyping followed standard procedures and were performed in two batches, in 2018 and 2020. Animals were genotyped with customized versions of the ProcineSNP60 v2 Genotyping Bead Chip. SNP markers common to both batches were extracted and mapped to the *Sus scrofa* 11.1 genome assembly. A number of quality control procedures were conducted in PLINK v 1.9 (Chang et al., 2015) depending on the desired type of analysis. Unmapped and non-autosomal SNPs were broadly removed as were 2 (SM) and 1 (RM) animals whose genomic relationship coefficient with other pairs exceeded 0.95. Individual and marker genotyping rate thresholds were both set to 90%, and SNPs with minor allele frequency (MAF) lower than 0.05 or that deviated from Hardy Weinberg Equilibrium (HWE: 10^{-6}) were removed. Specifically, for ROH analyses, SNP call rate was set to 95% for easy comparison (Gorssen et al., 2021) and there was neither MAF nor HWE pruning as recommended in previous studies (Meyermans et al., 2020). MAF pruning was also not conducted on data destined for Fixation index (F_{st}) analysis (Weir and Cockerham 1984) and the quality control steps were performed for each breed separately. Furthermore, a minor linkage disequilibrium (LD) pruning was applied to the dataset destined for population structure analysis by using the PLINK (Chang et al., 2015) command "--indep-pairwise 50 25 0.5". After all filtering steps, 23 BM, 24 RM, and 30 SM genotypes remained for further analyses.

Diversity and population structure

Within-breed genetic diversity was investigated based on observed heterozygosity estimates calculated as a difference between the number of homozygous and non-missing genotypes, expressed as a proportion of non-missing genotypes. Also, the relationship between BM, RM, and SM was investigated using principal component analyses. LD-pruned SNPs totaling 10,323 were used to compute Plink-based (Chang et al., 2015) eigenvectors for the first 20 components for each individual. Subsequently, eigenvectors of the first (PC1) and second (PC2) principal components were visualized in R (R Core Team 2020) using breed as color code. This investigation was further expanded to include the 20 HUMA outgroup data and the analysis was based on 8,624 quality-controlled SNPs common to all animals. The population structure was also studied using ADMIXRURE (Alexander et al., 2009). For this, a 5-fold cross-validation procedure was performed for a range of k between 1 and 20, and the k with the lowest cross-validation error was considered as the optimal number of clusters for the data. Cluster assignments ranging from k = 2 to k = 9 were visualized using Pophelper (Francis 2017).

To detect genetic differentiation over time, the F_{st} (Weir and Cockerham 1984) between breeds was calculated for all loci using PLINK (Chang et al., 2015). Subsequently, an empirical *p*-value for each SNP was estimated following a z-score calculated from the distribution of F_{st} values. The F_{st} values were visualized using Manhattan plots implemented in the qqman r package (Turner 2018), and the top 0.1% were suggested as signatures of selection.

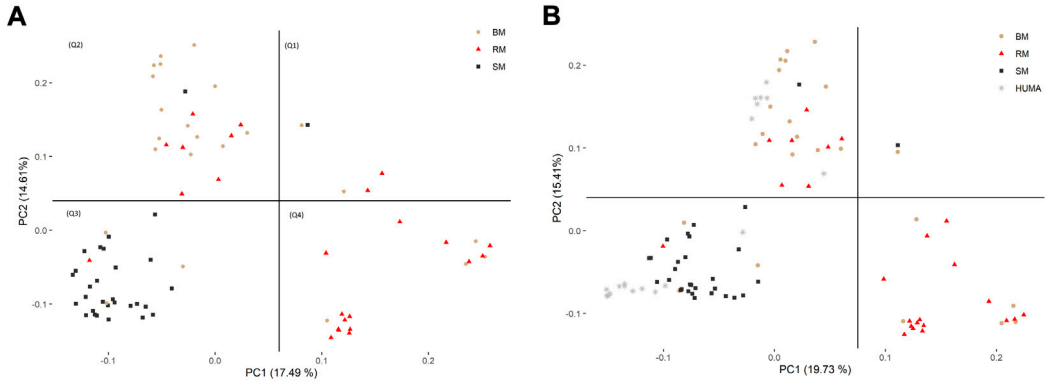


FIGURE 2
Distinguishing BM, RM and SM through principal component analysis based on 10,323 SNP markers **(A)**. The inclusion of HUMA in the analysis **(B)** was based on 8,624 SNP markers mapped to the Sus scrofa 10.2 genome assembly. The plotted area is divided into quadrants based on the occurrence of clusters and breeds are distinguished by color and shape.

ROH calling and analysis

The number of SNPs available for ROH analyses after quality control was 36,617 (BM), 38,085 (RM), and 36,964 (SM). ROH calling was performed using an R-script developed for standardized breed-by-breed quality control and analysis (Gorssen et al., 2021), which is available at Open Science Framework (https://doi.org/10.17605/OSF.IO/XJTKV). Default parameter settings as described in Gorssen et al. (2021) were followed in defining ROH, ROH incidence and ROH islands. Briefly, we considered a sliding window with minimal number of SNPs determined by an L-parameter (Purfield et al., 2012; Meyermans et al., 2020). Within this window, 1 SNP could be missing but no heterozygous SNP was allowed, and there was at least 1 SNP every 150 kb. Additionally, the maximal gap between two consecutive homozygous SNPs was set to 1,000 kb.

Genomic inbreeding coefficient (F_{ROH}) was calculated for each animal as the total length of all ROH in the genome of an individual expressed as a proportion of the length of autosomal genome coverage expressed by SNPs in the analysis (McQuillan et al., 2008). F_{ROH} was calculated considering all ROH (F_{ROH_all}) in an individual but also for different ROH length categories including 1–2 Mb ($F_{ROH1\text{-}2}$), 2–4 Mb ($F_{ROH2\text{-}4}$), 4–8 Mb ($F_{ROH4\text{-}8}$), 8–16 Mb ($F_{ROH8\text{-}16}$) and >16 Mb (F_{ROH16}).

For a given breed, the percentage of individuals with a specific SNP in ROH was defined as ROH incidence. From the distribution of ROH incidences, a threshold (*p*-value) was calculated based on standard normal z-scores. ROH islands were defined as the top 0.1% of SNPs with a *p*-value higher than 0.999 using a z-score table for ROH incidence (Purfield et al., 2017; Gorssen et al., 2020). Finally, a ROH must be present in at least 30% of the population to be part of a ROH island. ROH incidences and thresholds were visualized for each breed *via* Manhattan plots using the qqman package (Turner 2018).

Identification of candidate gene

Significant genome variants from both F_{st} and ROH analyses were annotated to the *Sus scrofa* genome version 11.1 reference assembly, and the "sscrofa_gene_ensembl" dataset was explored using biomaRt v. 2.50.3 (Durinck et al., 2009). Furthermore, genes within 100 kb distance on either side of these variants were identified as candidate genes that may have played a role in the development of the breeds.

Results

Population structure and fixation index

In the principal component analysis, PC1 and PC2 together explained 32,1% of variation in the three breeds (Figure 2A). There was a clustering along the lines of breeds, but notably, clusters were not well defined. In the third quadrant (Q3) of the plotted area, SM was predominantly separated from RM and BM by PC1 and PC2, respectively. In the second quadrant, what seemed to be a BM cluster harbored several genotypes of RM. The RM breed formed a cluster in the fourth quadrant with the highest level of dispersion. This latter cluster also harbored genotypes of BM. By including HUMA genotypes in the analysis (Figure 2B), 35.14% of the variation in the data was explained. Majority of the HUMA clustered close to SM, and a few were in the

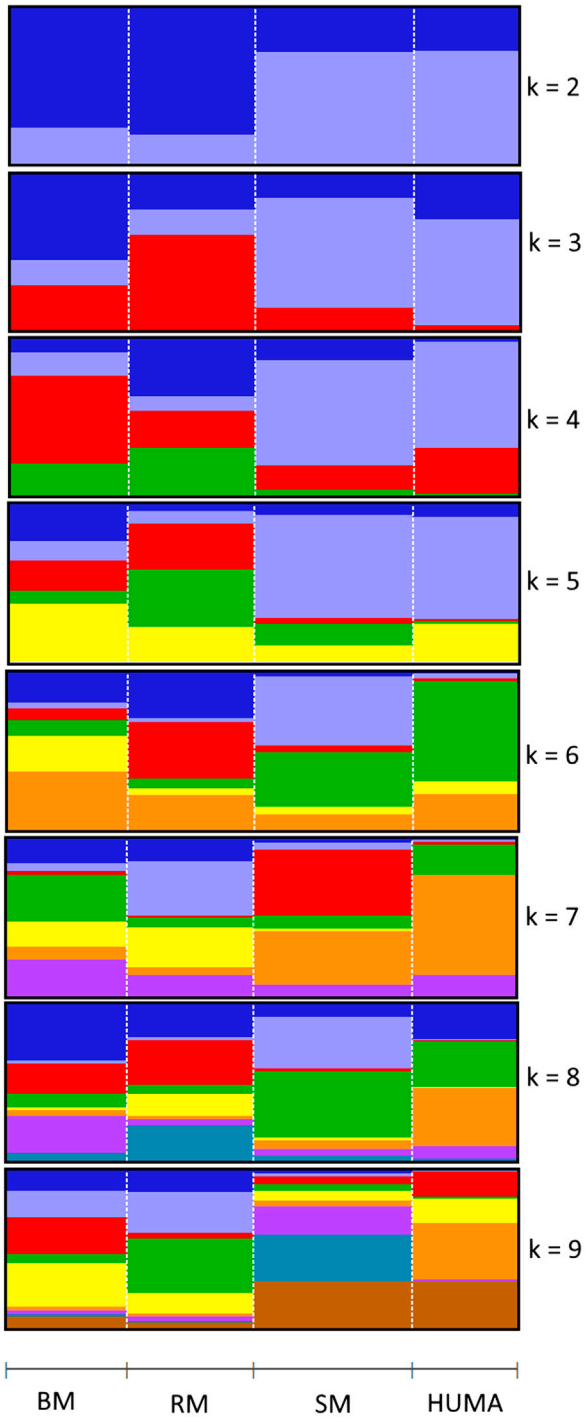


FIGURE 3
Admixture analysis of four Mangalitsa pigs populations (BM, RM, SM and HUMA) with graphs representing cluster levels 2 through 9. The optimal cluster level was found at k = 6.

cluster of BM. The ADMIXTURE analysis for low levels of k, especially, k = 2 showed a high degree of similarity of genetic background between BM and RM on one hand, and on the other hand, similarity between SM and HUMA (Figure 3). However, there was no complete separation of all breeds as demonstrated by traces of admixture at all cluster levels including k = 6, which produced the lowest cross-validation error estimate (Supplementary Figure S1).

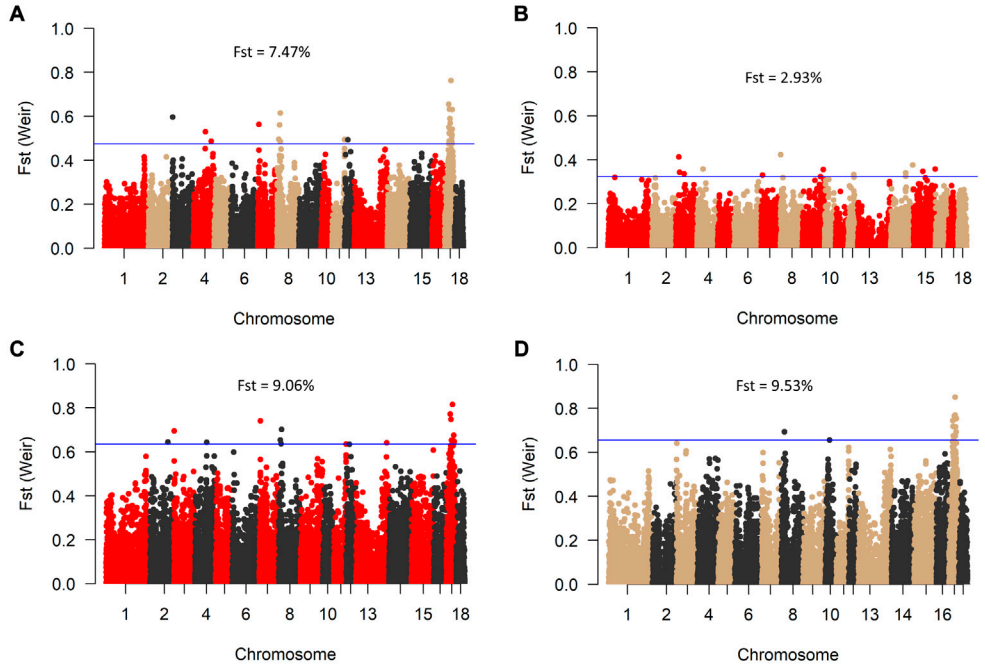


FIGURE 4
Manhattan plot of genome-wide Fst values between BM, RM and SM **(A)**; RM and BM **(B)**; RM and SM **(C)**, and BM and SM **(D)**. The blue line indicates genome-wide significant threshold above which SNPs were considered significant for candidate gene identification. Mean Fst values for each comparison is specified in percentage.

Additionally, genetic background was highly heterogeneous in all breeds.

Consistent with the principal component analysis, the lowest mean F_{st} (0.029) was found between RM and BM, while these two breeds were relatively distant from SM (Figure 4). By comparing all three breeds in a single F_{st} analysis, 30 genome-wide significant variants were detected across SSC 3, 4, 7, 8, 11, 12, and 17. Most of the significant variants (67%) were located on SSC17, which also had the topmost SNP (DRGA0016741) at position 38908507. The restriction of the analysis to a pairwise comparison revealed 14 significant variants between RM and BM with none occurring on chromosome 17. A number of significant variants were common to the comparison between RM and SM and between BM and SM. Five of these variants are depicted in Table 1 with superscript letters a—e. Additionally, Table 1 provides candidate genes located within 100 kb on either side of the F_{st}-based significant variants.

Within-breed diversity

Average Ho estimates were 0.27, 0.28 and 0.29 for BM, SM and RM, respectively. The heterozygosity estimates correlated significantly and negatively with total genomic inbreeding (r = -0.88; *p*-value < 2.2e-16) as shown in Figure 5. Across breeds, Ho ranged from 0.12 at the highest level of F_{ROH} (67.25%) to 0.42 at the lowest inbreeding level (0.39%).

The computed values of F_{ROH} for various ROH length considerations are presented with respect to breed in Figure 6. Considering all ROH, average inbreeding was highest in BM (24.11%) and lowest in RM (16.34%). This is also true for other ROH length considerations. Larger ROH segments played a major role in autozygosity in all breeds, and ROH segments below 2 Mb in size were completely absent in all animals. In general, fewer number of ROH segments were found in RM than in BM and SM (Table 2). Although BM and SM had about the same average number of ROH segments, the former had the highest sum of ROH (543626.6 kb) per individual. Across all breeds, 2715 ROH segments were found in this study.

The incidence of ROH and ROH islands varied across breeds (Figure 7). BM and SM had higher baseline ROH levels with ROH islands detected on SSC16 in BM, and on SSC11 and SSC17 in SM. The most significant variant in ROH island was found in 24 out of 30 SM animals on SSC17. In RM, ROH incidence levels were generally low, however, ROH islands were found on SSC7, SSC13 and SSC15. ROH incidence plots per chromosome show seemingly similar patterns across breeds (Supplementary Figures S2–S4). Furthermore, candidate genes proximal to ROH islands are presented in Table 3. Significant

TABLE 1 Summary information of significant SNPs and candidate genes from F_{st} comparison of three Mangalitsa pig breeds (BM, SM and RM). Superscripts a-e denote common SNPs across pairwise comparisons.

Breeds	CHR	SNP	Position	F_{ST}	Candidate genes
BMSM	17	DRGA0016741[a]	38908507	0.85	CEP250, ERGIC3, SPAG4, CPNE1, NFS1
	17	MARC0026961	39771108	0.77	DLGAP4, TGIF2, MYL9
	17	MARC0017379	33921649	0.76	SIRPB2, NSFL1C, SDCBP2, SNPH
	17	ALGA0095426[b]	46764111	0.75	OSER1, GDAP1L1, FITM2, R3HDML, HNF4A
	17	MARC0032380[c]	25017561	0.74	KIF16B
	17	H3GA0048218	26643149	0.72	ZNF133, DZANK1, POLR3F, RBBP9
	17	ALGA0094584[d]	33341575	0.72	STK35
	17	H3GA0049036	41739085	0.72	SLC32A1, ACTR5, PPP1R16B
	8	ALGA0046856[e]	20576000	0.69	STIM2
	17	ASGA0077154	45575325	0.69	PTPRT
	17	ALGA0094114	27748357	0.68	SLC24A3, RIN2
	17	ALGA0095121	41180101	0.67	RPRD1B, TGM2, KIAA1755
	17	ASGA0076251	30236910	0.66	THBD, CD93
	10	H3GA0029711	26091506	0.66	ERCC6L2
RMSM	17	DRGA0016741[a]	38908507	0.82	CEP250, ERGIC3, SPAG4, CPNE1, NFS1
	17	MARC0032380[c]	25017561	0.77	KIF16B
	7	ALGA0038333	7188826	0.74	TFAP2A
	8	ALGA0046856[e]	20576000	0.70	STIM2
	3	H3GA0008443	2734718	0.70	SDK1
	17	ALGA0095426[b]	46764111	0.68	OSER1, GDAP1L1, FITM2, R3HDML, HNF4A
	17	ALGA0094584[d]	33341575	0.65	STK35
	17	ASGA0077190	46189837	0.65	SRSF6, SGK2, IFT52, MYBL2
	17	M1GA0022187	49729457	0.65	NCOA3, SULF2
	2	H3GA0007369	113554653	0.64	FBXL17
	4	ALGA0026041	75779884	0.64	PLAG1, MOS, RPS20, LYN
	13	M1GA0017756	193064698	0.64	GRIK1
	8	ALGA0046640	16400498	0.64	KCNIP4
RMBM	8	M1GA0011680	2332002	0.42	ADRA2C
	3	ALGA0017963	20715185	0.41	HS3ST4
	4	MARC0024274	39449105	0.36	TSPYL5, CPQ
	15	MARC0036536	137538236	0.36	RBM44, RAMP1, SCLY
	9	ALGA0055521	130209873	0.35	RPS6KC1
	15	ALGA0107554	60110316	0.35	FMNL2
	3	ALGA0018104	25852288	0.34	GPR139, GPRC5B, IQCK
	3	ALGA0019015	53054275	0.34	CREG2, RNF149, CNOT11, TBC1D8
	12	MARC0039239	35751427	0.33	DHX40, CLTC
	7	ASGA0031021	8122178	0.33	NEDD9, TMEM170B

variants with no genes found within 100 kb distance on either side were not included.

In the separate analysis, where SNP positions were based on *Sus scrofa* genome version 10.2, ROH islands were detected on SSC7 (113–117 Mb) and SSC15 (125–134 Mb) in RM, and on SSC11 (80–86 Mb) and SS17 (17–29 Mb) in SM (Supplementary Table S1). For BM, no SNP reached a *p*-value > 0.999 for ROH due to the breed's high level of inbreeding. Therefore, we set a maximal threshold of 55% (*p*-value $\geqslant$ 0.998) for ROH island detection. Hence, ROH islands were found on SSC7 (104–106 Mb), SSC8 (29–31 Mb) and SSC16 (21–24 Mb; 60–62 Mb) in BM.

Discussion

The Mangalitsa pig with its dense curly hair and forward falling ears is considered precious, offering advantages such as

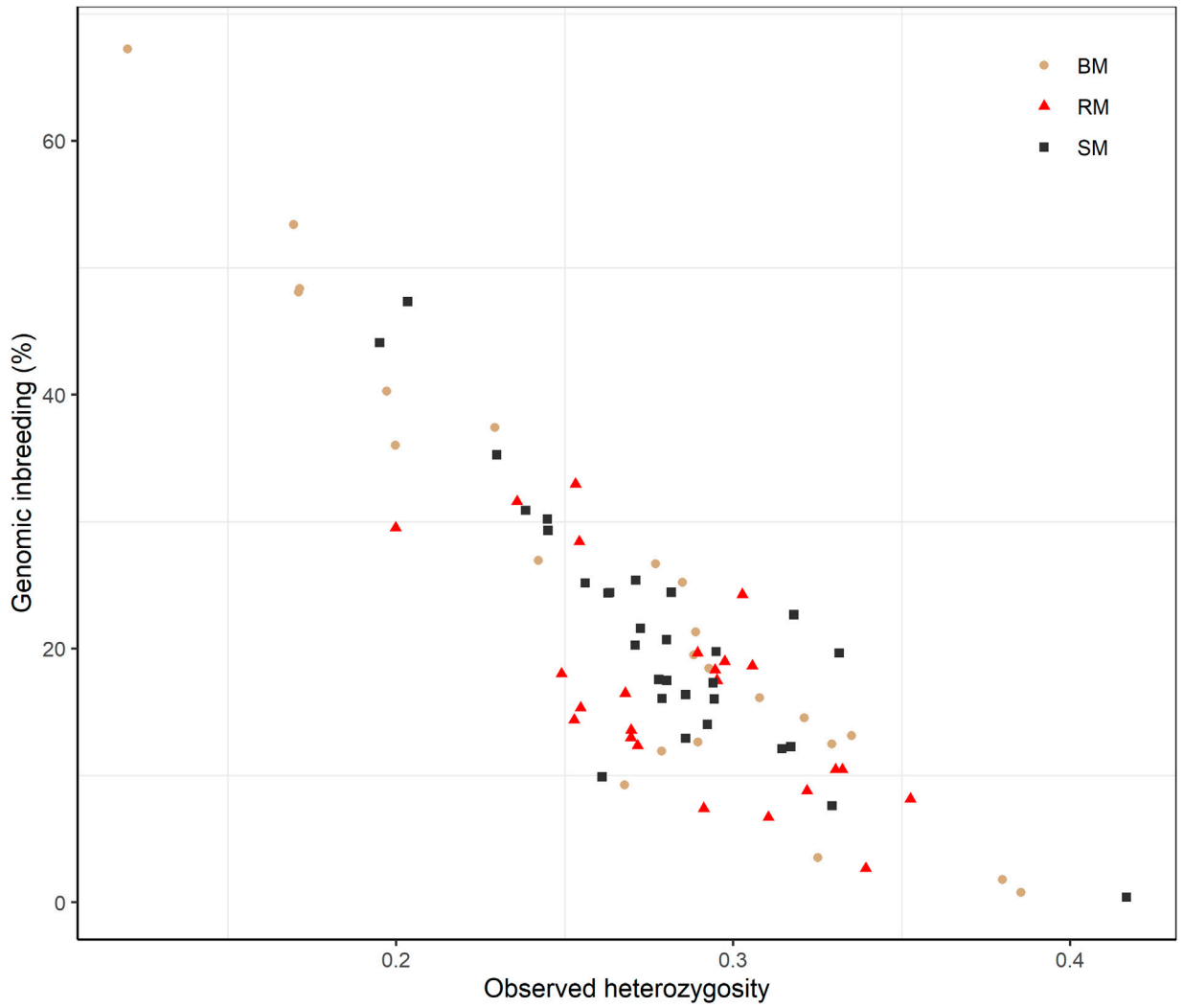


FIGURE 5
Scatterplot of the correlation between observed heterozygosity and genomic inbreeding (FROH) for Blond, Red and Swallow-bellied Mangalitsa pigs. The Pearson correlation coefficient was significant across breeds (r = 0.88; p-value < 2.2e-16).

resistance to challenging weather conditions and diseases (Egerszegi et al., 2003). Also, it is not very demanding on housing conditions and feeding. Therefore, breeders in Balkan countries, Austria, Switzerland, and Germany aim to conserve the original Mangalitsa without deleterious effects of inbreeding (Egerszegi et al., 2003). There is a growing interest in the exploration of the genome resource of Mangalitsa pigs, relying on an array of genetic markers including microsatellites (Zsolnai et al., 2006; Druml et al., 2012; Kharzinova and Zinovieva 2020), mitochondrial DNA D-loop sequences (Marincs et al., 2013; Molnár et al., 2013; Frank et al., 2017), SNPs (Bâlteanu et al., 2019; Bâlteanu et al., 2021; Zorc et al., 2022) and whole genome sequences (Molnár et al., 2014). Previous studies have attempted to discriminate between Mangalitsa pigs of different colors, highlighting a rejection of the hypothesis that Mangalitsa individuals belong to one indistinguishable group (Zsolnai et al., 2006). The current study did not only investigate genetic differentiation between the breeds, but also probed the occurrence of candidate genes in the vicinity of significantly differentiated marker loci. Additionally, we listed genes that are proximal to genomic regions displaying high levels of autozygosity. Whether or not these genes are differentially expressed in the studied breeds remains a subject for discussion.

