GENETIC DIVERSITY AND SELECTION SIGNATURES IN COMPOSITE BREEDS

EDITED BY: Tiago do Prado Paim, Luiz Brito, El Hamidi Hay and Johann Sölkner

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-099-6
DOI 10.3389/978-2-83250-099-6

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

GENETIC DIVERSITY AND SELECTION SIGNATURES IN COMPOSITE BREEDS

Topic Editors:
Tiago do Prado Paim, Goiano Federal Institute (IFGOIANO), Brazil
Luiz Brito, Purdue University, United States
El Hamidi Hay, United States Department of Agriculture, United States
Johann Sölkner, University of Natural Resources and Life Sciences Vienna, Austria

Citation: Paim, T. d. P., Brito, L., Hay, E. H., Sölkner, J., eds. (2022). Genetic Diversity and Selection Signatures in Composite Breeds. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-099-6

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 17 August 2022
DOI 10.3389/fgene.2022.992609

OPEN ACCESS

EDITED AND REVIEWED BY
Jiuzhou Song,
University of Maryland, College Park, United States

*CORRESPONDENCE
Tiago do Prado Paim,
tiago.paim@ifgoiano.edu.br

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

RECEIVED 12 July 2022
ACCEPTED 18 July 2022
PUBLISHED 17 August 2022

CITATION
Paim TdP, Hay EH and Brito LF (2022), Editorial: Genetic diversity and selection signatures in composite breeds. *Front. Genet.* 13:992609. doi: 10.3389/fgene.2022.992609

Editorial: Genetic diversity and selection signatures in composite breeds

Tiago do Prado Paim[1]*, El Hamidi Hay[2] and Luiz F. Brito[3]

[1]Instituto Federal Goiano, Rio Verde, GO, Brazil, [2]USDA Agricultural Research Service, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT, United States, [3]Department of Animal Sciences, Purdue University, West Lafayette, IN, United States

KEYWORDS

ancestry, hair sheep, genomic breed composition, genetic resources, environmental adaptability, genetic structure

Editorial on the Research Topic
Genetic diversity and selection signatures in composite breeds

Since domestication of livestock species, numerous breeds have been developed through natural and artificial selection for specific traits. In addition to organized crossbreeding schemes (e.g., in poultry and swine breeding industries), composite breeds have also been created through crossing breeds with complementary characteristics for multiple generations. As composite breeds also undergo intensive selection, signatures of selection could be formed in their genome, which could provide important background knowledge on their selection history and genomic structure. However, most previous studies focused on pure breeds. Therefore, this Research Topic was designed to compile original studies investigating the genetic diversity of composite breeds and potential footprints of selection in their genome and demonstrate the usefulness of genomic information to understand population structure, breed formation, and better understand the genetic background of local genetic resources. This Research Topic includes 15 papers, in which three of them investigated composite breeds of beef cattle (Crum et al.; Mulim et al.; Vahedi et al.), eight papers focused on Chinese genetic resources, including cattle (Zhang et al.; Jin et al.; Liu et al.; Luo et al.; Sun et al.), pigs (Xu et al.; Yang et al.), and sheep (Guo et al.). The other papers focused on Korean synthetic pigs (Kim et al.), Copy Number Variation (CNV) in chicken (Chen et al.), and hybrid ass (Dong et al.). Lastly, Qiao et al. reported a new hair sheep reference genome based on the Dorper breed, which originated in South Africa through crossing of Dorset Horn and Blackhead Persian sheep.