Population structure and diversity

The high degree of clustering of individual Mangalitsa pigs by breed in this study (Figure 2) is a notable observation echoing the findings of previous studies. Based on 10 microsatellite markers,

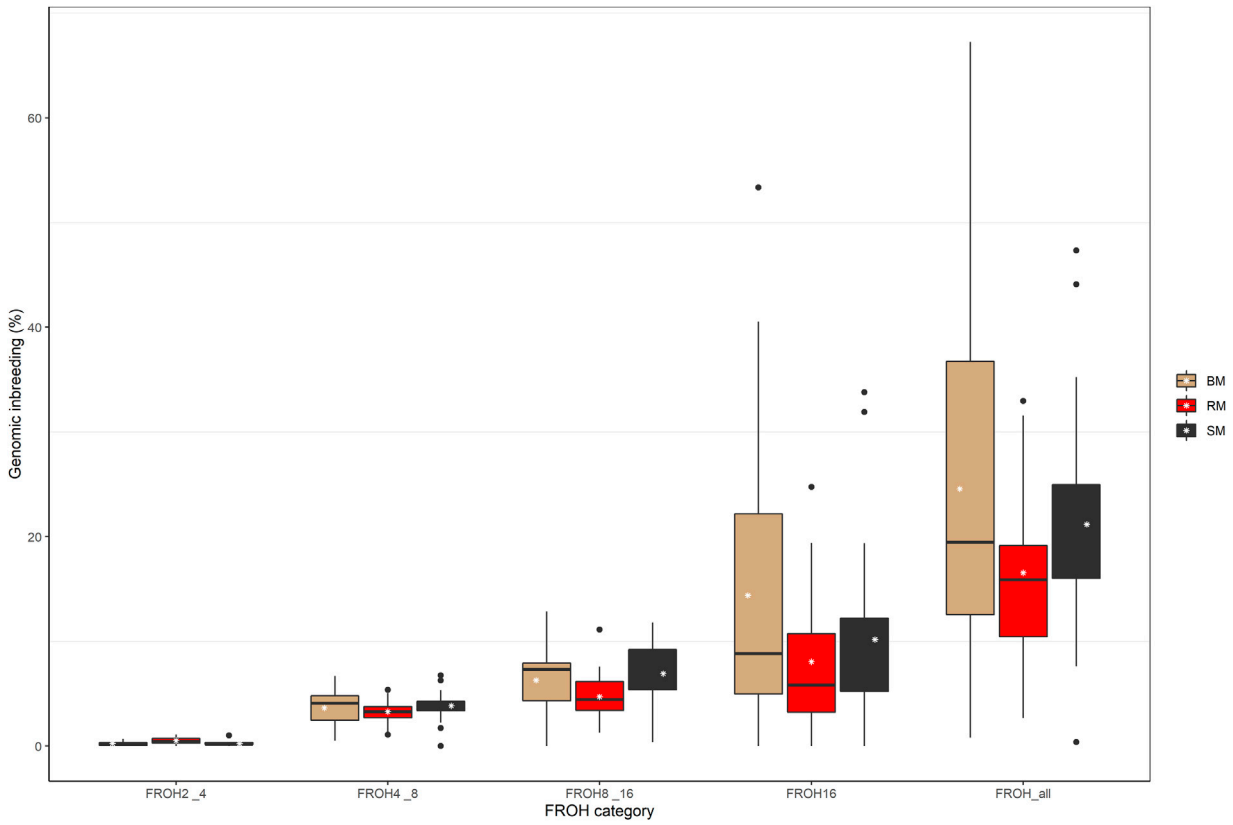


FIGURE 6
Boxplot showing the distribution of genomic inbreeding for various ROH length categories in Blond, Red and Swallow-bellied Mangalitsa pigs. The white stars represent estimate of average inbreeding.

TABLE 2 The average and total number of ROHs, and average sum of ROH length in BM, RM, SM and across all three breeds (ALL). Minimum and maximum values are provided in brackets.

Breed	Mean Nr. ROH (min + max)	Total Nr. ROH	Mean ∑ROH in kb (min + max)
BM	37.00 (3–64)	851	543626.6 (17568–1487860)
RM	30.83 (9–50)	740	368431.3 (59347–733062)
SM	37.47 (1–56)	1,124	469469.6 (8,582–1049150)
ALL	35.26 (1–64)	2,715	460128.0 (8,582–1487860)

Zsolnai et al. (2006) separated the three breeds *via* genetic distance and Structure analyses. Bâlteanu et al. (2019) also reported some degree of differentiation between the three breeds and other pig types through Admixture and principal component analysis using the Porcine SNP60 Beadchip. The persistency of traces of common genetic background across all cluster levels in our study reflects the developmental history of the breeds—RM and SM derived from BM. This is also consistent with previous studies that partly suggest that a present-day Mangalitsa population evolved by introgression of other European breeds and wild boars (Marincs et al., 2013). Explaining their observation of a close relationship between Mangalitsa and Turopolje pigs, Frank et al. (2017) mentioned a common wild boar, the Siska pig, assumed to be amongst the ancestors of both breeds. Another study suggests a possible geneflow from Slavonian Black and Pietrain pigs into both Mangalitsa and Turopolje pig populations, and reported lower levels of differentiation (from F_{st} = 0.05 to F_{st} = 0.09) between Mangalitsa and Slavonian Black (Druml et al., 2012). The closeness between Mangalitsa (SB) and Slavonian Black was

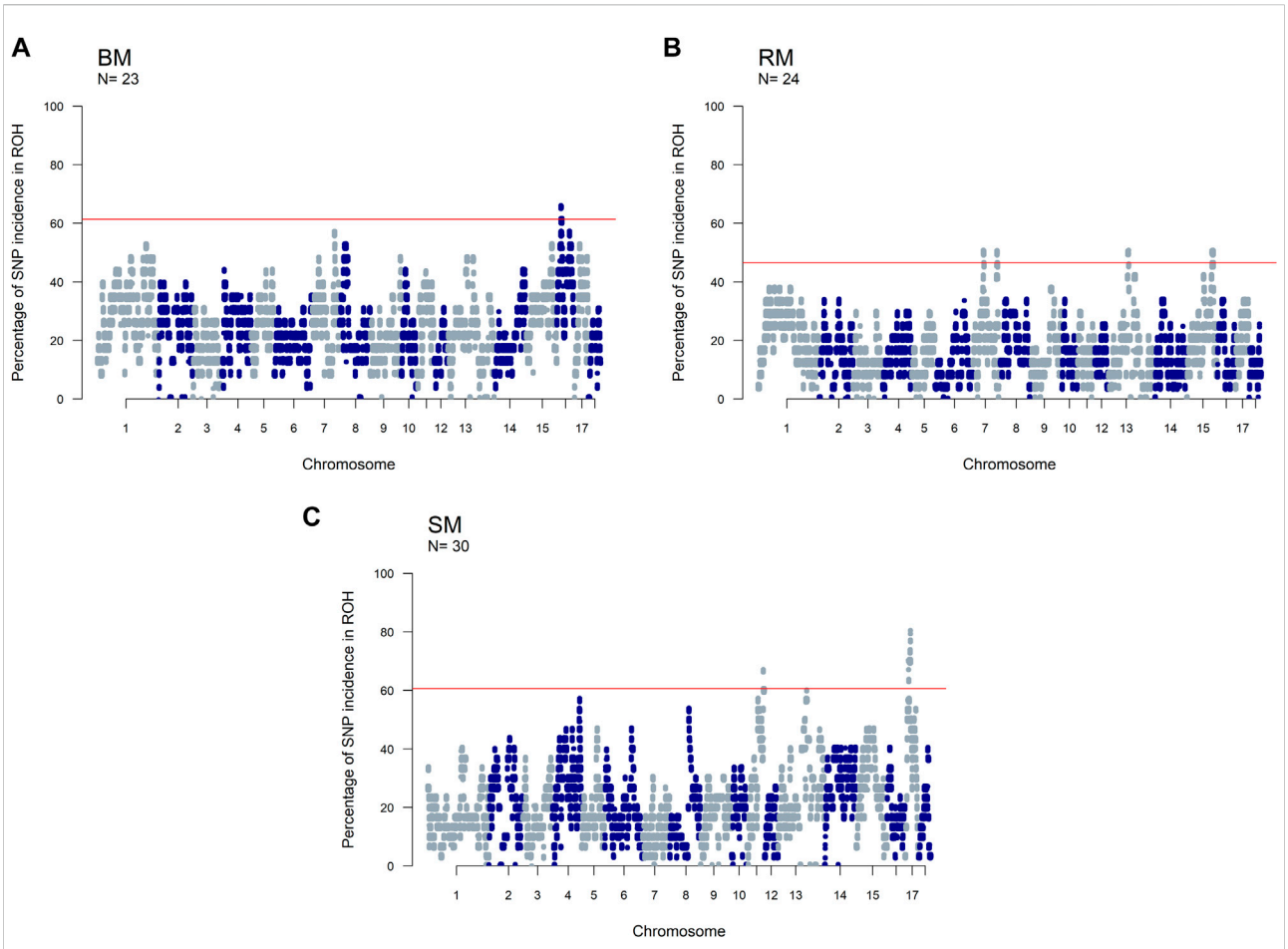


FIGURE 7
Manhattan plots of the incidence of SNPs in ROH for three Mangalitsa pig breeds: Blond **(A)**, Red **(B)** and Swallow-bellied **(C)**. The red line corresponds to a population specific p-value above which SNPs are considered to be in ROH islands.

confirmed by Zorc et al. (2022) who also showed potential gene flow between SB and Moravka pigs using both microsatellite and SNP data. What remains inconsistent across studies is the degree of differentiation among the breeds. We found in the current study, close connectedness between BM and RM than each of them to SM, which was corroborated by our F_{st} analysis, i.e., 0.029 (BMRM) vs. 0.091 (RMSM) or 0.095 (BMSM). On the contrary, Zsolnai et al. (2006) found BM and SM to be closer to each other (F_{st} = 0.064) than each of them to RM (F_{st} equals 0.094 and 0.099, respectively). Similarly, Marincs et al. (2013) found lower estimates of both F_{st} and Nei's distance for BMSM than for either BMRM or RMSM. Their findings are consistent with the observation of a high degree of similarity of genetic background between BM and SM at low cluster levels in Bâlteanu et al. (2019). Differences between our findings and those of previous studies could stem from differences in breeding focus and management across populations since the genotypes in the previous study were of Hungarian and Romanian origin. It is therefore, not surprising that the 20 HUMA genotypes, which we later included in our analysis, formed clusters predominantly with SM, and with BM to a lesser extent. Thus, we suspect the HUMA outgroup genotypes to belong to the SM breed, however, this cannot be immediately confirmed. Albeit, we consider the high resemblance between BM and RM (our finding) symbolic of the crossing of BM with Szalonta to develop RM (Zsolnai et al., 2013; Jukes 2017). Worth mentioning is that Szalonta is already one of the progenitors of BM (Egerszegi et al., 2003; Jukes 2017).

The dispersion of RM genotypes in our analysis reveals either a comparatively high level of genetic diversity or population substructures in this breed. Consistently, RM had the highest heterozygosity estimate (0.29) in our study. Even so, for the same breed types, observed heterozygosity estimates in previous studies are slightly higher than in ours, i.e., 0.32–0.38 vs. 0.29 for RM, 0.31 vs. 0.27 for BM and 0.29 vs. 0.28 for SM

TABLE 3 Summary information of significant SNPs and candidate genes from ROH islands in three Mangalitsa pig breeds (BM, SM and RM).

Breed	CHR	SNP	Position	ROH (%)	Candidate genes
SM	17	MARC0086847	25118032	80.08	KIF16B
	17	H3GA0048106	23564839	73.89	MACROD2
	17	MARC0030901	24049827	73.54	MACROD2
	17	ALGA0093805	22510190	70.63	ESF1, NDUFAF5, SEL1L2
	17	MARC0109531	20937630	70.63	BTBD3
	17	H3GA0048183	26170282	70.40	RRBP1, BANF2
	17	H3GA0047989	17680444	63.92	PLCB1
	11	H3GA0032842	77609850	60.66	ARHGEF7
	11	DRGA0011540	74100734	60.65	EFNB2
BM	16	MARC0086770	20525463	65.77	RAI14, BRIX1, RAD1, AGXT2, PRLR
	16	H3GA0046210	21372379	61.39	IL7R, CAPSL, LMBRD2, SKP2
	16	ALGA0115908	22998327	60.99	WDR70, GDNF
RM	15	DRGA0015508	120520780	50.81	CATIP, SLC11A1, CTDSP1, VIL1, USP37, CNOT9
	15	CASI0006663	117450632	50.81	ABCA12
	15	MARC0003725	119092405	50.81	TNP1
	15	ALGA0086843	114044775	50.80	ERBB4
	7	DIAS0001981	46539039	50.80	TMEM14A, GSTA1, GSTA4
	13	ASGA0058507	87695642	50.73	ZIC4, ZIC1
	13	H3GA0037098	89899499	50.72	WWTR1, COMMD2, ANKUB1
	7	ALGA0041512	47324213	50.71	CHRNA3, CHRNA5
	13	ALGA0071380	86992936	50.70	PLSCR5
	7	ASGA0033712	48095614	50.57	MORF4L1, CTSH, RASGRF1
	15	ASGA0070454	116910010	50.38	VWC2L
	15	DRGA0015443	113969520	49.67	ERBB4
	13	CASI0008610	90639691	46.65	TSC22D2
	7	DIAS0001022	110412608	46.65	SPATA7, PTPN21, ZC3H14, EML5,
	15	MARC0014079	115472905	46.50	IKZF2
	7	ALGA0044724	111026023	46.45	FOXN3
	7	ALGA0041861	49748787	46.44	CEMIP, MESD, TLNRD1

(Bâlteanu et al., 2019). Higher heterozygosity values in Bâlteanu et al. (2019) were associated with relatively low levels of inbreeding. Thus, across the three breeds, average F_{ROH} ranged from 0.09 to 0.14 in Bâlteanu et al. (2019) compared to our range of values being 0.16 to 0.24. A lower SNP-based observed heterozygosity value of 0.26 with a corresponding F_{ROH} estimate of 0.29 was previously estimated for SM by Zorc et al. (2022). They found SM to be second (after Turopolje: F_{ROH} = 0.51), to have low level of genetic diversity amongst other local Balkan pig breeds. Contrary to previous finding (Bâlteanu et al., 2019) of RM exhibiting the strongest and most recent inbreeding, BM in our study had the highest level of autozygosity regardless of the age of inbreeding (Figure 6). Nevertheless, consanguinity increased in all three breeds in more recent generations, probably, about three generations ago as demonstrated by predominance of ROH larger than 16 Mb (Meszaros, 2015). More generally, Mangalitsa pigs are known to have undergone a serious demographic decline in the past (Egerszegi et al., 2003; Posta et al., 2015; Bâlteanu et al., 2019). For the outgroup HUMA breed, observed heterozygosity estimated by Gorssen et al. (2021) was 0.26 with a corresponding high level of F_{ROH} estimate of 0.41. Since Gorssen et al. (2021) mentioned Yang et al. (2017) as source of the HUMA dataset used in their ROH analysis, we compared the estimate of F_{ROH} obtained for HUMA in the two studies. Average F_{ROH} for the same individuals was almost halved (0.22) in Yang et al. (2017). Variability in ROH statistics across studies are a direct consequence of differences in ROH calling criteria (Addo 2020; Meyermans et al., 2020) In this study, we used an R-script identical to Gorssen et al. (2021) for ROH calling, making it possible to compare our results to those available at the Open Science Framework (https://doi.org/10.17605/OSF.IO/XJTKV). For suitability of ROH island comparison, we further performed a separate analysis of our data using a previous genome assembly (*Sus scrofa* 10.2 genome build) as was used in the previous study. Autozygosity across breeds in our population of Mangalitsa pigs (n = 77) was by far lower (~0.21) than was found

for the Hungarian population (n = 20). The HUMA genotypes were sampled between 2008 and 2010 while we sampled our animals about a decade later, between 2018 and 2020. It could be that breed conservation strategies are being effectively implemented to save these endangered animals in recent years. Moreover, Bâlteanu et al. (2019) argued that ROH statistics recorded in Mangalitsa pigs are comparable to those measured in most cosmopolitan breeds, adding the difficulty in predicting significant statistical differences due to high dispersion of F_{ROH} data in Mangalitsa pigs. Nevertheless, comparing F_{ROH} value of 0.14 (Bâlteanu et al., 2019) to that of 0.29 in SB, Zorc et al. (2022) emphasized recent increase in inbreeding due to severe reductions in census numbers.

The inverse relationship between inbreeding coefficients and heterozygosity was generally studied across breeds in the current study. Our finding of a strongly negative correlation (−0.88) between the two variables is not limited to this study. Slate et al. (2004) investigated the use of multi-locus heterozygosity as a robust surrogate for inbreeding coefficients for subsequent investigation of inbreeding depression, which is especially important for captive populations. Based on a limited number of microsatellite markers, not only Slate et al. (2004) but also other studies, e.g., Hedrick et al. (2001), Overall et al. (2005), Jensen et al. (2007) and Alho et al. (2009) generally reported negative weak correlation coefficient values between heterozygosity and pedigree-based inbreeding coefficients, although the estimate for one of the populations (Scandinavian wolves) reached −0.72. The authors identified the effect of marker density and pedigree errors among other factors that impact the estimate of correlation coefficients between the two variables. Higher estimate of correlation in our study could be legitimate owing to high maker density and a better precision of genomic inbreeding estimates. Our findings reemphasize the utility of heterozygosity as proxy for inbreeding in the study of inbreeding depression in light of the availability of current genetic markers and where phenotypes are available.

Genes under selection

The detection of ROH island in all our breeds allows for the investigation of selection signatures, which are a consequence of selection, recombination and genetic drift (Ceballos et al., 2018; Gorssen et al., 2020). Surprisingly, identified ROH islands were not shared across the three breeds in our study, rather, SM and the outgroup HUMA shared the same ROH island on SSC11 (80–86 Mb). By contrast, in an investigation of breed substructures in Pietrain pigs, Gorssen et al. (2020) found several ROH islands common to different populations of the same pig type, and even reported a large ROH island on SSC8 (34–126 Mb), which appears to be fixed. For our findings, it is compelling that BM, RM, and SM have been managed as separate breeds whose ROH patterns have been predominantly shaped through within-breed selection. Assuming that the HUMA genotypes are from Swallow-bellied Mangalitsa as suggested by our principal component analysis, the shared ROH island region of this breed with SM indicates a breed-specific signature of selection present in both the Hungarian and German populations. The shared ROH island between SM and HUMA was mapped to SSC11 (73–78 Mb) in our original analysis, where an updated reference assembly (*Sus scrofa* genome version 11.1) was used. One out of nine previously detected ROH islands in SM was mapped to SSC11, however, the exact position (33,516,188–36,493,548) differs from our finding (Zorc et al., 2022). The authors identified PCDH20 which is associated with brain activity and tameness in this region. They also found many other genes in regions some of which overlapped with those in Black Slavonian, Banija spotted and Moravka pig breeds. To investigate candidate genes in our study, we limited our scope to 100 kb distance flanking the significant SNPs defining ROH islands (Table 3). The proximal, genes ARHGEF7 (SSC11: 77,427,368–77,530,995) and EFNB2 (SSC11: 74,186,64–4,233,354) appear to be important in the SM and HUMA. ARHGEF7 is associated with the storage of materials, including nutrients, pigments and waste products (GO:0000322), and might influence the coat color pattern in SM. Among others, EFNB2 plays a role in the negative regulation of neuron projection development (GO:0010977), a positive regulation of neuron death (GO:1901216) and the movement of cells in response to specific chemical signals (GO:0050920). The gene has been suggested as a candidate for hearing and visual impairment, and pigmentary anomalies in human (Lévy et al., 2018). Distinguishing between Landrace and Yorkshire pigs *via* fixation index analysis, Wang et al. (2021) found EFNB2 to have been under intense selection pressure.

The second ROH island in SM (SSC17: 16–26), which was neither shared by HUMA nor BM and RM revealed a high level of selection for the kinesin family member 16B (KIF16B) gene. Interestingly, the same gene also popped up while comparing SM to both BM and RM (Table 1). Located on SSC17 (24,871,985–25,188,204), the 31,6219 BP length KIF16B gene is predominantly involved in microtubule-based movement (GO: 0007018). By this, it plays an important role in regulating early development and organogenesis such as in embryos, kidneys and in stem cells (Hirokawa and Tanaka 2015). KIF16B gene was found to significantly influence wool length and greasy yield in fine wool sheep (Wang et al., 2014; Zhao et al., 2021). In humans it has been associated with synaptic signaling, which confers intellectual abilities (Loo et al., 2012). Therefore, SM may differ from the other breeds in terms of cognitive ability. To our knowledge, this has not yet been investigated in our breeds. Nevertheless, several studies have shown general differences in cognitive abilities in pigs (Gieling et al., 2011). The ROH region in SM also harbors other important genes—BANF2 and BTBD3 are associated with male (Omolaoye et al., 2022) and female (Kim et al., 2012) fertility in pigs; ESF1 is associated with

meat quality in pigs (Puig-Oliveras et al., 2014); and MACROD2 is reported to be associated with disease resistance in cattle (González-Ruiz et al., 2019).

Autozygosity in RM is especially related to genes associated with meat quality including ABCA12 (Wang Xiao et al., 2019), VIL1 (Zhang et al., 2015; Verardo et al., 2017), PLSCR5 (Wang Zezhao et al., 2019) and USP37 (Verardo et al., 2017). Meat quality assessment in Mangalitsa pigs has been very general and shown that meat from these pigs have higher intramuscular fat content compared to a commercial pig breed (Stanišić et al., 2016). Nistor et al. (2012) found significant differences in the ratio of saturated and unsaturated fatty acids in RM and BM meat (35.88%:62.76% and 38.42%:59.94%, respectively). Although not properly documented, we know that among German Mangalitsa pig breeders, RM is mostly preferred owing to a perceived higher meat quality in the breed. Also, of importance in RM are genes previously reported to be associated with characteristics such as feed efficiency—CTSH (Russo et al., 2008) and VWC2L (Kejun et al., 2015); growth—RASGRF1 (Magee et al., 2010); and fertility—SPATA7 (Marques et al., 2018), TNP1 (Gòdia et al., 2020) and WWTR1 (Budna et al., 2017). In Egerszegi et al. (2003), RM had the highest body weight of 220 kg and 180 kg compared to lowest values of 165 kg and 170 kg in SM boars and sows, respectively.

Identified in the ROH island of BM, the PRLR gene is reported to play a role in prolactin signaling (GO:0038161), lactation (GO:0007595), mammary gland development (GO:0060644) and response to bacterium stimuli (GO:0009617). Studies on lactation in Mangalitsa pigs are lacking, however, there is one that only reported a mean lactation length of 52.57 days in BM (Petrović et al., 2013).

The occurrence of identical loci under divergent selection for the pairwise comparison of BMSM and RMSM signals similarity between BM and RM on one hand, and on the other hand, differences between each of the pair and SM. Inferring from the associated genes, characteristics such as behavior influenced by SPAG4 (Bélteky et al., 2016; Bélteky et al., 2017); body size by STIM2 (Osei-Amponsah et al., 2017); carcass and fat deposition by HNF4A (Fan et al., 2010; Alexandre et al., 2014; MoreiraMonterio et al., 2018; Criado-Mesas et al., 2020) and CEP250 (Damon et al., 2012); muscularity by CPNE1 (Dawei et al., 2010); and disease resistance by NFS1 (Dauben et al., 2021) are more likely to be similar between BM and RM than between each of these and SM. However, this may not be entirely true if several other differentially expressed gene control these traits. The RMBM comparison showed the lowest differentiation in this study, for which the SNP with the highest signal (M1GA0011680) was found proximal to an autoregulatory α-adrenergic receptor 2C (ADRA2C) gene involved in the regulation of smooth muscle contraction among others (GO:0006940). Contrasting RM and BM, Bâlteanu et al. (2021) previously found the region SSC16 (18–20 Mb) to have a potential effect on hair pigmentation. Their findings were, however, based on selection scans with HapFLK, BAYESACN and GWAS, and furthermore, the analysis of gene content which revealed the solute carrier family 45-member 2 (SLC45A2) locus as a candidate gene.

Since their development, Mangalitsa pigs have evolved differentially as evidenced by both intra- and inter-population statistics in the current and previous studies. Adjudged by F_{st} estimates, the degree of differentiation, which peaked at about 9.5% (BMSM) in the current study is low compared to those in other breeds such as the Turopolje, being 21% on average (Druml et al., 2012). Besides, Mangalitsa pigs are collectively described as endangered and require conservation interventions. The maintenance of small populations under restricted geneflow between animals of different coat colors raises concerns about long-term implications for the conservation of genetic diversity. In our original data, 17 animals were said to be crossbreds from the three main breeds, meaning that some German breeders already practice crossbreeding amongst different Mangalitsa pig breeds. The lack of pedigree records makes it difficult to confirm this type of cross, and to use such animals in our analyses. The crossing of different Mangalitsa breeds by German breeders is not yet documented, however, possible reason for crossbreeding in these pigs can be the favorable effect of breed complementarity and the avoidance of inbreeding.

Conclusion

In this study, we provided an insight into the differences in the three Mangalitsa pigs breeds using a medium density SNP information. The Blond and Red Mangalitsa breeds are more similar to each other than each of them to the Swallow-bellied Mangalitsa. This finding sharply contrasts previous reports of the lowest differentiation between Blond and Swallow-bellied Mangalitsa pigs. Genetic diversity was highest in Red Mangalitsa; however, inbreeding was considerably high in all the breeds. Highly homozygous genomic regions were not shared across breeds and this, to a large extent, emphasizes the restriction of geneflow among the breeds. We found several breed specific signatures of selection including those that may underline growth and meat quality traits in Red Mangalitsa or that suggest intellectual ability in Swallow-bellied Mangalitsa. By providing a list of candidate genes for all genome-wide significant variants, we propose further investigations that would link these genes to actual phenotypes where available. Furthermore, we would add Manhattan plots of our ROH incidence to the repository of ROH islands for comparisons with future studies.

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 in the article/Supplementary Material.

Ethics statement

Ethical review and approval was not required for the animal study because Ethical review and approval were waived for this study, since no experimental handling of animals was involved. The pigs were kept in line with national law and guidelines. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

SA and LJ conceived the study. SA wrote the codes. SA and LJ performed the data analysis. SA drafted the manuscript and both finalized the writing.

Acknowledgments

We are sincerely grateful to the German Mangalitsa Breeders Association for their valuable support.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.909986/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Cross-validation for varying number of clusters ranging from k = 2 to k = 20 in Mangalitsa pigs.

SUPPLEMENTARY FIGURE 2
ROH incidence plots for each chromosome in Blond Mangalitsa pigs.

SUPPLEMENTARY FIGURE 3
ROH incidence plots for each chromosome in Red Mangalitsa pigs.

SUPPLEMENTARY FIGURE 4
ROH incidence plots for each chromosome in Swallow-bellied Mangalitsa pigs.

SUPPLEMENTARY TABLE 1
SNPs in ROH island in Blond (BM), Red (RM) and Swallow-bellied (SM) Mangalitsa pigs. SNP positions were mapped to the Sus scrofa genome version 10.2.

SUPPLEMENTARY DATA 1
Plink formatted files consisting of Blond (BM), Red (RM) and Swallow-bellied Mangalitsa (SM) genotypes.

References

Addo, S. (2020). Assessment of genetic diversity in local breeds. Dissertation thesis. Kassel, Germany: University of Kassel. https://kobra.uni-kassel.de/handle/123456789/12613

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109

Alexandre, P. A., Gomes, R. C., Santana, M. H. A., Silva, S. L., Leme, P. R., Mudadu, M. A., et al. (2014). Bovine NR1I3 gene polymorphisms and its association with feed efficiency traits in Nellore cattle. *Meta Gene* 2014, 206–217. doi:10.1016/j.mgene.2014.01.003

Alho, J. S., Lillandt, Bo-G., Jaari, S., and Merilä, J. (2009). Multilocus heterozygosity and inbreeding in the Siberian jay. *Conserv. Genet.* 10 (3), 605–609. doi:10.1007/s10592-008-9588-z

Bâlteanu, V. A., Cardoso, T. F., Amills, M., Egerszegi, I., Anton, I., Beja-Pereira, A., et al. (2019). The footprint of recent and strong demographic decline in the genomes of Mangalitza pigs. *Animal* 13 (11), 2440–2446. doi:10.1017/s1751731119000582

Bâlteanu, V. A., Cardoso, T. F., Amills, M., Luigi-Sierra, M. G., Egerszegi, I., Anton, I., et al. (2021). Red and blond Mangalitza pigs display a signature of divergent directional selection in the SLC45A2 gene. *Anim. Genet.* 52 (1), 66–77. doi:10.1111/age.13031

Bélteky, J., Agnvall, B., and Jensen, P. (2017). Gene expression of behaviorally relevant genes in the cerebral hemisphere changes after selection for tameness in Red Junglefowl. *PLOS ONE* 12 (5), e0177004. doi:10.1371/journal.pone.0177004

Bélteky, J., Agnvall, B., Johnsson, M., Wright, D., and Jensen, P. (2016). Domestication and tameness: Brain gene expression in red junglefowl selected for less fear of humans suggests effects on reproduction and immunology. *R. Soc. Open Sci.* 3 (8), 160033. doi:10.1098/rsos.160033

Budna, J., Bryja, A., Celichowski, P., Kahan, R., Kranc, W., Ciesiółka, S., et al. (2017). Genes of cellular components of morphogenesis in porcine oocytes before and after IVM. *Reprod. Camb. Engl.* 154 (4), 535–545. doi:10.1530/REP-17-0367

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., and Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nat. Rev. Genet.* 19 (4), 220–234. doi:10.1038/nrg.2017.109

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

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Fernández, A. I., and Folch, J. M. (2020). Identification of eQTLs associated with lipid metabolism in Longissimus dorsi muscle of pigs with different genetic backgrounds. *Sci. Rep.* 10 (1), 9845. doi:10.1038/s41598-020-67015-4

Damon, M., Wyszynska-Koko, J., Vincent, A., Hérault, F., and Lebret, B. (2012). Comparison of muscle transcriptome between pigs with divergent meat quality

phenotypes identifies genes related to muscle metabolism and structure. *PLOS ONE* 7 (3), e33763. doi:10.1371/journal.pone.0033763

Dauben, C. M., Pröll, C., Maren, J., Heuß, E. M., Appel, A. K., Henne, H., et al. (2021). Genome-wide associations for immune traits in two maternal pig lines. *BMC Genomics* 22 (1), 717. doi:10.1186/s12864-021-07997-1

Druml, T., Salajpal, K., Dikic, M., Urosevic, M., Grilz-Seger, G., and Baumung, R. (2012). Genetic diversity, population structure and subdivision of local Balkan pig breeds in Austria, Croatia, Serbia and Bosnia-Herzegovina and its practical value in conservation programs. *Genet. Sel. Evol.* 44 (1), 5–9. doi:10.1186/1297-9686-44-5

Durinck, S., Soellmen, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. doi:10.1038/nprot.2009.97

Egerszegi, I., Rátky, J., Solti, L., and Brüssow, K.-P. (2003). Mangalica – An indigenous swine breed from Hungary (review). *Arch. Anim. Breed.* 46 (3), 245–256. doi:10.5194/aab-46-245-2003

Fan, B., Du, Z-Q., and Rothschild, M. F. (2010). The Hepatocyte nuclear factor-1 alpha (HNF1A) gene is associated with fatness and loin muscle area in the pig. *Animal* 4 (10), 1619–1627. doi:10.1017/S175173111000087X

Flegler, J. (1999). *Das Wollschwein. Gefährdete Nutztierrasse des Jahres 1999*. Linz, Austria: Informationsbroschüre GEH.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 17 (1), 27–32. doi:10.1111/1755-0998.12509

Frank, K., Molnár, J., Barta, E., and Marincs, F. (2017). The full mitochondrial genomes of Mangalica pig breeds and their possible origin. *Mitochondrial DNA. B Resour.* 2 (2), 730–734. doi:10.1080/23802359.2017.1390415

García, D., Martínez, A., Dunner, S., Vega-Pla José, L., Fernández, C., et al. (2006). Estimation of the genetic admixture composition of Iberian dry-cured ham samples using DNA multilocus genotypes. *Meat Sci.* 72 (3), 560–566. doi:10.1016/j.meatsci.2005.09.005

Gieling, E. T., Nordquist, R. E., and Josef, F. (2011). Assessing learning and memory in pigs. *Anim. Cogn.* 14 (2), 151–173. doi:10.1007/s10071-010-0364-3

Gòdia, M., Reverter, A., González-Prendes, R., Ramayo-Caldas, Y., Castelló, A., Rodríguez-Gil, J-E., et al. (2020). A systems biology framework integrating GWAS and RNA-seq to shed light on the molecular basis of sperm quality in swine. *Genet. Sel. Evol.* 52 (1), 72. doi:10.1186/s12711-020-00592-0

González-Ruiz, S., Strillacci, M. G., Durán-Aguilar, M., Cantó-Alarcón Germinal, J., Herrera-Rodríguez Sara, E., Bagnato, A., et al. (2019): Genome-wide association study in Mexican holstein cattle reveals novel quantitative trait loci regions and confirms mapped loci for resistance to bovine tuberculosis animals : An open access journal from MDPI 9 (9). doi:10.3390/ani9090636

Gorssen, W., Meyermans, R., Janssens, S., and Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genet. Sel. Evol.* 53 (1), 2. doi:10.1186/s12711-020-00599-7

Gorssen, W., Meyermans, R., Buys, N., and Janssens, S. (2020). SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Piétrain pigs. *Anim. Genet.* 51 (1), 32–42. doi:10.1111/age.12888

Hedrick, P., Fredrickson, R., and Ellegren, H. (2001). Evaluation of d2, a microsatellite measure of inbreeding and outbreeding, in wolves with a known pedigree. *Evolution* 55 (6), 1256–1260. doi:10.1111/j.0014-3820.2001.tb00646.x

Herrero-Medrano, J. M., Megens, H-J., Martien, A. G., Bosse, M., Pérez-Enciso, M., and Crooijmans, R. P. M. A. (2014). Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics* 15 (1), 601–612. doi:10.1186/1471-2164-15-601

Hirokawa, N., and Tanaka, Y. (2015). Kinesin superfamily proteins (KIFs): Various functions and their relevance for important phenomena in life and diseases. *Exp. Cell Res.* 334 (1), 16–25. doi:10.1016/j.yexcr.2015.02.016

Jensen, H., Bremset, E. M., Ringsby, T. H., and Saether, B-E. (2007). Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation. *Mol. Ecol.* 16 (19), 4066–4078. doi:10.1111/j.1365-294X.2007.03452.x

Jukes, E. (2017). Mason's world encyclopedia of livestock breeds and breeding. *RR* 31 (5), 28–30. doi:10.1108/RR-02-2017-0039

Kejun, W., Dewu, L., Jules, H-S., Jie, C., Chengkun, L., Zhenfang, W., et al. (2015). Genome wide association analysis reveals new production trait genes in a male Duroc population. *PLOS ONE* 10 (9), e0139207. doi:10.1371/journal.pone.0139207

Kharzinova, V. R., and Zinovieva, N. A. (2020). The pattern of genetic diversity of different breeds of pigs based on microsatellite analysis. *Vavilovskii Zhurnal Genet. Sel.* 24 (7), 747–754. doi:10.18699/VJ20.669

Kim, M., Seo, H., Choi, Y., Shim, J., Kim, H., Lee, C-K., et al. (2012). Microarray analysis of gene expression in the uterine endometrium during the implantation period in pigs. *Asian-Australas. J. Anim. Sci.* 25 (8), 1102–1116. doi:10.5713/ajas.2012.12076

Lévy, J., Haye, D., Marziliano, N., Casu, G., Guimiot, F., Dupont, C., et al. (2018). EFNB2 haploinsufficiency causes a syndromic neurodevelopmental disorder. *Clin. Genet.* 93 (6), 1141–1147. doi:10.1111/cge.13234

Loo, S. K., Shtir, C., Doyle, A. E., Mick, E., McGough, J. J., McCracken, J., et al. (2012). Genome-wide association study of intelligence: Additive effects of novel brain expressed genes. *J. Am. Acad. Child. Adolesc. Psychiatry* 51 (4), 432–440. e2. doi:10.1016/j.jaac.2012.01.006

Magee, D. A., Sikora, K. M., Berkowicz, E. W., Berry, D. P., Howard, D. J., Mullen, M. P., et al. (2010). DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. *BMC Genet.* 11, 93. doi:10.1186/1471-2156-11-93

Manunza, A., Amills, M., Noce, A., Cabrera, B., Zidi, A., Eghbalsaied, S., et al. (2016). Romanian wild boars and Mangalitza pigs have a European ancestry and harbour genetic signatures compatible with past population bottlenecks. *Sci. Rep.* 6 (1), 29913–29919. doi:10.1038/srep29913

Marincs, F., Molnár, J., Tóth, G., Stéger, V., and Barta, E. (2013). Introgression and isolation contributed to the development of Hungarian Mangalica pigs from a particular European ancient bloodline. *Genet. Sel. Evol.* 45 (1), 22–26. doi:10.1186/1297-9686-45-22

Marques, D. B. D., Bastiaansen, J. W. M., Broekhuijse, M. L. W. J., Lopes, M. S., Knol, E. F., Harlizius, B., et al. (2018). Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genet. Sel. Evol.* 50 (1), 40. doi:10.1186/s12711-018-0412-z

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 (3), 359–372. doi:10.1016/j.ajhg.2008.08.007

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC genomics* 21 (1), 94. doi:10.1186/s12864-020-6463-x

Molnár, J., Nagy, T., Stéger, V., Tóth, G., Marincs, F., and Barta, E. (2014). Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. *BMC genomics* 15, 761. doi:10.1186/1471-2164-15-761

Molnár, J., Tóth, G., Stéger, V., Zsolnai, A., Jánosi, A., Mohr, A., et al. (2013). Mitochondrial D-loop analysis reveals low diversity in Mangalica pigs and their relationship to historical specimens. *J. animal Breed. Genet. = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie* 130 (4), 312–320. doi:10.1111/j.1439-0388.2012.01014.x

Moreira, G. C. M., Boschiero, C., Cesar, A. S. M., Reecy, J. M., Godoy, T. F., Trevisoli, P. A., et al. (2018). A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. *BMC Genomics* 19 (1), 374. doi:10.1186/s12864-018-4779-6

Nistor, E., Bampidis, V., Pentea, M., Prundeanu, H., and Ciolac, V. (2012). Nutritional quality of pork produced by Mangalitsa breed. *Bio Anim. Husb.* 2012(45), 386–389. doi:10.2298/BAH1002021P

Omolaoye, T. S., Hachim, M. Y., Du, P., and Stefan, S. (2022). Using publicly available transcriptomic data to identify mechanistic and diagnostic biomarkers in azoospermia and overall male infertility. *Sci. Rep.* 12 (1), 2584. doi:10.1038/s41598-022-06476-1

Osei-Amponsah, R., Skinner, B. M., Adjei, D. O., Bauer, J., Larson, G., Affara, N. A., et al. (2017). Origin and phylogenetic status of the local Ashanti Dwarf pig (ADP) of Ghana based on genetic analysis. *BMC Genomics* 18 (1), 193. doi:10.1186/s12864-017-3536-6

Overall, A. D. J., Byrne, K. A., Pilkington, J. G., and Pemberton, J. M. (2005). Heterozygosity, inbreeding and neonatal traits in Soay sheep on St Kilda. *Mol. Ecol.* 14 (11), 3383–3393. doi:10.1111/j.1365-294X.2005.02682.x

Petrović, M., Savic, R., Parunović, N., Radojkovic, D., and Radovic, C. (2013). "Reproductive traits of pigs of Mangalitsa breed," in 8th International Symposium on the Mediterranean Pig, Slovenia, Ljubljana, October 10th–12th, 2013.

Posta, J., Szabó, P., and Komlósi, I. (2015). Pedigree analysis of mangalica pig breeds. *Ann. Animal Sci.* 16 (3), 701–709. doi:10.1515/aoas-2015-0075

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D., Hudson, N. J., et al. (2014). Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. *PLOS ONE* 9 (6). e99720. doi:10.1371/journal.pone.0099720

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13, 70. doi:10.1186/1471-2156-13-70

Purfield, D. C., McParland, S., Wall, E., and Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLOS ONE* 12 (5), e0176780. doi:10.1371/journal.pone.0176780

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team.