All composite cattle breeds studied here were indicine (*Bos taurus indicus*) × taurine (*Bos taurus*) crossbred. Crum et al. identified taurine and indicine haplotype representation in three American Composite cattle breeds (Brangus, Beefmaster, Santa Gertrudis). Lower than expected levels of Brahman contribution were found across the genome of the composite breeds. The average Brahman genome content was 25.81 ± 8.01% (±standard deviation among sampled individuals) for Brangus; 27.60 ± 7.05% for

Santa Gertrudis; and 30.84 ± 7.48% for Beefmaster. These authors found strong evidence that selection for polledness, coat color, growth, calving ease, and intra-muscular fat content produced early-generation cattle with lower than expected indicine proportion in the genomes of all three breeds. Vahedi et al. studied the same three composite breeds using a different dataset and different analytical methods. These authors identified more than 90% of genomic regions underlying selection signatures had European taurine origin. Vahedi et al. explored three haplotype-based methods (iHS, iHH12, and nSL) for selection signatures aiming to identify more recent systematic artificial selection following the breed formation than old selective sweeps. Interestingly, most of the selection signatures and indicine-taurine differentiated genomic regions were breed specific in both studies, suggesting that differences in breeding objectives and selection intensities exist between the composite breeds. The only exception in most of the recent studies with composite cattle breeds is found in chromosome 5, which consistently had a high indicine ancestry (Paim et al., 2020; Crum et al.; Mulim et al.; Vahedi et al.). This warrants further exploration to elucidate the high indicine ancestry in chromosome 5. These papers demonstrated how complementarity and selection jointly contribute to shape the genetic architecture of the Composite breeds population.

Mulim et al. characterized a Brazilian composite beef cattle breed known as Purunã, which was formed by crossing Angus, Charolais, Canchim, and Caracu. Runs of homozygosity (ROH) analyses showed low inbreeding levels with low correlations with pedigree-based measures. These authors identified heterozygote islands harboring genes involved in growth pathways, carcass weight, meat and carcass quality, and marbling deposition. This four-breed composite population had low consistency of gametic phase with the founder breeds, therefore multi-breed genomic evaluation is likely not feasible (Mulim et al.). Composite breeds formed by more than two breeds and with multiple taurine and indicine founders can have a different genetic architecture than previously studied two-breed composite populations such as the Montana Tropical beef cattle (Grigoletto et al., 2020).

There are numerous local breeds in Asia, in which various of them were included in this Research Topic. Most of the Chinese cattle breeds have a complex breeding history and migration, with different combinations of Taurine (European and Asian) and Indicine (Chinese and Indian) origin (Freitas et al., 2021). These papers presented genome sequencing of different Chinese breeds, as Weining (Liu et al.), Dianzhong (Zhang et al.), Lincang Humped Cattle (Sun et al.), Dengchuan (Jin et al.); Xiangxi (Luo et al.). All of them identified genomic selection signatures related to environmental adaptation, such as cold adaptation associated with fat metabolism and blood pressure regulation (Liu et al.), adaptation to hot and humid climate (Zhang et al.), and body size, immunity, and heat tolerance (Sun et al.). These authors also reported missense mutations in the *HELB* gene that were specific to indicine cattle and were presumed to be associated with adaptation to hot environments. These studies provide new insights into the genetic background of Chinese cattle populations, which represent an important reservoir of cattle genetic diversity for future uses, especially considering the emerging challenges imposed by climate change. For instance, the Dengchuan cattle is the main local yellow dairy cattle breed in China with high milk fat percentage and a local specialty dairy product. This is an endangered population among Yunnan native cattle breeds, threatened mainly by crossing with the exotic Holstein breed (Jin et al.). Moreover, these authors showed that Yunnan has been one of the core regions for the migration of Indian indicine into the Chinese territory.

Chinese composite pig breeds (Xidu and Beijing black pig) were also studied as part of this Research Topic. Xu et al. reported genes within ROH islands related to reproduction, fat deposition, ear shape, and environmental adaptation in Xidu black pigs. Population genetic differentiation (F_{ST}) of Beijing Black pigs and other populations ranged from 0.10 to 0.27, which showed that Beijing Black pigs were more genetically similar to the commercial pig breeds than Chinese local pigs, retaining a small amount of Huainan and Min pigs (Yang et al.). This study provides new insights into the historical contribution of Western and Chinese ancestry to actual Beijing Black pigs.