Russo, V., Fontanesi, L., Scotti, E., Beretti, F., Davoli, R., Nanni Costa, L., et al. (2008). Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. *J. Anim. Sci.* 86 (12), 3300–3314. doi:10.2527/jas.2008-0920

Schachler, K., Distl, O., and Metzger, J. (2020). Tracing selection signatures in the pig genome gives evidence for selective pressures on a unique curly hair phenotype in Mangalitza. *Sci. Rep.* 10 (1), 22142. doi:10.1038/s41598-020-79037-z

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., et al. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Heredity* 93 (3), 255–265. doi:10.1038/sj.hdy.6800485

Stanišić, N., Parunović, N., Stajić, S., Petrović, M., Radović, Č., Živković, D., et al. (2016). Differences in meat colour between free-range Swallow Belly Mangalitsa and commercially reared Swedish Landrace pigs during 6 days of vacuum storage. *Arch. Anim. Breed.* 59 (1), 159–166. doi:10.5194/aab-59-159-2016

Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* 3 (25), 731. doi:10.21105/joss.00731

Verardo, L. L., Sevón-Aimonen, M-L., Serenius, T., Hietakangas, V., and Uimari, P. (2017). Whole-genome association analysis of pork meat pH revealed three significant regions and several potential genes in Finnish Yorkshire pigs. *BMC Genet.* 18 (1), 13. doi:10.1186/s12863-017-0482-x

Wang, Kai, Wu, Pingxian, Chen, Dejuan, Zhou, Jie, Yang, Xidi, Jiang, Anan, et al. (2021). Detecting the selection signatures in Chinese Duroc, Landrace, Yorkshire, liangshan, and qingyu pigs. *Funct. Integr. Genomics* 21 (5-6), 655–664. doi:10.1007/s10142-021-00809-5

Wang, Xiao, Cao, Chunwei, Li, Yongshun, Hai, Tang, Jia, Qitao, Zhang, Ying, et al. (2019a). A harlequin ichthyosis pig model with a novel ABCA12 mutation can be rescued by acitretin treatment. *J. Mol. Cell Biol.* 11 (12), 1029–1041. doi:10.1093/jmcb/mjz021

Wang, Zezhao, Zhu, Bo, Niu, Hong, Zhang, Wengang, Xu, Ling, Xu, Lei, et al. (2019b). Genome wide association study identifies SNPs associated with fatty acid composition in Chinese Wagyu cattle. *J. Anim. Sci. Biotechnol.* 10, 27. doi:10.1186/s40104-019-0322-0

Wang, Zhipeng, Zhang, Hui, Yang, Hua, Wang, Shouzhi, Rong, Enguang, Pei, Wenyu, et al. (2014). Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PloS one* 9 (9), e107101. doi:10.1371/journal.pone.0107101

Weir, B. S., and Cockerham, C. Clark (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38 (6), 1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x

Wilkinson, S., Lu, Zen H., Megens, Hendrik-Jan, Archibald, Alan L., Haley, Chris, Jackson, Ian J., et al. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genet.* 9 (4), e1003453. doi:10.1371/journal.pgen.1003453

Yang, Bin, Cui, Leilei, Perez-Enciso, Miguel, Traspov, Aleksei, Crooijmans, Richard P. M. A., Zinovieva, Natalia, et al. (2017). Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet. Sel. Evol.* 49 (1), 71–15. doi:10.1186/s12711-017-0345-y

Zhang, Chunyan, Wang, Zhiquan, Bruce, Heather, Kemp, Robert Alan, Charagu, Patrick, Miar, Younes, et al. (2015). Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. *BMC Genet.* 16, 33. doi:10.1186/s12863-015-0192-1

Zhao, Haiyu, Hu, Ruixue, Li, Fadi, and Yue, Xiangpeng (2021). Two strongly linked blocks within the KIF16B gene significantly influence wool length and greasy yield in fine wool sheep (*Ovis aries*). *Electron. J. Biotechnol.* 53, 23–32. doi:10.1016/j.ejbt.2021.05.003

Zorc, M., Škorput, D., Gvozdanović, K., Margeta, P., Karolyi, D., Luković, Z., et al. (2022). Genetic diversity and population structure of six autochthonous pig breeds from Croatia, Serbia, and Slovenia. *Genet. Sel. Evol.* 54 (1), 30. doi:10.1186/s12711-022-00718-6

Zsolnai, A., Radnóczy, L., Fésüs, L., and Anton, I. (2006). Do mangalica pigs of different colours really belong to different breeds? *Arch. Anim. Breed.* 49 (5), 477–483. doi:10.5194/aab-49-477-2006

Zsolnai, A., Tóth, G., Molnár, J., Stéger, V., Marincs, F., Jánosi, A., et al. (2013). Looking for breed differentiating SNP loci and for a SNP set for parentage testing in Mangalica. *Arch. Anim. Breed.* 56 (1), 200–207. doi:10.7482/0003-9438-56-019

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fgene.2022.886376

OPEN ACCESS

EDITED BY
Jianlin Han,
International Livestock Research Institute (ILRI), Kenya

REVIEWED BY
Vinzenz Lange,
DKMS Life Science Lab GmbH, Germany
Christian T. K.-H. Stadtlander,
Independent researcher, Destin, Florida, United States
Olympe Chazara,
Genomics plc, United Kingdom

*CORRESPONDENCE
Jun Heon Lee,
junheon@cnu.ac.kr

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

RECEIVED 28 February 2022
ACCEPTED 07 October 2022
PUBLISHED 28 October 2022

CITATION
Ediriweera TK, Manjula P, Cho E, Kim M and Lee JH (2022), Application of next-generation sequencing for the high-resolution typing of MHC-*B* in Korean native chicken.
Front. Genet. 13:886376.
doi: 10.3389/fgene.2022.886376

Application of next-generation sequencing for the high-resolution typing of MHC-*B* in Korean native chicken

Thisarani Kalhari Ediriweera[1], Prabuddha Manjula[2], Eunjin Cho[1], Minjun Kim[3] and Jun Heon Lee[1,3]*

[1]Department of Bio-AI Convergence, College of Engineering, Chungnam National University, Daejeon, Korea, [2]Department of Animal Science, Uva Wellassa University, Badulla, Sri Lanka, [3]Division of Animal and Dairy Science, College of Agriculture and Life Sciences, Chungnam National University, Daejeon, Korea

The major histocompatibility complex-*B* (MHC-*B*) region of chicken is crucially important in their immunogenesis and highly diverse among different breeds, lines, and even populations. Because it determines the resistance/susceptibility to numerous infectious diseases, it is important to analyze this genomic region, particularly classical class I and II genes, to determine the variation and diversity that ultimately affect antigen presentation. This study investigated five lines of indigenous Korean native chicken (KNC) and the Ogye breed using next-generation sequencing (NGS) data with Geneious Prime-based assembly and variant calling with the Genome Analysis Toolkit (GATK) best practices pipeline. The consensus sequences of MHC-*B* (*BG1-BF2*) were obtained for each chicken line/breed and their variants were analyzed. All of the Korean native chicken lines possessed an excessive number of variants, including an ample amount of high-impact variants that provided useful information regarding modified major histocompatibility complex molecules. The study confirmed that next-generation sequencing techniques can effectively be used to detect MHC variabilities and the KNC lines are highly diverse for the MHC-*B* region, suggesting a substantial divergence from red junglefowl.

KEYWORDS

assembly, Korean native chicken, MHC-*B*, NGS, variants

Introduction

The major histocompatibility complex (MHC) of chicken comprises two genetically disassociated gene clusters, MHC-*B* and MHC-*Y*, and the MHC-*B* contains the minimal essential region. It is largely responsible for the immune response and histocompatibility of chicken. It possesses classical MHC class I and II genes that produce MHC class I and class II molecules, which are extremely important biological agents in the recognition of foreign pathogenic peptides. It also carries BG family genes and C-type lectin like loci (Kaufman et al., 1999; Shiina et al., 2007; Kaufman, 2021).

The *BF1* and *BF2* genes belong to classical class I of the minimal essential region, while the *BLB1* and *BLB2* genes are in classical class II. Class I and II genes are extremely important in adaptive immune response, because they are recognized by cytotoxic and helper T lymphocytes, respectively; some are also important in innate immunity (i.e., the *BF1* gene, the minor class I gene that corresponds with natural killer cells as ligands). The *BG1* gene is also involved in adaptive immune response (Goto et al., 2009; Chen et al., 2018; Kaufman, 2021).

Numerous serological and molecular biological techniques have been used to study the MHC diversity of chicken. Although low-resolution MHC typing methods are capable of providing insights regarding diversity and haplotypes, they are still not able to describe the region as a whole, particularly class I and II variation (Fulton, 2020). Hence, the MHC haplotypes identified from different genetic sources based on single-nucleotide polymorphism (SNP)/microsatellite markers or direct sequencing need to be validated by high-resolution sequence typing to better understand the region.

Currently, next-generation sequencing (NGS) approach can generate the full-length sequences of highly polymorphic genomic regions such as MHC, with extremely high coverage and precision at each variant. Target polymerase chain reaction (PCR) and valid sequence information of the BF/BL region are crucial for haplotype identification and nomenclature.

A PCR-NGS approach for MHC haplotype identification has been developed and successfully applied in human leukocyte antigen (HLA) typing (Ozaki et al., 2015; Ka et al., 2017; Alizadeh et al., 2020). However, there are limited chicken MHC sequence data available with a sufficient sample count to describe the enormous haplotype diversity due to the forces of recombination and gene conversion within the MHC region. The databases containing such data were generated solely from MHC-*B* homozygous, inbred chicken lines. Local chicken breeds have not undergone such intensive selection for MHC polymorphism, which makes it difficult to obtain homozygous individuals. Recently, many DNA-based typing methods have been developed and successfully applied to local chickens of various origins (Fulton et al., 2016; Mwambene et al., 2019; Manjula et al., 2020). This difficulty can therefore be overcome by collecting primary data from low-resolution typing methods and then generating reliable nucleotide sequence data from high-throughput sequencing such as NGS.

The term "Korean native chicken" (KNC), as used in this study, comprises two major native chicken breeds: the KNC and Ogye chicken. The KNC breed carries five lines of gray, black, red, white, and yellow. The Ogye breed is also native to Korea and considered part of the national heritage. Hereafter, they are collectively referred to as six KNC lines.

These KNC lines are becoming increasingly popular due to their characteristic flavor and high meat quality (Manjula et al., 2020; Nawarathne et al., 2020). Although they were highly valued by Koreans in the past, their populations were substantially decreased during the Korean war. Based on collaborative efforts by the Korean government, scientific community, and livestock farmers, there is now a trend towards their commercial farming. Because disease resistance plays a vital role in chicken farming, it is very important to investigate the MHC region of KNC to prepare them as commercial breeds with a high disease resistance. Previous MHC studies on KNC lines have shown that they have a unique MHC diversity, while also share a few common haplotypes with commercial chicken breeds (Manjula et al., 2020). The presence of homozygous individuals with novel haplotypes in KNC pure line populations creates an opportunity to further investigate their haplotype diversity by developing reliable sequence information from a large section of MHC-*B*.

We analyzed six novel BSNP-MHC haplotypes (Manjula et al., 2020) of six KNC lines that are homozygous for both SNP and microsatellite markers, using long-range PCR and NGS methods. The consensus sequences of these novel BSNP haplotypes provide reference data to better understand new variants in KNC. Accordingly, an NGS-based MHC typing technique could be developed for local chicken breeds.

Materials and methods

DNA samples

Six birds, one each from six KNC lines, homozygous for both LEI0258 microsatellite marker and 90 MHC-*B* SNP panel described in our previous study (Manjula et al., 2020), from the National Institute of Animal Science (South Korea) were used. Before the investigation began, the birds were further analyzed for chromosome 16, which contains MHC genes, using the 600 K SNP chip to assess their MHC class I and class II variants (unpublished data). Birds that were homozygous for all markers were accepted for sequencing.

Haplotypes distinguished by the SNP panel and microsatellite typing included BSNP-B03 (249/249), BSNP-Kr11 (193/193), BSNP-Kr15 (193/193), BSNP-J06 (474/474), BSNP-B03 (249/249), and BSNP-Kr31 (417/417) from the Korean gray, black, red, white, yellow, and Ogye lines, respectively (Manjula et al., 2020). Genomic DNA was extracted as described previously (Manjula et al., 2020).

Long-range PCR (LR-PCR) amplifications

A total of 16 pairs of primers were used, including 10 LR-PCR pairs described in Hosomichi et al. (2008) and six new pairs designed using the Prime3 Software available on the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC = BlastHome). They were used for the LR-PCR amplifications of

TABLE 1 Raw NGS data and assembly statistics for six KNC lines with known and novel variants obtained for the MHC-*B* region (variants with an Ensembl-based rs-number were considered known variants).

Feature		KNC line					
		Black	Gray	Red	White	Yellow	Ogye
Total read bases		1,468,886,324	2,244,799,220	2,261,956,444	2,199,870,982	2,127,926,126	2,155,747,272
Total reads		9,727,724	14,866,220	14,979,844	14,568,682	14,092,226	14,276,472
GC (%)		59.54	60.26	60.28	60.23	60.74	60.42
Length (bp)		69,269	72,456	72,680	69,069	71,522	69,190
Identical sites		55,749	43,914	44,371	44,264	43,870	44,316
Pairwise identity (%)		81.5	70.8	75.1	70.1	70.2	85.6
Known variants (*BG1 – BF2*)	SNPs	477	402	427	484	428	373
	Indels	10	3	4	3	5	5
Novel variants (*BG1 – BF2*)	SNPs	92	185	148	106	108	132
	Indels	69	59	71	65	61	57
Total number of variants (*BG1 – BF2*)		648	649	650	658	602	567

15 MHC genes (*BG1—BF2*) belonging to the extended class I, classical class I and class II, covering approximately 69 kb of the chicken MHC-*B* region.

Initially, all 16 PCR products were amplified using LR-PCR primers designed based on the MHC-*B* reference sequence AB268588.1 (Shiina et al., 2007). All of the LR-PCR amplifications were carried out using Takara PrimeSTAR polymerase (Takara Bio, Japan) based on optimized two-step and three-step standard protocols (Cat. #R050A). In brief, the 20 μL PCR reaction volume included 100 ng template DNA, 1 μL (10 pmol) of each forward and reverse primers, 4 μL 5 × GL buffer, 1.6 μL dNTP, 0.4—0.5 μL Prime STAR polymerase, and distilled water. The size of the LR-PCR was 4,780 kb on average, with a range from 1,345 to 9,437 bp (Supplementary Table S1).

To measure the success of LR-PCR amplifications, PCR products were checked on a 0.8% agarose gel and visualized by staining with ethidium bromide. The PCR products were purified using a PCR purification kit (Genet Bio, Daejeon, South Korea) and the concentrations were obtained using a NanoDrop device (Thermo Fisher Scientific, NanoDrop, 2000C). The normalization of PCR product concentration and PCR product pooling were conducted based on the equimolar pooling method. Pooled PCR products of BSNP-Kr11 (193/193) from the black line were sent to the TNT research facility (South Korea) for library preparation and NGS.

After obtaining the NGS results, areas containing gaps or ambiguities were identified by the mapping procedure described later in the methodology. To further optimize the results for the remaining five lines, seven new LR-PCR primer pairs were prepared based on AB268588.1 (Shiina et al., 2007) and the NGS consensus of the BSNP-Kr11 (193/193) sample. Accordingly, a total of 23 LR-PCR products were obtained for the remaining five samples following the same PCR amplification protocol.

Library preparation and NGS for PCR amplicons

Purified and pooled PCR products were sent for NGS. After performing DNA quality control using PicoGreen (cat.#P7589; Invitrogen) and the DNA High Sensitivity Chip (Bioanalyzer), qualified samples were used for library construction using a TruSeq DNA PCR-Free (350) kit, according to the TruSeq DNA PCR-Free Sample Preparation Guide (Part #15036187 Rev. D) and sequenced with the 6000 S4 Reagent kit (ver. 1.5; NovaSeq; 300 cycles) using the Illumina NovaSeq platform.

NGS data pre-processing and assembly

The NGS reads were assembled using the Geneious Prime molecular biology tools and subjected to sequence analysis using RRID:SCR_010519 (ver. 2022.0.1; Geneious).

The raw NGS reads were imported into the Geneious Prime tool (Illumina read technology), and paired reads were set with the relative orientation of forward/reverse inward-pointing (Illumina paired-end) with an insert size of 350 bp. Then, they were trimmed and normalized using BBDuk Adapter/

Quality trimmer (version 38.84) and BBNorm (version 38.84), respectively.

In the trimming process, the default parameters were right end trimming with a K-mer size of 27, with the maximum substitutions set to one and maximum 'substitutions + indels' set to 0. The low-quality bases, i.e., quality lower than 20, were trimmed from both ends and the minimum overlap was set to 20 when trimming adapters on paired read overhangs. Short reads with a maximum length of 10 bp were discarded. Two-pass normalization was conducted by adhering to the default settings where the target coverage level/target normalization depth and minimum depth were adjusted to 40 and 6, respectively. The number of threads was set to 12 while the K-mer size was 31.

The *Gallus* genes, MHC region, and partial and complete coding sequences (CDSs) (accession number: AB268588.1) available in the NCBI and published by Shiina et al. (2007) were used as the reference and annotated directly against the NCBI database using the Geneious Prime GenBank accession tools for both genes and CDSs. Then the targeted MHC-*B* region was extracted.

Next, the normalized reads were assembled with the "map to reference" technique using Geneious mapper under medium sensitivity, as per the instructions provided in the product manual. Paired read overhangs were trimmed and gaps were allowed (maximum per read = 15%). However, trimming from the Geneious mapper prior to the mapping was restricted because the trimming was conducted with BBDuk in a previous step. The program was set to generate contigs, a consensus sequence, and an assembly report for each assembly. Once the assembly was completed, the consensus sequences were obtained as the DNA sequence for the relevant portion of MHC-*B* for each of the KNC lines separately. Then they were computationally annotated using the reference-based technique.

The same raw NGS data were analyzed again with the GATK best practices pipeline to obtain the variants. For this pipeline, the chromosome 16:GRCg6a:16:1:2844601:1, which is available in the Ensembl genome browser, was set as the reference (Ensembl.com, 2021). Then the variants were analyzed using the pandas library in Python.

Results

LR-PCR amplifications

Once the extracted DNA from the selected individuals of six KNC lines were amplified with the primers as described previously (targeting the *BG1—BF2*), the LR-PCR amplified products were checked *via* agarose gel electrophoresis. (Images for all six lines are shown in Supplementary Figure S1). The results confirmed that the products for all six lines were successfully amplified and had the expected fragment sizes.

NGS for PCR amplicons

The NGS reads were paired-end, with sequencing read length and insert size of 151 and 350 bp, respectively. Details of the sequencing results are presented in Table 1.

The use of a lesser number of primer pairs to produce PCR products resulted in a lesser number of total reads and read bases in the black line than the other five lines. The average total reads and read bases of the KNC lines except for the black line were 14,556,689 and 2,198,060,009, respectively. In all cases, the guanine-cytosine (GC) contents were between 59.54 and 60.74, whereas the adenine-thymine (AT) contents were close to 40%.

Pre-processing and assembly of NGS data

Assembly with Geneious Prime

All six lines resulted in read recoveries in excess of 93% after trimming, with the lowest value obtained for the yellow line (93.46%), and the highest obtained for the black line (96.58%). The reason behind the recovery of the highest percentage of reads after trimming in the black line may be due to the lesser number of reads generated compared to the other five lines, which was a consequence of the comparatively low number of primer pairs used.

When assembling the reads to the targeted MHC-*B* region of AB268588.1 (74,072 bp), the following statistics were obtained (Table 1). The numbers of identical sites between the KNC lines and the reference (AB268588.1) ranged from 43,870 in the yellow line to 55,749 in the black line. Furthermore, the pairwise identity percentages ranged from ~70% to ~86%, revealing noticeable deviations of KNC from the reference. An example of the visualization of a completed assembly, annotations, and the consensus sequence is given in Supplementary Figure S2.

Based on the Lander–Waterman equation (Lander and Waterman, 1988), the coverages of the assemblies for the KNC gray, black, red, white, yellow and Ogye breeds were approximately 146, 105, 138, 136, 136, and 123, respectively.

After their annotations, all of the lines were called for all of the expected genes, often with varying lengths due to insertions and deletions (indels; Table 2). The size of *Blec4* pseudogene (Shiina et al., 2007; Yuan et al., 2021) was the same for the reference and all six lines. In the case of *TAP2*, all KNC lines had a size of 2997 bp, unlike the reference *TAP2* (3037 bp). The same pattern was observed for *BRD2* (3,763 bp in all KNC lines vs. 3,762 bp in the reference). The sizes of the other 12 genes differed from the reference, at least in one KNC line.

TABLE 2 Summary of the sizes of MHC-*B* genes (from *BG1*-*BF2*) in KNC lines.

Gene	Reference (AB268588.1)	KNC line					
		Black	Gray	Red	White	Yellow	Ogye
BG1	4267 bp	4266 bp	4266 bp	4212 bp	4268 bp	4266 bp	4266 bp
Blec4	720 bp	720 bp	720 bp	720 bp	720 bp	720 bp	720 bp
Blec2	2493 bp	2490 bp	2481 bp	2490 bp	2543 bp	2481 bp	2484 bp
Blec1	2074 bp	2073 bp	2074 bp	2074 bp	2073 bp	2074 bp	2074 bp
BLB1	1364 bp	1357 bp	1359 bp	1364 bp	1364 bp	1358 bp	1365 bp
TAPBP	3450 bp	3443 bp	3445 bp	3443 bp	3449 bp	3445 bp	3452 bp
BLB2	1352 bp	1352 bp	1358 bp	1341 bp	1353 bp	1353 bp	1341 bp
BRD2	3762 bp	3763 bp	3763 bp	3763 bp	3763 bp	3763 bp	3763 bp
DMA	2087 bp	2088 bp	2089 bp	2089 bp	2087 bp	2089 bp	2089 bp
DMB1	1820 bp	1819 bp	1822 bp	1821 bp	1821 bp	1822 bp	1821 bp
DMB2	2937 bp	2982 bp	2938 bp	2957 bp	2946 bp	2945 bp	2946 bp
BF1	2034 bp	2023 bp	2009 bp	2024 bp	2013 bp	2024 bp	2032 bp
TAP1	4797 bp	4835 bp	4799 bp	4799 bp	4808 bp	4799 bp	4808 bp
TAP2	3037 bp	2997 bp	2997 bp	2997 bp	2997 bp	2997 bp	2997 bp
BF2	2017 bp	2014 bp	2008 bp	2022 bp	2019 bp	2017 bp	2017 bp

Variant calling with the GATK pipeline

The NGS reads were processed with the GATK pipeline in which the variants were called and annotated with SnpEff. The results were analyzed and simplified to aid understanding. We initially analyzed the known and novel variants for each line. The results are presented in Table 1.

The numbers of variants were 649, 648, 650, 658, 602, and 567 for the gray, black, red, white, yellow, and Ogye lines, respectively. All six KNC lines contained both known and novel variants. Accordingly, 37.6%, 24.9%, 33.7%, 26.0%, 28.1%, and 33.4% of the total variants were novel in the gray, black, red, white, yellow, and Ogye lines, respectively. More than 57% of them were SNPs (ranging from 57.2% in the black line to of 75.9% in the gray line) and others were indels.

The variants were further categorized by SnpEff (within the GATK pipeline) based on their impacts on protein synthesis. When the SnpEff analysis resulted in two or more transcripts for the same genes, the Ensembl canonical transcript was called for the gene of concern.

The genes in the MHC-*B* region (from *BG1* to *BF2*) were analyzed in terms of the impact of variants for each line (Figure 1 and Supplementary Table S2). The high-impact variants that caused immensely disruptive changes/modifications in the respective proteins could be found for only a few of the most important genes (*BF1*, *BF2*, *BLB1*, *BLB2*, and *BG1*) in all six lines.

BLB1, which is a minor class II gene (Jacob et al., 2000), was the only gene with high-impact variants in all six KNC lines. Other genes (*BF1*, *BF2*, *BLB2*, and *BG1*) carried high-impact variants only in some lines. Furthermore, *BLB1* had the most high-impact variants in all KNC lines, except for the red and white lines. Considering all impact types, *BLB2* always led *BLB1* in terms of the number of variants, confirming the higher number of polymorphisms in *BLB2*.

The other genes in which no high-impact variant was found did, however, derive moderate-, low-, and modifier-impact variants. Based on the impact variants (considering only the canonical transcripts), the fewest variants obtained for a single gene of a single line was recorded for *BG1* of the Ogye breed (70 variants), while the most was recorded for *TAP1* of the KNC gray line (250 variants).

For the variant types, upstream, downstream, or intron variants were obtained in the first three places based on the number of variants in each of the lines with different orders. The least number of variants per line was given by one or more of the following variant types: conservative inframe deletion, disruptive inframe insertion, disruptive inframe deletion, stop gain, 5′UTR premature start codon gain, splice donor, and splice acceptor frameshift. Stop gains were comparatively rare and accounted for the high-impact variants, while all lines had missense variants with moderate impacts.

Discussion

The MHC region in the chicken plays a key role in adaptive and innate immunity; in this region, the BF/BL

FIGURE 1
Impact of the variants on MHC-*B* genes for six KNC lines.

genes are predominantly involved in antigen presentation, and therefore determine the resistance/susceptibility to various diseases, such as Marek's disease (Kaufman et al., 1999) and avian influenza (Hunt et al., 2010).

NGS approaches have been applied in avian species that are closely related to chickens, such as quail (Kawahara-Miki et al., 2013) and turkeys (Dalloul et al., 2010), to study such variation. Although many scientists are currently working with NGS for chicken MHC analysis, to our knowledge, no previous study on the utility of LR-PCR combined with NGS for chicken MHC genotyping has been published.

When assembling the pre-processed NGS reads by mapping to the AB268588.1 reference, it was first directly annotated to the NCBI source data, particularly for its genes and CDSs. One of the annotation notes is shown in Supplementary Figure S2. This facilitated an easy extraction of the target reference region, thus improving assembly quality. The average pairwise identity between the assemblies of KNC lines and the reference provides evidence of a possible evolutionary divergence among them with respect to MHC-*B*.

As MHC class I and class II genes obtain significant importance among the MHC genes, some observations about them are highlighted. Among the MHC-class I genes, *BF2* is dominantly expressed (Parker and Kaufman, 2017) and highly polymorphic (Shaw et al., 2007). Considering the Ensembl canonical transcripts, *BF2* always had more variants than *BF1*, except for the gray line. We observed more polymorphisms in *BF1* than *BF2* in the KNC gray line, based on their canonical transcripts (Supplementary Table S2).

Because *BLB1* and *BLB2* are responsible for producing the MHC class II molecules, they are also vital for activating/initiating

adaptive immunity in chicken (Worley et al., 2008; Kaufman, 2021). The exon two of both *BLB1* and *BLB2* contributes to the fabrication of the peptide-binding groove/peptide binding region of the MHC class II molecules. Therefore, the polymorphisms in that region cause substantial changes in peptide binding affinity (Li et al., 2010). This study confirms the high number of polymorphisms in both MHC class II genes, and provides an insight into the modified form of MHC class II molecules in KNC lines compared to the reference red junglefowl.

These high-impact variants in highly polymorphic, functional, and important MHC genes provide clear insight of the differentiated immune responses in KNC lines compared to the reference red junglefowl. However, the actual effects of such variants on protein modifications are yet to be discovered, and it is not possible to comment on their benefits or drawbacks in effective immune responses.

Our data (Supplementary Table S2) showed that several variant types occur throughout the genes in all lines. These results suggest a very high diversity in KNC lines compared to the reference red junglefowl.

Consequently, the NGS technique can be used to reliably detect MHC-*B* variabilities, in contrast to some of the previously used marker-based methods. However, NGS reads with larger insert sizes may greatly increase the assembly quality by increasing the probability of alignment at the most accurate locations, and we therefore recommend the use of such NGS techniques.

Conclusion

We investigated a portion of the chicken MHC-*B* region (from *BG1* to *BF2*), including the classical class I and II genes, completely based on the NGS data of the six KNC lines. After quality control and assembly, their consensus sequences were derived from *BG1* to *BF2* for all six KNC lines. The variants were analyzed and numerous novel variants were revealed. The results indicated that all of the KNC lines were highly diverse and diverged from the reference red junglefowl.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found below: https://www.ncbi.nlm.nih.gov/genbank/, OM953772, OM953773, OM953774, OM953775, OM953776, and OM953777.

Ethics statement

The animal study was reviewed and approved by Committee on the Ethics of Animal Experiments, National Institute of Animal Science (Poultry Research Institute), South Korea. The blood sample collection procedures have been approved under the ethical approval number 2012-C-037 of the National Institute of Animal Science (Poultry Research Institute), South Korea. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

JHL administered in idea generation, conceptualization, and final manuscript reviewing. PM designed the primers, performed LR-PCR, and sent the purified PCR products for the next-generation sequencing. TKE conducted the NGS data pre-processing, assembly, variant calling, data analysis, data submission to GenBank, and manuscript writing, whereas EC and MK assisted with variant calling procedures and related troubleshooting. All authors read the article and approved the submitted version.

Funding

This study was supported by the National Research Foundation, Republic of Korea under the Grant Number MHC 2022R1F1A1064025 and the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-01441, Artificial Intelligence Convergence Research Center (Chungnam National University).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.886376/full#supplementary-material

References

Alizadeh, M., Picard, C., Frassati, C., Walencik, A., Gauthier, A. C., Bennasar, F., et al. (2020). A new set of reagents and related software used for NGS based classical and non-classical HLA typing showing evidence for a greater HLA haplotype diversity. *Hum. Immunol.* 81 (5), 202–205. doi:10.1016/j.humimm.2020.02.003

Chen, L., Fakiola, M., Staines, K., Butter, C., and Kaufman, J. (2018). Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. *Front. Immunol.* 9, 930. doi:10.3389/fimmu.2018.00930

Dalloul, R. A., Long, J. A., Zimin, A. V., Aslam, L., Beal, K., Ann Blomberg, L., et al. (2010). Multi-platform next-generation sequencing of the domestic Turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol.* 8 (9), e1000475. doi:10.1371/journal.pbio.1000475

Ensembl (2021). Ensembl.com. Available at: https://www.ensembl.org/Gallus_gallus/Info/Index. [Accessed January 18 2022].

Fulton, J. (2020). Advances in methodologies for detecting MHC-B variability in chickens. *Poult. Sci.* 99 (3), 1267–1274. doi:10.1016/j.psj.2019.11.029

Fulton, J. E., McCarron, A. M., Lund, A. R., Pinegar, K. N., Wolc, A., Chazara, O., et al. (2016). A high-density SNP panel reveals extensive diversity, frequent recombination and multiple recombination hotspots within the chicken major histocompatibility complex B region between BG2 and CD1A1. *Genet. Sel. Evol.* 48 (1), 1–15. doi:10.1186/s12711-015-0181-x

Goto, R. M., Wang, Y., Taylor, R. L., Wakenell, P. S., Hosomichi, K., Shiina, T., et al. (2009). *BG1* has a major role in MHC-linked resistance to malignant lymphoma in the chicken. *Proc. Natl. Acad. Sci. U. S. A.* 106 (39), 16740–16745. doi:10.1073/pnas.0906776106

Hosomichi, K., Miller, M. M., Goto, R. M., Wang, Y., Suzuki, S., Kulski, J. K., et al. (2008). Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. *J. Immunol.* 181 (5), 3393–3399. doi:10.4049/jimmunol.181.5.3393

Hunt, H. D., Jadhao, S., and Swayne, D. E. (2010). Major histocompatibility complex and background genes in chickens influence susceptibility to high pathogenicity avian influenza virus. *Avian Dis.* 54 (1), 572–575. doi:10.1637/8888-042409-ResNote.1

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 BF/BL region of the chicken major histocompatibility complex. *Immunogenetics* 51 (2), 138–147. doi:10.1007/s002510050022

Ka, S., Lee, S., Hong, J., Cho, Y., Sung, J., Kim, H. N., et al. (2017). HLAscan: Genotyping of the HLA region using next-generation sequencing data. *BMC Bioinforma.* 18 (1), 258. doi:10.1186/s12859-017-1671-3

Kaufman, J. (2021). Innate immune genes of the chicken MHC and related regions. *Immunogenetics* 74 (1), 167–177. doi:10.1007/s00251-021-01229-2

Kaufman, J., Milne, S., Göbel, T. W., Walker, B. A., Jacob, J. P., Auffray, C., et al. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401 (6756), 923–925. doi:10.1038/44856

Kawahara-Miki, R., Sano, S., Nunome, M., Shimmura, T., Kuwayama, T., Takahashi, S., et al. (2013). Next-generation sequencing reveals genomic features in the Japanese quail. *Genomics* 101 (6), 345–353. doi:10.1016/j.ygeno.2013.03.006

Lander, E. S., and Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* 2 (3), 231–239. doi:10.1016/0888-7543(88)90007-9

Li, X., Han, L., and Han, J. (2010). No specific primer can independently amplify the complete exon 2 of chicken *BLB1* or *BLB2* genes. *Int. J. Poult. Sci.* 9 (2), 192–197. doi:10.3923/ijps.2010.192.197

Manjula, P., Fulton, J. E., Seo, D., and Lee, J. H. (2020). Major histocompatibility complex B variability in Korean native chicken breeds. *Poult. Sci.* 99 (10), 4704–4713. doi:10.1016/j.psj.2020.05.049

Mwambene, P. L., Kyallo, M., Machuka, E., Githae, D., and Pelle, R. (2019). Genetic diversity of 10 indigenous chicken ecotypes from Southern Highlands of Tanzania based on Major Histocompatibility Complex-linked microsatellite LEI0258 marker typing. *Poult. Sci.* 98 (7), 2734–2746. doi:10.3382/ps/pez076

Nawarathne, S. R., Lee, S. K., Cho, H. M., Wickramasuriya, S. S., Hong, J. S., Kim, Y. B., et al. (2020). Determination and comparison of growth performance parameters between two crossbred strains of Korean native chickens with a white semi broiler chicken for 84 days post-hatch. *Korean J. Agric. Sci.* 47 (2), 255–262. doi:10.7744/kjoas.20200016

Ozaki, Y., Suzuki, S., Kashiwase, K., Shigenari, A., Okudaira, Y., Ito, S., et al. (2015). Cost-efficient multiplex PCR for routine genotyping of up to nine classical HLA loci in a single analytical run of multiple samples by next generation sequencing. *BMC Genomics* 16, 318. doi:10.1186/s12864-015-1514-4

Parker, A., and Kaufman, J. (2017). What chickens might tell us about the MHC class II system. *Curr. Opin. Immunol.* 46, 23–29. doi:10.1016/j.coi.2017.03.013

Shaw, I., Powell, T. J., Marston, D. A., Baker, K., van Hateren, A., Riegert, P., et al. (2007). Different evolutionary histories of the two classical class I genes *BF1* and *BF2* illustrate drift and selection within the stable MHC haplotypes of chickens. *J. Immunol.* 178 (9), 5744–5752. doi:10.4049/jimmunol.178.9.5744

Shiina, T., Briles, W. E., Goto, R. M., Hosomichi, K., Yanagiya, K., Shimizu, S., et al. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *J. Immunol.* 178 (11), 7162–7172. doi:10.4049/jimmunol.178.11.7162

Worley, K., Gillingham, M., Jensen, P., Kennedy, L., Pizzari, T., Kaufman, J., et al. (2008). Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. *Immunogenetics* 60 (5), 233–247. doi:10.1007/s00251-008-0288-0

Yuan, Y., Zhang, H., Yi, G., You, Z., Zhao, C., Yuan, H., et al. (2021). Genetic diversity of MHC BF/BL region in 21 chicken populations. *Front. Genet.* 12, 710–770. doi:10.3389/fgene.2021.710770

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.932272

OPEN ACCESS

EDITED BY
Lingyang Xu,
Institute of Animal Sciences (CAAS), China

REVIEWED BY
Sangang He,
Xinjiang Academy of Animal Science, China
Syed Shanaz Shafi,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

*CORRESPONDENCE
E. F. Dzomba,
Dzomba@ukzn.ac.za

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

RECEIVED 29 April 2022
ACCEPTED 22 November 2022
PUBLISHED 04 January 2023

CITATION
Dzomba EF, Van Der Nest MA, Mthembu JT, Soma P, Snyman MA, Chimonyo M and Muchadeyi FC (2023), Selection signature analysis and genome-wide divergence of South African Merino breeds from their founders.
Front. Genet. 13:932272.
doi: 10.3389/fgene.2022.932272

Selection signature analysis and genome-wide divergence of South African Merino breeds from their founders

E. F. Dzomba[1]*, M. A. Van Der Nest[2], J. N. T. Mthembu[1], P Soma[3], M. A. Snyman[4], M. Chimonyo[5] and F. C. Muchadeyi[2]

[1]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa, [2]Agricultural Research Council Biotechnology Platform, Private Bag X5 Onderstepoort, Pretoria, South Africa, [3]Agricultural Research Council, Animal Production and Improvement, Pretoria, South Africa, [4]Grootfontein Agricultural Development Institute, Middelburg, South Africa, [5]Discipline of Animal and Poultry Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa

Merino sheep are a breed of choice across the world, popularly kept for their wool and mutton value. They are often reared as a pure breed or used in crossbreeding and are a common component in synthetic breed development. This study evaluated genetic diversity, population structure, and breed divergence in 279 animals of Merino and Merino-based sheep breeds in South Africa using the Illumina Ovine SNP 50K BeadChip. The sheep breeds analysed included the three Merino-derived breeds of Dohne Merino (n = 50); Meatmaster (n = 47); and Afrino (n = 52) and five presumed ancestral populations of Merinos (Merino (n = 46); South African Merino (n = 10); and South African Mutton Merino (n = 8)); and the non-Merino founding breeds of Damara (n = 20); Ronderib Afrikaner (n = 17); and Nguni (n = 29). Highest genetic diversity values were observed in the Dohne Merino (DM), with H_o = 0.39 ± 0.01, followed by the Meatmaster and South African Merino (SAM), with H_o = 0.37 ± 0.03. The level of inbreeding ranged from 0.0 ± 0.02 (DM) to 0.27 ± 0.05 (Nguni). Analysis of molecular variance (AMOVA) showed high within-population variance (>80%) across all population categories. The first principal component (PC1) separated the Merino, South African Mutton Merino (SAMM), DM, and Afrino (AFR) from the Meatmaster, Damara, Nguni, and Ronderib Afrikaner (RDA). PC2 aligned each Merino-derived breed with its presumed ancestors and separated the SAMM from the Merino and SAM. The *iHS* analysis yielded selection sweeps across the AFR (12 sweeps), Meatmaster (four sweeps), and DM (29 sweeps). Hair/wool trait genes such as *FGF12*; metabolic genes of *ICA1*, *NXPH1*, *and GPR171*; and immune response genes of *IL22*, *IL26*, *IFNAR1*, and *IL10RB* were reported. Other genes include *HMGA*, which was observed as selection signatures in other populations; *WNT5A*, important in the development of the skeleton and mammary glands; *ANTXR2*, associated with adaptation to variation in climatic conditions; and *BMP2*, which has been reported as strongly selected in both fat-tailed and thin-tailed sheep. The DM vs. SAMM shared all six sweep regions on chromosomes 1, 10, and 11 with AFR vs. SAMM. Genes such as *FGF12* on OAR 1:191.3–194.7 Mb and *MAP2K4* on OAR 11:28.6–31.3 Mb were observed. The selection sweep on

chromosome 10 region 28.6–30.3 Mb harbouring the *RXFP2* for polledness was shared between the DM vs. Merino, the Meatmaster vs. Merino, and the Meatmaster vs. Nguni. The DM vs. Merino and the Meatmaster vs. Merino also shared an Rsb-based selection sweep on chromosome 1 region 268.5–269.9 Mb associated with the *Calpain* gene, *CAPN7*. The study demonstrated some genetic similarities between the Merino and Merino-derived breeds emanating from common founding populations and some divergence driven by breed-specific selection goals. Overall, information regarding the evolution of these composite breeds from their founding population will guide future breed improvement programs and management and conservation efforts.

KEYWORDS

Merino-type sheep, population genetic structure, breed divergence, SNP genotypes, selection sweeps, EHH signatures

Introduction

Sheep serves as an important source of mutton, manure, and wool, playing an important role in the economy of the country (Groeneveld et al., 2010; Department Agriculture Forestry and Fisheries–South Africa, 2015). Sheep is the ideal farm animal for smallholder farmers due to its small body size that makes it an easily disposable source of meat or cash (Hassan et al., 2014). Furthermore, sheep are sold to meet financial obligations, and their ability to survive in harsh weather conditions allows resource-poor farmers to depend on sheep for food and human livelihood (Hassan et al., 2014; Edea et al., 2017). Sheep also fulfil different socio-cultural roles (Wilson, 2011).

The Merino sheep breed is regarded as one of the oldest and most economically influential sheep breeds in the world (Al-Atiyat et al., 2016). The Merino breeds are known for their fine and soft wool (Department Agriculture Forestry and Fisheries–South Africa, 2015). In South Africa, the breed was introduced in the 1780s from Spain (Buduram, 2004; DAD-IS) and has become adapted to South African climatic and environmental conditions. The South African Merino (SAM) is believed to be a composite breed developed from Spanish, Saxony, Rambouillet, American, and Australian sheep breeds (Mason, 1996) that likely evolved to carry different genes that confer adaptation to particular production environments (Peters et al., 2010). Coupled to selection within breed, several Merino-based breeds have been developed for either wool or mutton or as dual-purpose breeds (Hlophe, 2011).

In South Africa, the Merino breed contributed to the development of composite breeds such as the Afrino (AFR), Meatmaster, and Dohne Merino (DM). The DM breed was developed by crossing the German Mutton Merino (commonly known as SAMM and the SAM ewes) as dual-purpose animals for both meat and wool production (Kotze, 1951; Buduram, 2004; Naidoo et al., 2005; Jordan, 2013). DM sheep are amongst the leading wool sheep breeds in South Africa; they are hardy animals that are well-adapted to their local environments (e.g., resistant to parasites, particularly *Heamonchus contortus*) (Dlamini et al., 2019; Synman and Fisher, 2019). The AFR sheep breed originated during the depressed wool market of the late 1960s when farmers began crossbreeding Merino ewes with mutton breeds. These dual-purpose animals reared for both meat and wool production originated from the crossbreeding of Ronderib Afrikaner (RDA), SAM, and SAMM in a targeted ratio of 25:25:50 of the respective contributing breeds that thrive in the harshest conditions (Bezuidenhout, 2012). The Meatmaster is a breed developed to improve the meat qualities of the fat-tailed sheep breeds (Mason, 1996). Literature indicates that the Meatmaster is a composite of many breeds, though it was predominantly developed from SAMM, SAM, Damara, and other indigenous breeds (Mason, 1996; Hlophe, 2011).

In South Africa, the Merino sheep were bred with local breeds in an effort to improve productivity and resilience of the breeds to the harsh local conditions whilst producing optimally (Bezuidenhout, 2012; Dlamini et al., 2019; Synman and Fisher, 2019). With climate change and other production challenges, the rationale of crossbreeding and developing new and composite breeds will prevail, thus requiring an investigation and documentation of the genomic architecture of the current composite breeds and their evolution from the ancestral populations. It is, therefore, worthwhile for inventory purposes and to guide future breed development initiatives to investigate the breed relationships and differentiations and the genomic regions targeted by selection through breed development of the SAM and Merino-derived breeds. The study by Ciani et al. (2015) suggested that intensive gene flow, founder effects, and geographic isolation are the main factors that determined the genetic makeup of current Merino and Merino-derived breeds.

The Ovine SNP50 genotyping array and other similarly designed bead chips provide unprecedented power to scan the genomes of livestock and investigate footprints of selection and their impact on the genetic potential of breeds to meet designed

production goals. With the advent of genome-wide SNP genotyping, different statistical methods have been developed to interrogate genomes for signatures of selection and the associated effects on phenotypes. Signatures of selection in a genome are usually associated with either high-frequency-derived alleles or highly differentiated allele frequencies between populations or long-range haplotypes with strong linkage disequilibrium (LD) (Grossman et al., 2010). Statistical methods such as the within-population-integrated haplotype homozygosity score (*iHS*) (Voight et al., 2006) and the between-population *Rsb* and *XP-EHH* test (Tang et al., 2007) have been developed to screen for high LD and long-range haplotype and infer on signatures of selection. The hapFLK package detects selection signatures based on differences in haplotype frequencies between all the populations (Fariello et al., 2014). A number of studies have used *iHS*, *Rsb*, and other similar methods in the investigation of selection sweeps in sheep (Paim et al., 2018; Alvarez et al., 2020), cattle (Chen et al., 2016; Tijani et al., 2019), and other livestock species.

A number of previous studies have contributed towards the genomic characterisation of Merino and Merino-derived breeds. The study by Liu et al. (2022) used whole-genome sequences of 10 South African Mutton Merino (SAMM) sheep together with 39 Australian Merino and Chinese Merino (wool-type Merino) sheep to identify selection signatures using the iHS and *XP-EHH* methods. On the other hand, the study by Megdiche et al. (2019) used a multi-cohort approach, comparing wool-type Merino-derived breeds with non-Merino-derived breeds raised in the same geographical regions using F_{ST} outlier methods, local ancestry approaches, and genome-wide patterns of distribution of runs of homozygosity (ROH) to infer on selection signatures. The South African Merino-derived breeds are a composite of Merino breeds and non-Merino-type breeds of predominantly fat-tailed Damara, Nguni, and Ronderib Afrikaner and other local breeds. In order to complement and add more information on Merino and Merino-derived breeds, our study included the presumed non-Merino and presumed ancestors of the Merino-derived breeds in the analysis. The addition of the non-Merino-type breeds, which is unique to this study, allowed for a comprehensive analysis and interpretation of the divergence of the Merino-derived breeds from their presumed ancestors. Therefore, this study aimed to investigate the population genetic structure, breed similarities, and divergence of Merino-derived sheep breeds in South Africa. As such, this study first investigated population structure and admixture levels by referencing the composite breeds of AFR, DM, and Meatmaster against their presumed ancestral breeds of Merino, South African Merino (SAM), South African Mutton Merino (SAMM), Ronderib Afrikaner (RDA), Damara, and Nguni. Based on this analysis, the study went on to investigate regions habouring selection sweeps within the composite breeds and between each composite breed and its presumed ancestors. In addition to the iHS (Voight et al., 2006) and *Rsb* and *XP-EHH* (Tang et al., 2007) methods used by Liu et al. (2022), which investigated signatures of selection within populations and between pairs of populations, respectively, this study used the hapFLK (Fariello et al., 2014) method for a global analysis of signatures of selection in a data set consisting of both Merino-type and the non-Merino presumed ancestors. SNP genotypes for this analysis were limited to those generated in previous studies (Nxumalo et al., 2018; Dzomba et al., 2020) and the International Sheep Genomics Consortium HapMap data (http://www.sheephapmap.org). The study hypothesised that crossbreeding, followed by intensive selection towards breed-specific production goals, resulted in genomic divergence between the South African Merino and Merino-derived breeds.

Materials and methods

Animal genotypes

The study used Ovine SNP50K genotype data from a total of 279 animals obtained from five different sheep populations consisting of Merino (n = 46), SAMM (n = 8), SAM (n = 10), DM (n = 50), AFR (n = 52) and Meatmaster (n = 47), Nguni = (n=29), RDA (n =17) and Damara (n =20). The AFR, DM, Meatmaster, SAM, and SAMM genotypes were obtained from samples kept in a biobank at Grootfontein College of Agriculture, South Africa, together with the Nguni sheep that were sampled from the KwaZulu-Natal region of South Africa, and the SNP genotype data were generated and reported in a previous study by Dzomba et al. (2020). The Damara sheep genotypes were provided from a separate study (Nxumalo et al., 2018), while the RDA and Merino genotypes were extracted from the ISGC (http://www.sheephapmap.org). The AFR, DM, and Meatmaster are the Merino-derived composite breeds, and their presumed ancestral breeds based on literature (Synman, 2014a-d) are presented in Table 1.

Genotype data quality control

In this study, the Illumina OvineSNP 50K BeadChip genotypes (as reported by Nxumalo et al. (2018); Dzomba et al. (2020); and ISGC, http://www.sheephapmap.org) were subjected to quality control using PLINK v1.07 (Purcell et al., 2007) and Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) to ensure all SNPs had less than 5% missing genotypes, a call rate more than 95%, a minor allele frequency (MAF) less than 5%, and in Hardy–Weinberg equilibrium ($p < 0.0001$). Additional quality control measures ensured that individual animals had an IBD <0.025. High levels

TABLE 1 South African Merino-derived breeds and their ancestral breeds.

Composite breed	SAMM (8)	SAM (10)	Merino (46)	RDA (17)	Damara (20)	Nguni (29)
Afrino (52)	X	X	X	X		
Dohne Merino (50)	X	X	X			
Meatmaster (47)	X	X			X	X

of pairwise linkage disequilibrium (LD) between SNPs may affect both the performance and efficiency of genomic prediction models. Therefore, in order to minimise bias for SNPs in strong linkage disequilibrium ($r^2 > 0.2$), the second SNP was removed, leaving a dataset with SNP pairs whose $r^2 < 0.2$ (Purcell et al., 2007).