Korean synthetic pig breed (Woori-Heukdon—Korean native pig x Duroc) had more stable genomic breed composition in the first generations. Short ROH reduced while medium and long ROH increased from F1 to WRH, suggesting that more recent inbreeding is happening at a higher rate in WRH (Kim et al.). Therefore, the authors indicated the need for better inbreeding management in these composite breeds.

Guo et al. studied Yunnan semi-fine wool sheep based on whole-genome resequencing data, using Tajima´s D, iHS, and fisher test (comparing groups of one or two lambs per gestation). These authors identified genomic regions associated with environmental adaptation (cold climate, high altitude, and hypoxic conditions) and litter size.

Dong et al. sequenced one Mongolian Kulan and 29 Kulan hybrids. These authors identified the important contribution of the *KITLG* gene to coat color. Mongolian Kulan is an essential part of Asiatic wild ass, but hunting and deteriorating living conditions have caused their numbers to plummet leading them to nearly the level of a threatened species in the International Union for Conservation of Nature Red list (Dong et al.).

Chen et al. studied copy number variation (CNV) in six chicken breeds (four Chinese indigenous breeds and two commercial breeds), which provided an interesting perspective on the evolutionary spectrum of CNVs under artificial selection during chicken domestication and breed formation. Important candidate genes contributing to fast growth, high reproduction, and distinct breed characteristics were identified by Chen et al. This study is a valuable resource to facilitate genetic and functional investigation of domestication and economic traits in chickens.

Qiao et al. provided the first hair sheep reference genome, representing a valuable resource for sheep genetic studies, and provided a pipeline for mining genetic information of composite breeds based on detection of allele-specific expression genes. According to these authors, Dorset sheep had a greater impact in the growth rate, carcass quality, and carcass yield of Dorper sheep than the Persian breed. The Persian breed seems to have contributed more to traits related to fat deposition.

Overall, this Research Topic is a first step towards better characterizing breed formation, genetic architecture, and selection signatures in composite livestock populations. Moreover, many new research insights may arise from the results and discussion presented in the studies included in this Research Topic.

Author contributions

TP drafted the first version of the editorial manuscript. EH and LB edited and approved the final version of the manuscript.

Acknowledgments

The authors acknowledge multiple expert researchers who reviewed the papers included in this Research Topic.

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

References

Freitas, P. H. F., Wang, Y., Yan, P., Oliveira, H. R., Schenkel, F. S., Zhang, Y., et al. (2021). Genetic diversity and signatures of selection for thermal stress in cattle and other two Bos species adapted to divergent climatic conditions. *Front. Genet.* 12, 102. doi:10.3389/fgene.2021.604823

Grigoletto, L., Ferraz, J. B. S., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic architecture of carcass and meat quality traits in Montana Tropical® composite beef cattle. *Front. Genet.* 11, 123. doi:10.3389/fgene.2020.00123

Paim, T., Do, P., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., et al. (2020). Genomic breed composition of selection signatures in brangus beef cattle. *Front. Genet.* 11, 710. doi:10.3389/fgene.2020.00710

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fgene.2021.720081

Genome-Wide Assessment of Runs of Homozygosity and Estimates of Genomic Inbreeding in a Chinese Composite Pig Breed

*Zhong Xu[1], Shuqi Mei[1], Jiawei Zhou[1], Yu Zhang[1], Mu Qiao[1], Hua Sun[1], Zipeng Li[1], Lianghua Li[1], Binke Dong[1], Favour Oluwapelumi Oyelami[2], Junjing Wu[1]*and Xianwen Peng[1]**

[1]Hubei Key Laboratory of Animal Embryo and Molecular Breeding, Institute of Animal Husbandry and Veterinary, Hubei Provincial Academy of Agricultural Sciences, Wuhan, China, [2]Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Tiago do Prado Paim, Goiano Federal Institute (IFGOIANO), Brazil

Reviewed by:
Ran Li, Northwest A and F University, China
Lei Zhou, China Agricultural University, China