Determination of within-breed genetic diversity

To determine the expected (H_E) and observed heterozygosity (*Ho*) values, PLINK v1.07 (Purcell et al., 2007) was used by running the command "*--hardy*" on the data for each breed. Inbreeding coefficients were calculated as the difference between expected (H_E) and observed heterozygosity (*Ho*) values divided by the expected heterozygosity (H_E) values also in PLINK v1.07. The mean H_E and *Ho* values per breed were calculated using the PROC MEANS procedure in SAS (2013).

Analysis of molecular variation

Analysis of molecular variance (AMOVA) was used to determine the genetic variance within populations (F_{IS}), among populations within groups (F_{SC}), and among groups (F_{CT}) using Arlequin v3.5 (Excoffer and Lischer, 2009). The populations were categorised into (i) all nine breeds; (ii) composite breeds of the AFR, DM, and Meatmaster; (iii) the presumed ancestral populations of the Nguni, Damara, RDA, SAMM, SAM, and Merino; and (iv) each composite breed and its presumed ancestors, which included (a) the AFR and Ronderib Afrikaner, Merino, and SAMM; (b) the Meatmaster, Damara, Nguni, SAMM, and SAM; and (c) the DM, SAM, SAMM, and Merino.

Analysis of population structure

Principal component analysis (PCA) was carried out using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The eigen values and eigen vectors for the principal components were estimated using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com).

Inference of local genomic ancestry (PCAdmix)

PCAdmix v1.0 (Brisbin et al., 2012) was used to infer local genomic ancestry in the composite breeds. The program utilises haplotypes from ancestral representatives to infer the ancestry of focal individuals. In this study, the SAMM, SAM, Merino, and RDA were treated as the ancestral representatives of the AFR sheep, the SAMM, SAM, and Merino breeds were treated as the ancestral representatives of the DM, and the SAMM, SAM, Damara, and Nguni were treated as the ancestral representatives of the Meatmaster. The software algorithms perform the inference chromosome-wide through PCA, *via* short windows along each chromosome. Using a hidden Markov model, PCAdmix then returns the posterior probability (PP) of ancestry from each reference population for each haploid individual in each window. PCAdmix requires phased genotypes, which were obtained using fastPHASE v1.2 (Scheet and Stephens, 2006) with default parameters.

Analysis of selective sweeps and differentiating genomic regions

Selection signature analysis was used to assess genome-wide signatures of selection in the composite Merino-derived sheep of South Africa. The hapFLK package v1.2 was used to detect selection signatures based on differences in haplotype frequencies between all the Merino-derived breeds included in this study (Fariello et al., 2014). The number of haplotype clusters (*K*) was calculated using the imputeqc R package and accompanied scripts (Khvorykh and Khrunin, 2020). Using the number of haplotype clusters, the hapFLK values and the kinship matrix were calculated in the fastPHASE model (*-K* 40). Since the implementation of this approach required the construction of a neighbour-joining (NJ) tree from using a kinship matrix, Reynolds genetic distances were converted into the kinship matrix using an R script supplied with the package. In the construction of the NJ tree, Afrino was used as the outgroup population. The hapFLK statistic was computed as the average value across 40 expectation maximization (EM) runs to fit the LD model. The *p*-values were obtained by running a Python script "Scaling_chi2_hapFLK.py" available at forge-dga.

jouy.inra.fr/documents/588, which fits a chi-squared distribution to the empirical distribution. The Manhattan plot was rendered in R using the qqman package (Turner, 2014) (settings for the suggestive line q = 0.03 and genome-wide line q = 0.01) to indicate selection.

The selection signature analysis also included complementary extended haplotype-based statistics (EHH-based statistics). The integrated haplotype score (iHS) approach compares EHH between alleles at an SNP within a population (Voight et al., 2006), while the extended haplotype-based homozygosity (*Rsb*) and cross population extend haplotype homozygosity (*XP-EHH*) approaches compare EHH patterns of the same allele between two populations (Tang et al., 2007). The ancestral alleles required for the computation were inferred as the most common alleles in the entire dataset, as previously described (Bahbahani et al., 2015). Haplotypes were phased using Beagle (Browning and Browning, 2007) and used to calculate *iHS* scores for each SNP/haplotype within a breed/population and *Rsb* and *XP-EHH* scores for each SNP/haplotype between breeds/populations. Haplotype frequencies were calculated using sliding windows of 20 SNPs that overlapped by five SNPs. For each locus, the *iHS*, *XP-EHH*, and *Rsb* scores were computed using the REHH package (Gautier and Vitalis, 2012) in R. For the analysis of within-population (*iHS*) and between-population differences (*Rsb* and *XP-EHH*), a score > 3 (i.e., a $-\log_{10}$ 3 score corresponding to a two-sided p-value < 0.001) was used to infer the candidate genomic regions under selection.

Mapping the region of differentiation to find genes

Under selection SNPs *($p<0.001$)* were mapped for genes using the Ensembl genome browser and NCBI (NCBI; www.ncbi.nlm.nih.gov). Ensembl Ovine (*Ovis aries*) genome build OAR3 was implemented in Golden Helix SVS v8 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Candidate genes were considered if their boundaries fell within 75 kb up or downstream of the selection sweep region defined. The associated genomic regions were also annotated using the Sheep QTL database (www.animalgenome.org/cgi-bin/QTLdb/OA/summary).

Results

Genetic diversity

All 277 animals proceeded for further analysis following quality control. The number of SNPs retained for analysis ranged from 36,976 in the Ronderib Afrikaner to 37,671 in the AFR (Table 1). Highest genetic diversity values were observed in the DM with H_o = 0.39 ± 0.01 followed by the Meatmaster and SAM with H_o = 0.37 ± 0.03. Lowest diversity was observed in the Nguni with H_o = 0.28 ± 0.02. Inbreeding estimates ranged from 0.00 ± 0.02 in the DM to 0.27 ± 0.05 in the Nguni (Table 2).

Analyses of molecular variation in pure and developed breeds

Table 3 illustrates the partitioning of variation within breeds, among breeds, and among breeds within the categories of (i) ancestral breeds, (ii) composite breeds, and (iii) each composite breed and its presumed ancestors. Within-population variation was found to be 90% in the composite breeds and 84% within the presumed ancestral breeds, whilst it was 83% within all breeds (Table 3). A high level of molecular variation was observed within populations in comparison to among populations and among individuals within populations. In the category of a breed and its presumed ancestors, the highest within-population variation was observed in the AFR and its presumed ancestors (92%), followed by the DM category (90%) and least in the Meatmaster category (77%). Among-breed variation within groups was highest in the Meatmaster (21%) and its presumed ancestors, followed by the group of ancestral breeds (15%), and least in the AFR and its presumed ancestors (6.4%). Within-breed variation was highest in the category consisting of all eight breeds (17%) and least in the category with ancestral breeds.

Analysis of population structure

The PCA results that explained the population structure (i.e., PC1 and PC2) explained 43% of the total variance (Figure 1). While PC1 (28% of the variation) separated the Merino, SAM, SAMM, DM, and AFR from the Meatmaster, Damara, Nguni, and Ronderib Afrikaner, PC2 (15% of the variation) separated the Damara, Meatmaster, Merino, SAM, and DM from the SAMM, Ronderib Afrikaner, and AFR. In PC1, the AFR was on the same axis as its Merino ancestors, but separated from the RDA. The DM clustered with the two Merinos, while the Meatmaster, on the other hand, clustered in the same axis with its Nguni and Damara presumed ancestors separated from the Merinos.

PCAdmix-based analysis of co-ancestry

The PCAdmix results are illustrated in Figure 2. Using the PCAdmix algorithm, the genome of each composite Merino-derived breed was partitioned into segments of inferred ancestry at a resolution of chromosomal level. The PCAdmix of the AFR yielded tracts of ancestry consistent to predominantly SAMM

TABLE 2 Expected and observed heterozygosity in five sheep breeds of South Africa.

Breed	Number of animals	Number of SNPs	$H_E \pm SD$	$H_O \pm SD$	F_{IS}
Afrino	52	37,671	0.39 ± 0.00	0.36 ± 0.01	0.06 ± 0.02
Meatmaster	47	36,586	0.39 ± 0.00	0.37 ± 0.03	0.05 ± 0.07
Merino	46	37,686	0.39 ± 0.00	0.35 ± 0.01	0.07 ± 0.03
SA Merino	10	37,452	0.39 ± 0.00	0.37 ± 0.03	0.04 ± 0.08
SA Mutton Merino	8	37,614	0.39 ± 0.00	0.34 ± 0.02	0.12 ± 0.04
Dohne Merino	50	37,638	0.39 ± 0.00	0.39 ± 0.01	0.00 ± 0.02
Damara	20	37,626	0.39 ± 0.00	0.31 ± 0.02	0.19 ± 0.04
Ronderib Afrikaner	17	36,976	0.39 ± 0.00	0.33 ± 0.03	0.14 ± 0.07
Nguni	29	37,634	0.39 ± 0.00	0.28 ± 0.02	0.27 ± 0.05
All breeds	279	37,381	0.39 ± 0.00	0.35 ± 0.03	0.08 ± 0.09

TABLE 3 Analysis of molecular variation.

Dataset	Variance component (%)		
	Among breeds (F_{CT})	Among breeds within groups (F_{SC})	Within breeds (F_{IS})
All eight breeds	10.62 (17.23%)	-	91.52 (82.77%)
Ancestral breeds	0.011 (1.06%)	0.149 (14.77%)	0.158 (84.17%)
Composite breeds	0.029 (2.97%)	0.071 (6.91%)	0.099 (90.11%)
Afrino and presumed ancestors	0.014 (1.42%)	0.065 (6.43%)	0.079 (92.15%)
Meatmaster and presumed ancestors	0.021 (2.12%)	0.218 (21.38%)	0.235 (76.50%)
Dohne Merino and presumed ancestors	0.048 (90.21)	0.052 (4.95%)	0.098 (90.21%)

(49.7%, 2.8 ± 0.03) followed by Merino (28.3%, 2.5 ± 0.03) and RDA (21.9%, 2.3 ± 0.02) (Figure 2A), consistent with a targeted ratio of 50:25:25 of the respective contributing breeds. The DM that was developed from crossing SAMM and the SAM ewes was predominantly Merino (37.4%, 2.4 ± 0.03) and SAMM (39.8%, 2.6 ± 0.03) and less of the SAM (22.8%, 2.4 ± 0.02) (Figure 2B). The Meatmaster, a composite of many breeds (Peters et al., 2010), was largely Nguni (41.2%, 2.4 ± 0.04) and Damara (32.0%, 2.9 ± 0.03) and less of the Merino (26.8%, 1.8 ± 0.02) breeds (Figure 2C).

Signatures of selection

The analysis of within-population iHS >3.0 identified 21 selection sweep regions distributed across 14 chromosomes (OAR 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16, and 19) in the DM sheep (Figure 3; Table 4; Supplementary Table S1). Using this method, fewer selection sweep regions in the Afrino and Meatmaster were identified. Only nine selection sweep regions distributed across eight chromosomes (OAR 1, 2, 3, 4, 6, 8, 13, and 19) were identified in the Afrino, while only four selection sweep regions across four chromosomes (OAR 1, 2, 3, and 9) were identified in the Meatmaster (Figure 3).

The analysis between the composite Merino-derived breed and each of its presumed ancestors (Rsb and XP-EHH >3.0) identified 25 selection sweep regions distributed across 15 chromosomes (OAR 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, and 19) (Figure 4; Table 4; Supplementary Table S2, 3). Of the 25 selection sweep regions, 18 were identified using the *iHS* (>3.0) method. hapFLK results between the composite Merino-derived breed and each of its presumed ancestors identified eight selection sweep regions distributed across five chromosomes (OAR 1, 5, 10, 13, and 25) (Figure 5; Table 4; Supplementary Table S4). All five of the selection sweep regions identified using hapFLK analyses were also identified using the Rsb and XP-EHH methods (>3.0).

Numerous genes and QTLs were identified in selection sweep regions. These included genes *CAPN7*, *IFNAR1*, *IL10RB*,

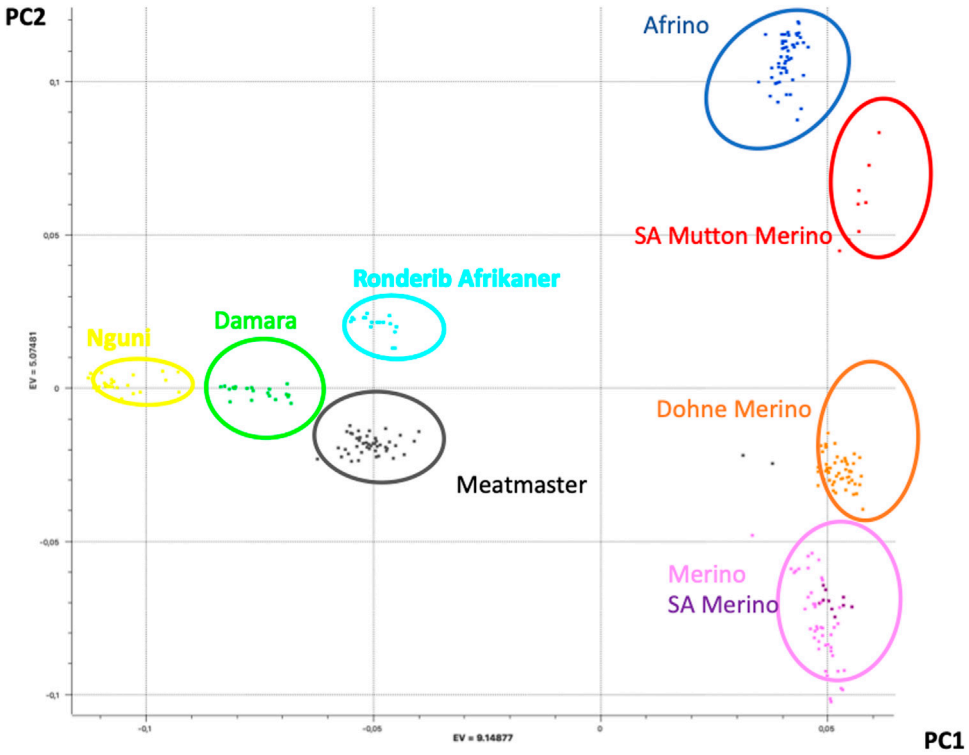


FIGURE 1
PCA-based clustering of Merino, Merino-derived breeds, and non-Merino presumed ancestors.

SLC2A2, *SLC5A3*, *SLC7A14*, *CRYZL1*, *FGF12*, *GPR171 GHSR*, *and SPATA16* that were within the sweep on the OAR1 region and *HMGA*, *IL22*, and *IL26* on OAR3, while genes *ICA1* and *NXPH1* were observed within the OAR4 sweep; *FGF5* and *ANTXR2* on OAR6; *RXFP2*, *Y2*, *SLITRK4*, and *GPC* on OAR10; *MAP2K4*, *SPTSSB*, *PPMIL*, and *B3GALNT* on OAR11; *BMP2* on OAR13; and the *WNT5A* gene on OAR19. The associated genomic regions annotated using the Sheep QTL database (release 48) (www.animalgenome.org/cgi-bin/QTLdb/OA/summary) identified several QTLs associated with important health and production traits. This included QTLs associated with reproductive traits (e.g., reproductive seasonality, QTLID:16602, QTLID:16603, and QTLID:195222 and total lambs born, QTL: 130451); skeletal morphology and body size (e.g., carcass bone percentage, QTLID:14293), body weight at birth, QTLID:12934; body weight at 20 weeks, QTLID:193069; muscle density, QTLID:95864; carcass fat percentage, QTLID:14277; ear size, QTL:159964); milk yield and quality traits (e.g., milk fat yield, QTL:169583); horn size and type traits (e.g., horn circumference, QTL:161397 and horn type, QTL:161480), and immune response (e.g., *Trichostrongylus colubriformis* FEC, QTLID:12884 and QTLID:4155; *Haemonchus contortus* FEC, QTLID:19803; *Trichostrongylus* adult and larva count, QTLID:12899 and QTLID:12900).

Discussion

In conducting this study, we made use of available sheep genotypes from previous projects to make inferences on genetic diversity, breed relations, and divergence amongst the Merino-derived breeds and their presumed ancestors. Our data were drawn from previous studies reported for the SAMM, SAM, DM, Meatmaster, and AFR (Dzomba et al., 2020); for the Damara sheep (Nxumalo et al., 2018); and for the Ronderib Afrikaner and Merino sheep (Kijas et al., 2012). These Merino and Merino-derived sheep breeds dominate the South African sheep industry, contributing to mutton, wool, and other sheep by-products. Merino sheep originate from Spain and are primarily useful and highly prized for their wool. In South Africa, their use in livestock farming can be traced to the late 18th century when few founder ewes and rams were donated by the Dutch government for experimental purposes to the Cape Town government (merinosa.co.za/history/).

The current study focused to make inferences on genetic diversity. Observed and expected heterozygosity values together with the inbreeding coefficient were used to explain genetic diversity within the studied sheep breeds. The highest genetic diversity (H_E) was observed in the DM followed by the Meatmaster, SAM, and AFR. The DM, AFR, and Meatmaster

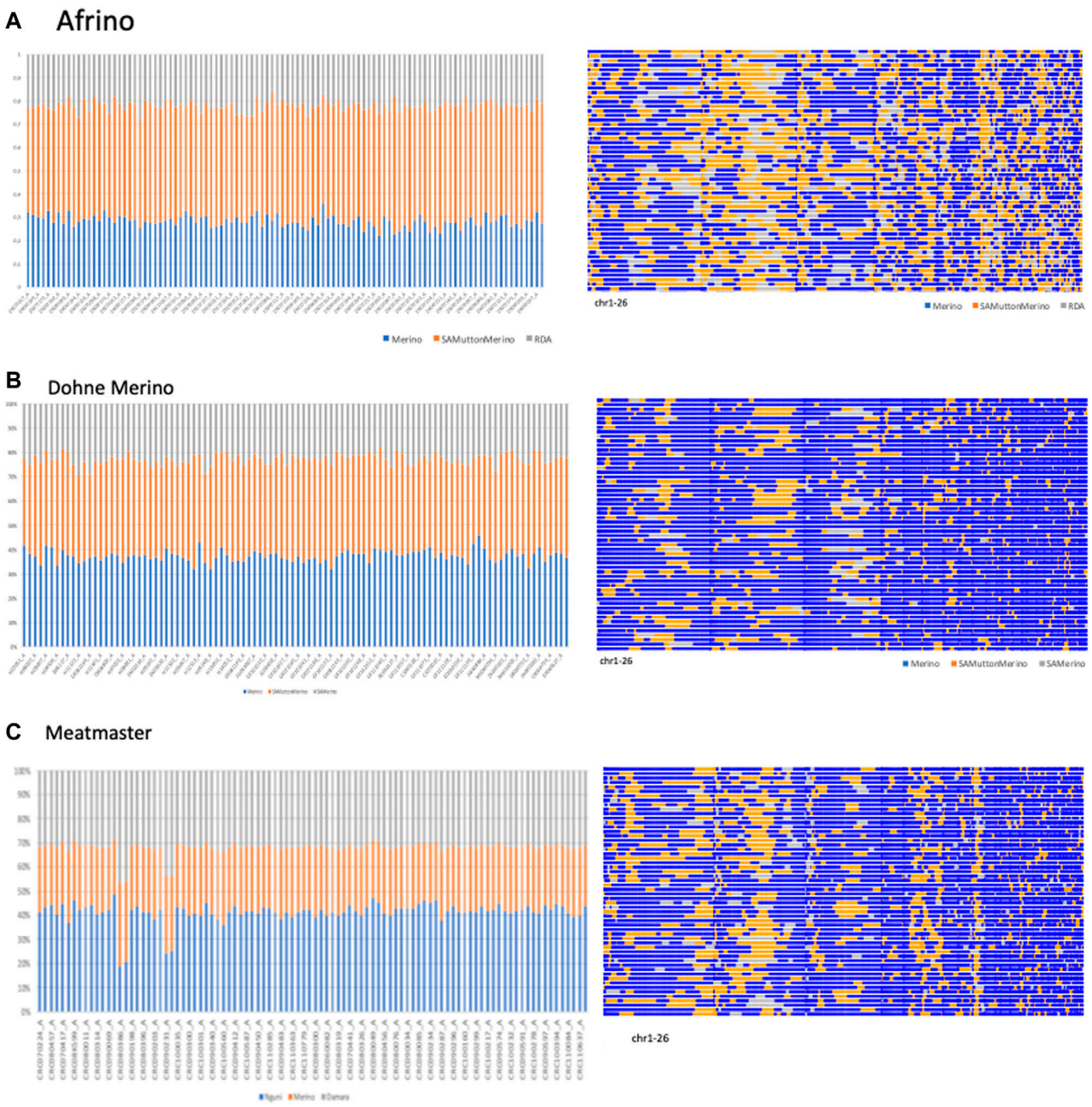


FIGURE 2
PCAmix of Merino-derived breeds [Afrino, **(A)**; Dohne Merino, **(B)**; and Meatmaster, **(C)**] and their presumed ancestors.

are the three Merino-derived composite breeds and are of high genetic diversity, similar to that reported in Spanish and Australian Merino breeds by Ciani et al. (2015). The Nguni, Damara, and Ronderib Afrikaner, which are the indigenous ancestral populations, are raised by few fragmented communities (Qwabe, 2011; Nxumalo et al., 2018; Selepe et al., 2018), which would explain the low levels of within-population genetic diversity. Although the founding populations of the Nguni, Damara, and Ronderib Afrikaner had low genetic diversity, crossbreeding them with Merino breeds to develop composite breeds resulted in increased diversity observed in the AFR and Meatmaster breeds, which can be attributed to the combination of two or more genetic pools. The results of this study clearly demonstrated that

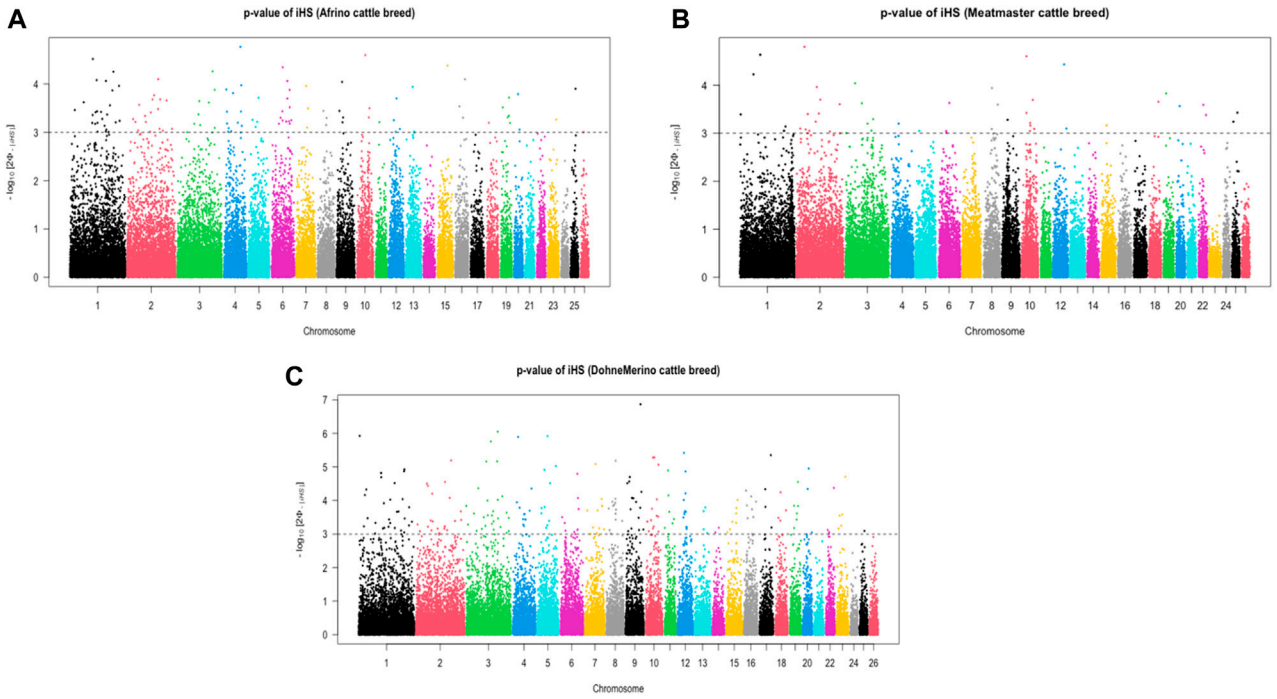


FIGURE 3
iHS scores of **(A)** Afrino, **(B)** Meatmaster, and **(C)** Dohne Merino.

significant genetic variation is maintained in the composite Merino-derived sheep breeds.

To gain an insight into the genetic structure of breeds, analysis of molecular variation (AMOVA) was used to determine the portioning of variance within and between populations and within populations amongst groups. In panmictic populations, the variance is expected to come from within samples (Excoffer and Lischer, 2009). If the variance occurs among samples within the population or comes from among populations, this would be regarded as evidence of the existence of population structure (Excoffer and Lischer, 2009). As expected, within-population variation was high in the composite breeds. The AFR and DM were established from predominantly well-managed commercial breeds of SAMM, Merino, and SAM and, to a small extent, the Ronderib Afrikaner in the case of the AFR. These breeds have moderate-to-high genetic diversity (Table 2), which explains the high within-population genetic variation in this category. The Meatmaster, on the other hand, is based on the small and less diverse breeds of the Damara and the Nguni, which is reflected in the relatively lower (77%) within-population diversity in this group. Significant population substructure was, therefore, evident in the Meatmaster and its presumed ancestors' category, with an among-breed diversity of 21%. These Merino-derived sheep breeds that exist as widely distributed admixed populations represent economically and historically important genetic resources (Ciani et al., 2015).

PCAdmix confirmed the presumed ancestry of the Merino-derived breeds of AFR, DM, and Meatmaster (Ciani et al., 2015). PC1 separated the Merino breeds from non-Merino breeds, with the exception of Meatmaster. While the Meatmaster was bred from fat-tailed sheep, there is intensive and directional selection against fat localisation and long tails in the breed (www.meatmasters.co.za). Coupled to this, part of the breed standards for the Meatmaster is that it should be 50% Damara (Synman, 2014c-d). These selection criteria explain its clustering with the Damara, Nguni, and Ronderib Afrikaner away from the Merinos and other Merino-derived breeds under PC1. In PC2 (15% of variation), the AFR clustered with the SAMM was separated from the DM, Merino, and SAM. According to the breed standards (http://www.afrino.org.za), 80% of the income from AFR is generated through meat production and 20% through wool production. This would be regarded as a biased selection objective towards growth and meat production traits, which explains why the AFR clustered with the SAMM. The Ronderib Afrikaner sheep are an improved form of the Namaqua Afrikaner sheep (Epstein, 1960), and together with the Damara and Nguni sheep are fat-tailed sheep (Peters et al., 2010), which could have formed the basis of their clustering

TABLE 4 *iHS*-based selection sweep regions and associated genes in Afrino, Dohne Merino, and Meatmaster sheep breeds and Rsb, XP-EHH and hapFLK-based selection sweep regions for Afrino, Dohne Merino, Meatmaster sheep breeds and their presumed ancestors.

	Start	Stop	Method	QTL trait (QTLID)*	Candidate gene
1	119.6	138.1	iHSDM; iHSAfrino	Muscle weight in carcass (14276), carcass fat percentage (14277, 14251, and 14277), lean meat yield percentage (14278), and bone weight in carcass (14275)	SLC5A3; MRPS6; IFNGR2; IFNAR1; IL10RB, PAXBP1; SYNJ1; MIS1BA; HUNK; SCAF4, S15; CHODL; BTG3; CXADR
1	168.6	201.5	iHSDM; iHSAfrino; Rsb_DM/SAMM; Rsb_Afrino/SAMM	Lean meat yield percentage (14278 and 14252), carcass fat percentage (14277 and 14251), reproductive seasonality (16602), and muscle weight in carcass (14320, 14270, and 14250)	*ALCAM*; *CBLB*; *HEG1*; *ZNF148*; *LRCH3*; *SLC49A4*; *SLC51A*; *AP; FGF12*; *GMNC*; *PLAAT1; IL1RAP; TPRG1; ATP13A4*
1	212.1	244.1	iHSMM; iHSDM; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_Afrino/Merino; Rsb_Afrino/RDA; XP-EHH_DM/Merino; XP-EHH_Afrino/Merino; XP-EHH_MM/Merino	Carcass fat percentage (14277), lean meat yield percentage (14252), and reproductive seasonality (16603)	*SPATA16; ECT2; GHSR; SLC2A2; SLC7A14; SHOX2; VEPH1; PTX3; CCNL1*
1	263.4	285.0	Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_MM/Nguni; XP-EHH_MM/Nguni; Rsb_MM/Merino; XP-EHH_MM/Merino; XP-EHH_Afrino/Merino hapFLK	Carcass fat percentage (14277), reproductive seasonality (16603), average daily gain (13948, 13955, and 13964) , and *Trichostrongylus colubriformis* FEC (12884)	*IGSF10; P2RY12; P2RY14; WWTR1; RNF13; GPR171; DIPK2A; SLC9A9; CH5T2; SETD4; CBR3; CAPN7*
2	6.2	26.0	iHSMM; iHSDM; Rsb_Afrino/RDA; XP-EHH_DM/Merino; XP-EHH_Afrino/Merino; XP-EHH_Afrino/SAMM; XP-EHH_MM/Merino	Body weight (57659, 14171, and 14280), average daily gain (57776), hot carcass weight (14279), hindquarter weight (14161), subcutaneous fat thickness (13722), loin fat weight (13732); and meat color (14163, 14167, and 14165)	*TMC1; ALDH1A1; FAM189A2; HPF1; CLCN3; NEK1*
2	62.0	109.5	iHSMM; iHSDM; XP-EHH_Afrino/RDA	No hit	
2	130.9	168.6	iHSDM; iHSAfrino XP-EHH_DM/Merino; XP-EHH_DM/SAMerino; XP-EHH_Afrino/Merino	Hot carcass weight (14279), meat color (14163, 14165, 14169, 14164, and 14168), longissimus muscle area/width/weight (13728, 13726, 13729), loin fat thickness (13730), and subcutaneous fat weight (13731 and 13738)	*LRP1B*
3	15.4	17.2	Rsb_DM/SAMerino; XP-EHH_DM/SAMerino	*Trichostrongylus colubriformis* FEC (14155), and body weight (56 weeks) (13927)	*SOX11; RSAD2; RNF144A*
3	45.9	74.3	Rsb_Afrino/RDA; Rsb_DM/Merino; XP-EHH_DM/Merino; XP-EHH_MM/Merino; XP-EHH_Afrino/Merino	Internal fat amount (14281) and body weight (56 weeks) (13927)	*LRRTM4; XPO1; FAM161A; NRXN1; U6*
3	114.5	153.3	iHSAfrino; iHSDM; iHSMM	Internal fat amount (14281 and 14255), *Trichostrongylus colubriformis* FEC (12885) , meat-conjugated linoleic acid content (17220), and body weight at birth (17230)	*SYT1; PPP1R12A; ZNF641; PFKM; SLC48A1; COL2A1; HDAC7; SLC38A4; SLS38A2; SLC38A1; ARID2; YEATS4; LYZ; KCNMB4; SLC35E3; RAP1B; CCT2; MYRFL; IL22; IL26*
4	16.1	63.5	iHSAfrino; iHSDM	Body weight (17232), *Haemonchus contortus* FEC (19803)	*ICA1; NXPH1; AVL9; NT5C3A; FKBP9; PDE1C*
5	23.8	85.5	iHSDM; Rsb_MM/Merino; hapFLK<	Body weight (birth) (12934) and body weight (20 weeks) (193069)	*SLC12A2; MEGF10; PHAX; TEX43; PRSS57; PLPPR3; RNF126; STK11; REDX01*
6	26.1	38.5	iHSDM; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA; XP-EHH_Afrino/Merino; XP-EHH_DM/Merino; XP-EHH_MM/Damara	Fat weight in carcass (95819 and 95820), total fat area (95836, 95837, 95826, 95827, and 95828), fat density (95850), and body weight (14261, 14284, 193062, 193068, and 193063)	*STPG2, CCSER1; MMRNA1; SNCA; FAM13A; ABCG2; SPP1; LCORl*
6	41.3	64.4	iHSAfrino; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA XP-EHH_DM/Merino; XP-EHH_MM/Nguni	Fat weight in carcass (95822 and 95823), total fat area (95838, 95839, and 95841), fat density (95840 and 95842), and muscle density (95864)	*GBA3; PPARGC1A; DHX15; SOD3; LG12*
6	93.5	102.3	iHSAfrino; iHSDM	Body weight (slaughter) (14284), *Trichostrongylus colubriformis* FEC (12887), reproductive seasonality (195222), and lean meat yield percentage (14286)	*GUF1; YIPF7; GNPDA2; GABRA4; ANTXR2; FGF5; GK2; FGFS; SLC10A6; PTPN13; MAPK10*
7	37.2	93.0	XP-EHH_MM/Damara; XP-EHH_DM/Merino	No hit	
8	47.4	51.8	iHSAfrino; iHSDM; XP-EHH_DM/Merino	*Trichostrongylus* adult and larva count (12899 and 12900) and internal fat amount (14288)	*LYRM2; BACH2; GJA10; GABRR1; RARS2; SLC35A; SPACA1*

(Continued on following page)

TABLE 4 (*Continued*) *iHS*-based selection sweep regions and associated genes in Afrino, Dohne Merino, and Meatmaster sheep breeds and Rsb, XP-EHH and hapFLK-based selection sweep regions for Afrino, Dohne Merino, Meatmaster sheep breeds and their presumed ancestors.

	Start	Stop	Method	QTL trait (QTLID)*	Candidate gene
9	31.8	40.7	iHSMM; iHSDM; XP-EHH_MM/Nguni	Hot carcass weight (14290) and longissimus muscle area (14323)	*SPIDR; PRKDC; MCM4, RGS20; SOX17; TGS1; TCEA1;*
10	14.5	32.5	Rsb_MM/Nguni; XP-EHH_MM/Nguni; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_MM/Merino; XP-EHH_MM/Merino; XP-EHH_DM/SAMerino; XP-EHH_Afrino/Merino	Tail fat deposition (127009), milk fat yield (QTL: 169583), ear size (QTL:159964), total lambs born (QTL:130451), horn circumference (QTL: 161397), and horn type (QTL:161480)	*RXFP2; RXFP2; FRY; BSGLCT*
10	34.5	59.4	iHSDM; Rsb_Afrino/RDA; Rsb_MM/Nguni; Rsb_DM/SAMM; Rsb_Afrino/SAMM; XP-EHH_DM/SAMM XP-EHH_MM/SAMM; P-EHH_Afrino/SAMM XP-EHH_Afrino/Merino; XP-EHH_MM/SAMerino; XP-EHH_Afrino/RDA; XP-EHH_MM/Damara XP-EHH_MM/Nguni	Fat weight in carcass (14292), carcass bone percentage (14293), carcass fat percentage (14294), and lean meat yield percentage (14295)	*SPRY2; SLITRK4; SLITRK5; SLC16A9; BORA; PIBF1; MZT1*
10	59.6	72.1	iHSDM; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA; Rsb_MM/Nguni; XP-EHH_MM/Nguni; XP-EHH_Afrino/Merino XP-EHH_Afrino/SAMM XP-EHH_MM/Damara XP-EHH_MM/SAMM; hapFLK	Body weight (57656), fat weight in carcass (14292), carcass bone percentage (14293), carcass fat percentage (14294), and lean meat yield percentage (14295)	*GPC5; SLITRK5; SLC16A9; GPR180; GPC5; TGDS; DCT; CTNNA3; GCD; TGDS; RUFY2; SLC25A16; COX20P*
10	78.6	81.7	Rsb_DM/Merino XP-EHH_DM/Merino Rsb_MM/Merino XP-EHH_MM/Merino XP-EHH_Afrino/Merino	Fat weight in carcass (14264), carcass fat percentage (14265), and lean meat yield percentage (14266)	*POGLUT2; SLC10A2*
11	21.5	38.2	Rsb_Afrino/SAMM; XP-EHH_Afrino/Merino; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_DM/SAMerino; Rsb_DM/SAMM; XP-EHH_DM/SAMM; Rsb_MM/Merino; XP-EHH_MM/Merino	Internal fat amount (14298), body weight (14297), average daily gain (13945 and 13966), and hot carcass weight (14296)	*MAP2K4; DNAH9; MYH3; MYH4; MYH8; NCOR1; ZNF624; SLC47A2; SLC5A12; MAPK7; FSHB; SOX15; SAT2; SNORA62; SOX; SLC13A5; TEKT1*
12	28.3	78.1	iHSDM; XP-EHH_Afrino/Merino; XP-EHH_DM/Merino	Body weight (yearling) (213860)	*VAMP4; FMO4; TNFSF18; ACOT7; NOL9; HES3; RNF207*
13	36.5	53.7	iHSAfrino; iHSDM; Rsb_Afrino/RDA; XP-EHH_ Afrino/RDA; XP-EHH_Afrino/Merino; hapFLK	Muscle weight in carcass (14301); tail fat deposition (127011)	BFSP1; RRBP1; DSTN; SNX5; KAT14; OVOL2; RASSF2; SLC23A2; BMP2
14	27	30.9	Rsb_MM/Nguni; XP-EHH_MM/Nguni	Dressing percentage (14304 and 14270), bone weight in carcass (14302), and fat weight in carcass (14269)	*CDH8*
16	33.5	45.8	iHSDM; XP-EHH_DM/SAMerino	Subcutaneous fat thickness/area (14309/14308), body weight (slaughter) (14306), and Dressing percentage (14305)	
19	25.8	27.3	iHSDM XP-EHH_DM/SAMerino	Average daily gain (193081)	
19	41.2	48.8	iHSAfrino; iHS|DM; Rsb_MM/Nguni; XP-EHH_MM/Nguni	Entropion (193397, 193378, and 193385)	*FAM3D; FAM107A; C3or167; ACOX2; WNT5A; IL17RD; IL17RB; ACTR8; DCP1A; SELENOK; CACNA1D; NEK4; SPCS1*

*The associated genomic regions were annotated using the Sheep QTL database (www.animalgenome.org/cgi-bin/QTLdb/OA/summary).

together in PC1. Even though PCA is a statistical method commonly used in population genetics to identify structure in the distribution of genetic variation across populations, PCA projections are strongly influenced by uneven sampling (McVean, 2009), which might have been the case in this study. Despite the uneven sample sizes, the PCA clustering observed in this study resembles that observed in a prior study (Dzomba et al., 2020) that included more populations and sample sizes from which the investigated data set was sub-sampled, which validates the results. Similar clustering of Merino-type and non-Merino-type breeds was reported in other studies (Gifford-Gonzalez and Hanotte, 2011; Kijas et al., 2012). In spite of the importance of these breeds, there is limited information on the genomic influence of either Merino or other indigenous sheep on the different composite Merino-derived breeds. Results of this study that increase our knowledge regarding Merino-based breeds will, therefore, inform and guide future breed improvement programs and management and conservation efforts.

iHS analysis is used to infer recent and generally segregating selection sweeps (Voight et al., 2006) and has been used in humans (Liu et al., 2013) and a number of livestock studies

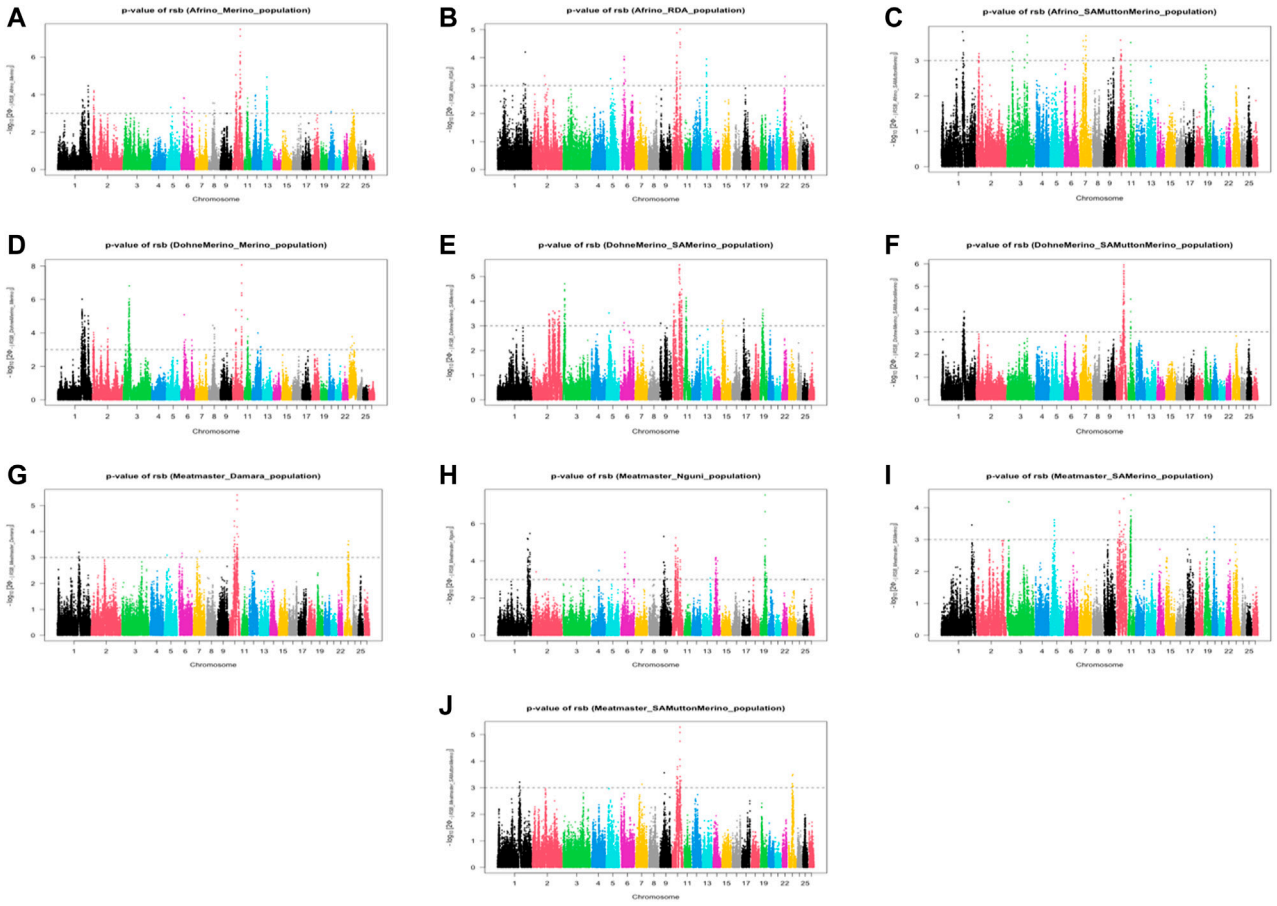


FIGURE 4
Genome-wide XP-EHH and Rsb scores of Merino-derived breeds and against their presumedancestors. Afrino and Merino **(A)**, Afrino and RDA **(B)**, Afrino and SAMuttonMerino **(C)**, Dohne Merino and Merino **(D)**, Dohne Merino and SAMerino **(E)**, Dohne Merino and SAMuttonMerino **(F)**, Meatmaster and Damara **(G)**, Meatmaster and Nguni **(H)**, Meatmaster and SAMerino **(I)** and Meatmaster and SAMuttonMerino **(J)**.

for dairy cattle (Cheruiyot et al., 2018), pigs (Chen et al., 2016),etc. For AFR, *FGF12,* a candidate gene for hair follicle development (Lv et al., 2020) and reproductive traits (An et al., 2018), and *ICA1 and NXPH1*, associated with metabolic pathways (Akanno et al., 2018), were some of the key candidate genes identified. Although the AFR is predominantly selected or weighted for meat quality traits, the breed was established as a white-woolen breed for use as a terminal sire when crossed with Merino ewes (Synman, 2014) in response to the presence of kemp (coloured fibre) in crosses of Merino ewes with mutton breeds. The signature for hair follicle development might be a reflection of this selection. *ICA1* and *NXPH1*, on the other hand, are signatures of the intensive selection weight put on the AFR for meat traits (Synman, 2014).