***Correspondence:**
Junjing Wu
jeanne1106@126.com
Xianwen Peng
pxwpal@163.com

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

Received: *03 June 2021*
Accepted: *06 August 2021*
Published: *01 September 2021*

Citation:
Xu Z, Mei S, Zhou J, Zhang Y, Qiao M, Sun H, Li Z, Li L, Dong B, Oyelami FO, Wu J and Peng X (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

The primary purpose of the current study was to assess the genetic diversity, runs of homozygosity (ROH) and ROH islands in a Chinese composite pig and explore hotspot regions for traces of selection. First, we estimated the length, number, and frequency of ROH in 262 Xidu black pigs using the Porcine SNP50 BeadChip and compared the estimates of inbreeding coefficients, which were calculated based on ROHs (F_{ROH}) and homozygosity (F_{HOM}). Our result shows that a total of 7,248 ROH exceeding 1 Mb were detected in 262 pigs. In addition, *Sus scrofa* chromosome (SSC) 8 and SSC10, respectively, has the highest and lowest chromosome coverage by ROH. These results suggest that inbreeding estimation based on total ROH may be a useful method, especially for crossbreed or composite populations. We also calculated an inbreeding coefficient of 0.077 from the total ROH. Eight ROH islands were found in this study. These ROH islands harbored genes associated with fat deposition, muscular development, reproduction, ear shape, and adaptation, such as *TRAF7*, *IGFBP7*, *XPO1*, *SLC26A8*, *PPARD*, and *OR1F1*. These findings may help to understand the effects of environmental and artificial selection on the genome structure of composite pigs. Our results provide a basis for subsequent genomic selection (GS), and provides a reference for the hybrid utilization of other pig breeds.

Keywords: pig, crossbreed, runs of homozygosity, inbreeding, linkage disequilibrium

INTRODUCTION

Crossbreeding is a common strategy to improve livestock production because it can explore complementarities of additive genetic effects as well as heterosis caused by non-additive genetic effects (Howard et al., 2016). The Enshi black pig, a typical native black breed in China, mainly lives in mountainous areas of southwest China at an average altitude of more than 800 m. It is well-known for its adaptability to a mountainous environment, excellent meat quality, fat storage ability, and cold-wet tolerance (China National Commission of

Animal Genetic Resources, 2011). Since the 1990s, under the impact of exotic germplasm with high growth rate and lean meat rate, the Enshi black pig has been facing extinction due to its low growth rate. To overcome these deficiencies and conserve the Enshi black pig, crossbreeding programs have been implemented to increase productivity, and the Xidu black pig is a new composite breed that has been developed for this situation. Crossbreeding combines the cold-wet environmental adaptation of the Enshi black pig with the high fertility of the Meishan pig and the fast growth rate of the Hubei white pig. When the three-way crossbreeding [Hubei white × (Meishan × Enshi black)] was formed, they were inter-se mating and selected to become the Xidu black pig breed, which having about 50% Hubei white, 25% Meishan, and 25% Enshi black pig inheritance. It is worth noting that the Xidu black pig is now well established and can be used as a purebred without the need for any ongoing crossbreeding programs.

Generally, crossbred offspring can be mated among themselves in each generation, and selection for specific traits and genetic improvement can be applied during this process. Therefore, it is essential to manage genetic diversity by avoiding high inbreeding rates in composite breed, which will retain high levels of heterozygosity and heterosis (Peripolli et al., 2020). It is difficult to assess genetic diversity using pedigree data mostly because the genealogical relationships between parental breeds that are used in crossbreeding cannot be established. However, the analysis of genomic data could solve this problem (Ganteil et al., 2020). There are several methods used in estimating genetic diversity from genomic data. Some of these methods include the use of observed and expected heterozygosity, runs of homozygosity (ROH), and linkage disequilibrium (LD).

Runs of homozygosity are long continuous homozygous segments in the genome that are formed in an individual by the combination of two identical haplotypes from a common ancestor (Ceballos et al., 2018). As an important genome feature, ROH provides an essential reference for the study of the genome structure. Besides, in animal genetics, the presence of homozygous segments in the genome can be influenced by intensive selection, population history, and consanguinity levels (Peripolli et al., 2017). Also, inbreeding estimates based on ROH is usually considered to be more accurate for estimating individual inbreeding levels when compared with other existing methods (Keller et al., 2011). In addition, ROH hotspots are known to be non-randomly distributed across the genome, and can reveal selection pressure events since selection is one of the main causes of homozygous stretches on the genome. Recently, ROH has been mostly employed in estimating the genomic inbreeding and selection signatures of many livestock populations (Zhang et al., 2018b; Xu et al., 2019; Shi et al., 2020), but less commonly used in crossbred or composite populations.

The LD analysis is also an efficient approach for determining the level of genetic diversity within a studied population. Generally, LD can be defined as the non-random genetic relationship between two loci in a population (Saravanan et al., 2020). Thus, exploring the pattern and extent of LD in the genome can provide essential insights for guiding genome-wide association studies (GWAS) and genome selection (GS; Mokry et al., 2014).

The main objectives of this study were: (1) to investigate the characteristics of ROH on the genome of the Xidu black pig, and also identify the genomic regions with high ROH frequency; and (2) to estimate genetic diversity parameters, such as inbreeding rates and LD in this composite pig population.

MATERIALS AND METHODS

Ethical Statement

All experimental procedures were approved by the Institutional Animal Care and Use Committee of the Hubei Academy of Agriculture Sciences, and all methods involved pigs were in accordance with the agreement of Institutional Animal Care and Use Committee of the Hubei Academy of Agriculture Sciences (Permit number: 36/2016).

Sample Collection, SNP Genotyping, and Quality Control

The animal genomics dataset used in the current study were gotten from pigs raised in a composite swine breeding farm located in Enshi, Hubei, China. In order to make the sample representative, we chose individuals from the core group to avoid full siblings. At last, there were 262 individuals consisting of approximately three generations. The 262 animals were genotyped using the Porcine Single nucleotide polymorphism SNP50 BeadChip (Illumina, United States), consisting of 51,315 Single nucleotide polymorphism (SNPs) evenly distributed along the pig genome. Genotype quality control was carried out using PLINK v1.90 (Chang et al., 2015) software based on the following filtering criteria: (1) the call rate of SNPs and individuals were higher than 0.9; (2) the minor allele frequency (MAF) was greater than 0.01; and (3) only SNPs mapped to autosomes were included. The latest version of the pig genome, *Sus scrofa* 11.1 was used in this study.

Runs of Homozygosity Detection and Classification

We identified ROH in individuals using PLINK v1.90 software, which uses a sliding window approach to detect autozygous segments. The algorithm is as follows: take a window of X SNPs and slide them across the genome. Determine at each window position whether the window looks sufficiently "homogeneous" (yes/no). Then, for each SNP, calculate the proportion of the "homozygous" window that overlaps that position. Call segmentation based on this metric, such as a threshold based on the average value. To define a ROH, the criterion and thresholds were as follows: (1) a minimum ROH length of 1 Mb; (2) at least 50 homozygous SNPs included in a ROH; (3) a minimum density of a SNP in 100 Kb; (4) a sliding window of 50 SNPs across the genome that moves one SNP at a time; and (5) up to one heterozygous SNP and five missing SNPs were allowed in a sliding window. Detected ROHs were later classified into three different classes based

on their length: 1–5, 5–10, and >10 Mb. The total number and length of ROHs were counted for all individuals.

Detection of Common Runs of Homozygosity and Gene Annotation

We identified the genomic regions that were mostly associated with ROHs, by calculating the proportion of SNPs in ROH. This was done by counting the number of times the SNP was detected in those ROH across individuals. Afterward, we selected the top 1% of SNPs that were commonly observed in ROHs. Adjacent SNPs that were above this threshold were finally merged into genomic regions which are called ROH islands, which is characterized by being shared by a majority of individuals in the population (Dixit et al., 2020). We used the database provided by NCBI to