The four selection sweep regions identified in the Meatmaster breed corresponded to *GPR171* on chromosome 1 which is associated with feed and metabolism (Ruiz-Larranaga et al., 2018). The *HMGA* gene on chromosome 3 was also observed as a selection signature in Sardinian ancestral black sheep (Kijas et al., 2012) and in Spanish breeds (Manunza et al., 2016). *HMGA2* is involved in skeletal morphology and body size (Kijas et al., 2012) and has been shown to be under selection in dogs with divergent stature (Jones et al., 2008; Akey et al., 2010). According to the breed society standards (www.meatmastersa.co.za/Breed-Standard.htm), Meatmaster sheep must be of average size and have a functional, efficient body conformation and well-placed legs with excellent walking ability. Such selection for body size and skeletal morphology could be the signature presented through the *HMGA2* gene, which, together with the *GPR171* gene associated with feed and metabolism, could ensure optimal performance for mutton production. Also, *IL22* and *IL26* on chromosome 3 are immune response genes that have been reported as under selection in some studies, including

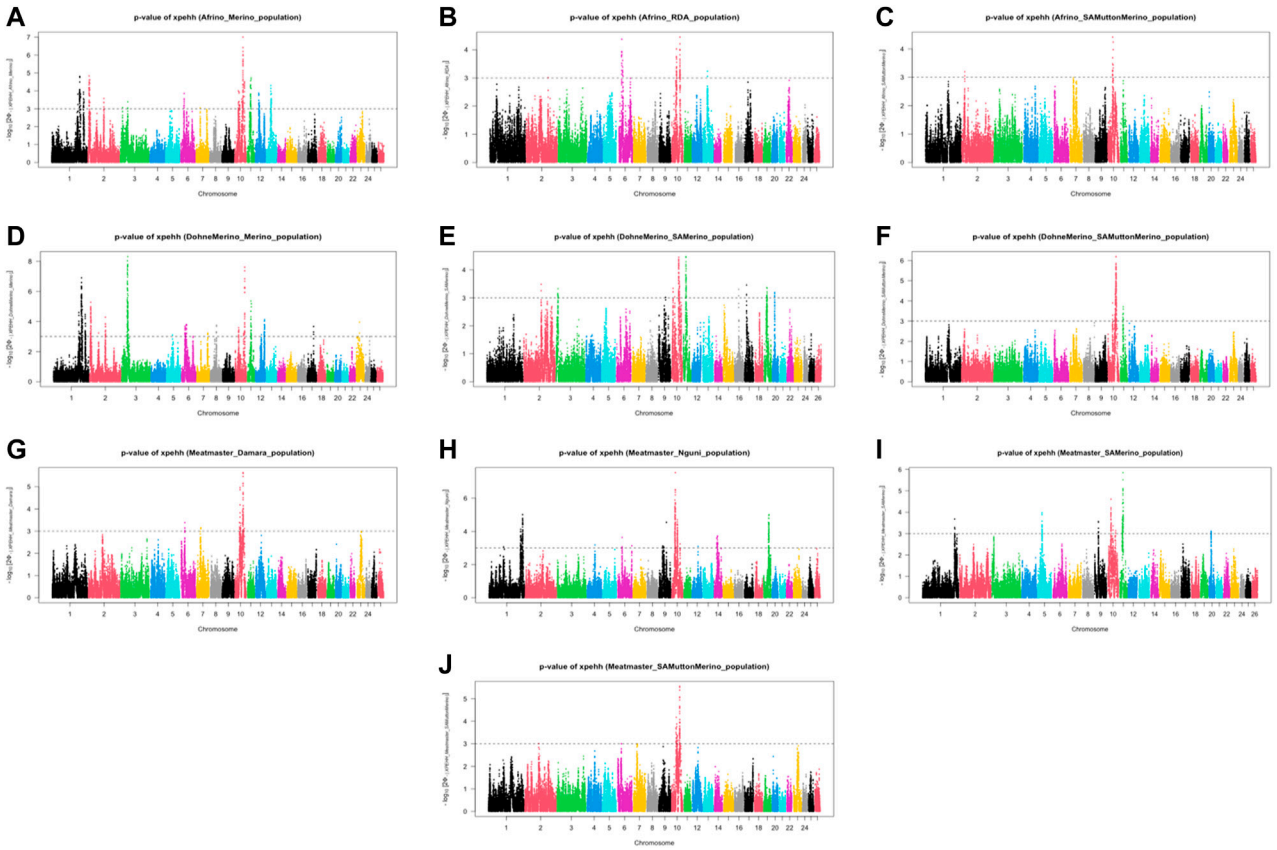


FIGURE 5
XP-EHH scores of Merino-derived breeds against presumed ancestors. Genome-wide XPGenome-wide XP-EHH scores of Merino-derived breeds and against ancestors. Afrino and Merino **(A)**, Afrino and RDA **(B)**, Afrino and **(C)**, Dohne Merino and Merino **(D)**, Dohne Merino and SAMerino Dohne Merino and SAMuttonMerino **(F)**, Meatmaster and Damara Meatmaster and Nguni **(H)**, Meatmaster and SAMerino **(I)** and Meatmaster and SAMuttonMerino **(J)**.

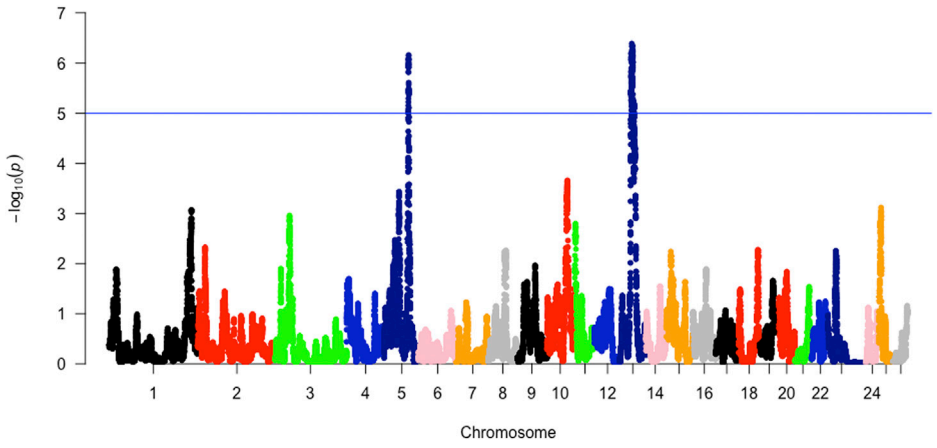


FIGURE 6
hapFLK scores of Merino-derived breeds against presumed ancestors.

Fariello et al. (2014). Detection of immune response genes is especially expected in breeds raised in arid environments with harsh and compromised production systems (Mdladla et al., 2018).

Interestingly, more selection sweep regions were identified in the DM sheep relative to the AFR and Meatmaster. DM shared two sweeps with AFR, one on chromosome 1 in regions 119, 0–121, and 0 Mbs, which harbours *IFNAR1, IL10RB, SLC5A3*, and *CRYZL1* associated with anti-inflammatory and immune response (Uemoto et al., 2020) and reproductive traits such as implantation of the conceptus to the uterus (Zhang et al., 2013), and one on chromosome 19 position in regions 46, 7–48, and 1 Mbs, which harbours the gene *WNT5A* (45 and 5 Mb), important in morphology, particularly the development of limbs and skeleton (Fariello et al., 2014) and reproductive traits (including mammary gland development) (Hao et al., 2019). Other sweeps on chromosome 6 (93, 5–95, and 2 Mb) associated with the *FGF5* gene are reported as a signature of selection in worldwide sheep breeds by Kijas et al. (2012) and Fariello et al. (2014), and the *ANTXR2* gene is associated with adaptation to variation in climatic conditions (Lv et al., 2020). A selection sweep on chromosome 13 on regions 46, 5–48, and 1 Mb in the AFR and 49, 1–50, and 9 Mb in the DM was associated with the gene *BMP2*, which has been reported as a signature of selection by Kijas et al. (2012) and Fariello et al. (2014) and is strongly selected in both fat-tailed and thin-tailed sheep (Dong et al., 2020).

Rsb, XP-EHH, and hapFLK results presented selection sweeps between a composite Merino-derived breed and each of its presumed ancestors. Sweeps on OAR 1 yielded genes such as the *GHSR* important for growth and carcass traits in sheep (Bahrami et al., 2012); *SPATA16* is associated with environmental variables in goats (Mdladla et al., 2018) and male fertility in cattle (Wang et al., 2014); and the *SLC7A14* and *SLC2A2* are involved in nutrient transport and absorption (Wiedemar et al., 2015). The selection sweep on chromosome 10 in regions 28, 6–30, and 3 Mb is associated with the *RXFP2* gene associated with polledness (Wang et al., 2014; Wiedemar and Drogenmuller, 2015). The region carried other genes such as the *FRY* gene, which is associated with lambing percentage, ear size, and coat phenotypes (Wei et al., 2015), and the *BSGLCT* gene, which is associated with wool traits. Zhang et al. (2013) suggested the role of *FRY* in sheep wool development. Overall, the Rsb, XP-EHH, and hapFLK analyses revealed the direction of selection when these breeds were selected, which focused on meat and wool production and robustness of breed through body confirmation, disease resistance, and adaptability to the harsh production conditions in South Africa (Kim et al., 2016; Molotsi et al., 2017).

Although fairly documented, there is limited information on the genomic influence of either Merino or other indigenous sheep on the different composite Merino-derived breeds. This study provided requisite information on the evolution of these composite breeds from their founding populations, which will inform and guide future breed improvement programs and management and conservation efforts.

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 the Animal Ethics Board of the University of KwaZulu-Natal.

Author contributions

Conceptualization: ED; formal analysis: ED, MV, and FM; investigation: ED and JM; original draft preparation: ED; review and editing: ED, FM, MV, JM, PS, MS, and MC. All authors have read and agreed to the final version of the manuscript.

Funding

The work was funded by the National Research Foundation of South Africa and the University of KwaZulu-Natal Competitive Grant.

Acknowledgments

The Grootfontein Agricultural Development Institute (GADI), South Africa, provided some DNA samples for some of the sheep breeds from their biobank. We also used DNA samples from Dr. Pranisha Soma's sheep research studies. SNP genotyping was carried out at the Agricultural Research Council - Biotechnology Platform. We acknowledge the International Sheep Genome Consortium for the Ovine SNP50K genotypes from global populations used in this study.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.932272/full#supplementary-material

References

Akanno, E. C., Chen, L., and Abo-Ismail, M. K., (2018). Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. *Genet. Sel. Evol.* 50, 48. doi:10.1186/s12711-018-0405-y

Akey, J. M., Ruhe, A. L., Akey, D. T., Wong, A. K., Connelly, C. F., Madeoy, J., et al. (2010). Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1160–1165. doi:10.1073/pnas.0909918107

Al-Atiyat, R., Flood, W., Franklin, I., Kinghorn, B., and Ruvinsky, A. (2016). Microsatellite-based genetic variation and differentiation of selected Australian Merino sheep flocks. *Small Rumin. Res.* 136, 137–144. doi:10.1016/j.smallrumres.2016.01.018

Álvarez, I., Fernández, I., Traoré, A., Perez-Pardal, L., Menendez-Arias, N. A., and Goyache, F. (2020). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Sci. Rep.* 10, 2824. doi:10.1038/s41598-020-59839-x

An, X., Ma, H., Han, P., Zhu, C., Cao, B., and Bai, Y. (2018). Genome-wide differences in DNA methylation changes in caprine ovaries between oestrous and dioestrous phases. *J. Anim. Sci. Biotechnol.* 9, 85. doi:10.1186/s40104-018-0301-x

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

Bahrami, A., Miraei-Ashtiani, S. R., and Mehrabani-Yeganeh, H. (2012). Associations of growth hormone secretagogue receptor (GHSR) genes polymorphisms and protein structure changes with carcass traits in sheep. *Gene* 505 (2), 379–383. doi:10.1016/j.gene.2012.06.009

Bezuidenhout, R. (2012). Bringing back the Afrino. Farmers weekly. Available at: www.farmersweekly.co.za/animals/cattle/bringing-back-the-afrino/.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., et al. (2012). PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* 84, 343–364. doi:10.3378/027.084.0401

Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi:10.1086/521987

Buduram, P. O. (2004). *Genetic characterization of Southern African sheep breeders using DNA markers*. MSc thesisDepartment of Animal, Wildlife and Grassland Sciences. University of the Free State.

Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., et al. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genet. Sel. Evol.* 48, 76. doi:10.1186/s12711-016-0254-5

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., and Mujibi, F. D. N. (2018). Signatures of selection in admixed dairy cattle in Tanzania. *Front. Genet.* 9, 607. doi:10.3389/fgene.2018.00607

Ciani, E., Lasagna, E., D'Andrea, M., Alloggio, I., Marroni, F., Ceccobelli, S., et al. (2015). Merino and merino-derived sheep breeds: A genome-wide intercontinental study. *Genet. Sel. Evol.* 47, 64. doi:10.1186/s12711-015-0139-z

Department Agriculture Forestry and Fisheries–South Africa (DAFF) (2015). A profile of the South African wool market value chain. E-print. Available from www.daff.gov.za/daffweb3/branches/economic-development-trade-marketing/marketing/annual-publications.

Dlamini, N. M., Visser, C., Snyman, M. A., Soma, P., and Muchadeyi, F. C. (2019). Genomic evaluation of resistance to *Haemonchus contortus* in a South African Dohne Merino flock. *Small Rumin. Res.* 175, 117–125. doi:10.1016/j.smallrumres.2019.04.020

Dong, K., Yang, M., Han, J., Ma, Q., Han, J., Song, Z., et al. (2020). Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. a major target of fat-tail selection in sheep. *BMC Genomics* 21, 800. doi:10.1186/s12864-020-07210

Dzomba, E. F., Chimonyo, M., Snyman, M. A., and Muchadeyi, F. C. (2020). The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. *Anim. Genet.* 51, 910–923. doi:10.1111/age.12991

Edea, Z., Dessie, T., Dadi, H., Do, K. T., and Kim, K. S. (2017). Genetic diversity and population structure of Ethiopian sheep populations revealed by high-density SNP markers. *Front. Genet.* 8, 218. doi:10.3389/fgene.2017.00218

Epstein, H. (1960). History and origin of the Ronderib and Namaqua afrikaner sheep. *Z. für Tierzüchtung Züchtungsbiologie* 74, 267–292. doi:10.1111/j.1439-0388.1960.tb00132.x

Excoffier, L., and Lischer, H. E. L. (2009). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Fariello, M. I., Servin, B., Tosser-Klopp, G., Rupp, R., and Moreno, C. (2014). Selection signatures in worldwide sheep populations. *PLos One* 9, e103813. doi:10.1371/journal.pone.0103813

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

Gifford-Gonzalez, D., and Hanotte, O. (2011). Domesticating animals in Africa: Implications of genetic and archaeological findings. *J. World Prehist.* 24, 1–23. doi:10.1007/s10963-010-9042-2

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B,, Pilling, D, et al. (2010). Genetic diversity in farm animals - a review. *Anim. Genet.* 41, 6–31. doi:10.1111/j.1365-2052.2010.02038.x

Grossman, S. R., Shlyakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327, 883–886. doi:10.1126/science.1183863

Hassan, D. I., Mbap, S. T., and Naibi, S. A. (2014). Socio-economic characteristics of Yankasa sheep and West African dwarf goat's farmers and their production constraint in Lafia Nigeria. *Int. J. Food, Agric. Veterinary Sci.*, 227782–209X93. http://www.cibtech.org/jfav.htm2015Vol.5(1) January-April.

Hlophe, S. R. (2011). *Genetic variation between and within six selected south African sheep breeds using random amplified polymorphic DNA and protein markers*. PhD Thesis. Faculty of Science and Agriculture University of Zululand. www.semanticscholar.org/paper/Genetic-variation-between-and-within-six-selected-Hlophe/8fbfb389f50090df8b0d5f655f3ff24015b73c65.

Jones, P., Chase, K., Martin, A., Davern, P., Ostrander, E. A., and Lark, K. G. (2008). Single- nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* 179, 1033–1044. doi:10.1534/genetics.108.087866

Jordan, W. (2013). *Enhancing the breed analysis of the DM by accounting for heterogeneous variances and phantom parents*. Stellenbosch, South Africa: Stellenbosch University.

Khvorykh, G. V., and Khrunin, A. V. (2020). imputeqc: an R package for assessing imputation quality of genotypes and optimizing imputation parameters. *BMC Bioinforma.* 21, 304–314. doi:10.1186/s12859-020-03589-0

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R. P., 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, e1001258. doi:10.1371/journal.pbio.1001258

Kim, E., Elbeltagy, A., Aboul-Naga, A., 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

Kotzé, J. J. J. (1951). The development of a mutton woolled sheep for the sour-grassveld area. *Farming S. Afr.* 28, 110–113.

Liu, X., Ong, R. T-H., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., et al. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *Am. J. Hum. Genet.* 92 (6), 866–881. doi:10.1016/j.ajhg.2013.04.021

Liu, Z., Bai, C., Shi, L., He, Y., Hu, M., Sun, H., et al. (2022). Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data. *Anim. Genet.* 53, 224–229. doi:10.1111/age.13173

Lv, X., Chen, W., Sun, W., Hussain, Z., Wang, S., and Wang, J. (2020). Analysis of lncRNAs expression profiles in hair follicle of hu sheep lambskin. *Animals* 10, 1035. doi:10.3390/ani10061035

Manunza, A., Cardoso, T. F., Noce, A., Martinez, A., and Pons, A. (2016). Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. *Sci. Rep.* 6, 27296. doi:10.1038/srep27296

Mason, I. L. (1996). *A world dictionary of livestock breeds, types and varieties*. 4th Edition. CAB International Wallingford UK.

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genet.* 165 (10), e1000686. doi:10.1371/journal.pgen.1000686

Mdladla, K., Dzomba, E., and Muchadeyi, F. (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* 120, 369–378. doi:10.1038/s41437-017-0044-z

Megdiche, S., Mastrangelo, S., Ben Hamouda, M., Lenstra, J. A., and Ciani, E. (2019). A combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in merino and merino-derived sheep breeds. *Front. Genet.* 10, 1025. doi:10.3389/fgene.2019.01025

Molotsi, A. H., Taylor, J. F., Cloete, S. W., Muchadeyi, F., Decker, J. E., Whitacre, L. K., et al. (2017). Genetic diversity and population structure of South African smallholder farmer sheep breeds determined using the OvineSNP50 beadchip. *Trop. Anim. Health Prod.* 49 (8), 1771–1777. doi:10.1007/s11250-017-1392-7

Naidoo, P., Cloete, S. W. P., and Fossey, A. (2005). South African merinos divergently selected for multiple rearing ability: A preliminary study of divergence based on RAPD markers. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 16, 254. doi:10.4314/sajas.v44i4.1

Nxumalo, K. S., Grobler, J. P., Ehlers, K., Banga, C. B., and Mapholi, N. O. (2018). "Genomic diversity and population structure of four South African indigenous Sheep," in *Proceedings of the world congress of genetics applied to livestock production*. http://www.wcgalp.org/system/files/proceedings/2018/genomic-diversity-and-population-structure-four-south-african-indigenous-sheep.pdf.

Paim, T. P., Ianella, P., Paiva, S. R., Caetano, A. R., and Pimentel, C. M. M. (2018). Detection and evaluation of selection signatures in sheep. *Pesq. Agropec. Bras.* 53, 527–539. doi:10.1590/s0100-204x2018000500001

Peters, F. W., Kotze, A., Van der Bank, F. H., Soma, P., and Grobler, J. P. (2010). Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. *Small Rumin. Res.* 90, 101–108. doi:10.1016/j.smallrumres.2010.02.005

Purcell, S. N., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., 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

Qwabe, S. O. (2011). *Genetic and phenotypic characterisation of the South African Namaqua Afrikaner sheep breed*. Doctoral dissertation. University of Pretoria.

Ruiz-Larrañaga, O., Langa, J., Rendo, F., Manzano, C., Iriondo, M., and Estonba, A. (2018). Genomic selection signatures in sheep from the Western Pyrenees. *Genet. Sel. Evol.* 50, 9. doi:10.1186/s12711-018-0378-x

SAS (2013). *Statistical Analysis Systems user's guide (SAS)*. SAS Institute Inc. Cary N.C., US, 9.

Scheet, P., and Stephens, M. A. (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

Selepe, M. M., Ceccobelli, S., Lasagna, E., and Kunene, N. W. (2018). Genetic structure of South African Nguni (Zulu) sheep populations reveals ADMIXTURE with exotic breeds. *PloS one* 13, e0196276. doi:10.1371/journal.pone.0196276

Snyman, M. A., and Fisher, A. D. (2019). Genetic parameters for traits associated with resistance to *Haemonchus contortus* in a South African Dohne Merino sheep flock. *Small Rumin. Res.* 176, 76–88. doi:10.1016/j.smallrumres.2019.01.004

Snyman, M. A. (2014). *South African sheep breeds: Afrino sheep*. Grootfontein Agricultural Development Institute. Info-pack ref. 2014/01. Available at: http://gadi.agric.za/InfoPacks/2014013%20South%20African%20Sheep%20breeds%20-%20Afrino.pdf.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, e171. doi:10.1371/journal.pbio.0050171

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

Turner, S. D. (2014). *qqman: an R package for visualizing GWAS results using QQ and manhattan plots*. Biorxiv, 005165.

Uemoto, Y., Takeda, M., Ogino, A., Kurogi, K., Ogawa, S., Satoh, M., et al. (2020). Genetic and genomic analyses for predicted methane-related traits in Japanese Black steers. *Anim. Sci. J.* 91, e13383. doi:10.1111/asj.13383

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72. doi:10.1371/journal.pbio.0040072

Wang, X. L., Zhou, G. X., Li, Q., Zhao, D. F., and Chen, Y. L. (2014). Discovery of SNPs in *RXFP2* related to horn types in sheep. *Small Rumin. Res.* 116, 133–136. doi:10.1016/j.smallrumres.2013.10.022

Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., et al. (2015). Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16, 194. doi:10.1186/s12864-015-1384-9

Wiedemar, N., and Drögemüller, C. A. (2015). A 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. *Anim. Genet.* 46, 457–461. doi:10.1111/age.12309

Wilson, R. T. (2011). Populations and production of fat-tailed and fat-rumped sheep in the Horn of Africa. *Trop. Anim. Health Prod.* 43, 1419–1425. doi:10.1007/s11250-011-9870-9

Zhang, L., Mousel, M. R., Wu, X., Michal, J. J., Zhou, X., Ding, B., et al. (2013). Genome-wide genetic diversity and differentially selected regions among suffolk, rambouillet, columbia, polypay, and targhee sheep. *PLos One* 8, e65942. doi:10.1371/journal.pone.0065942

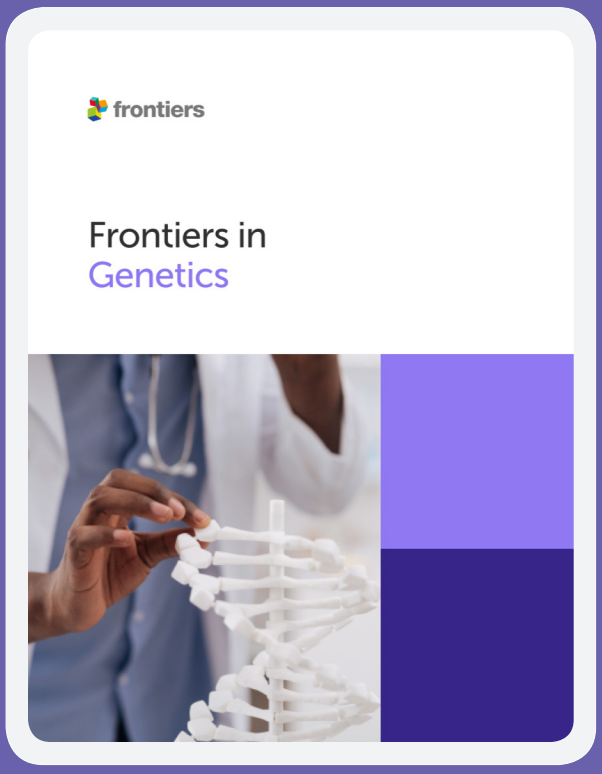
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact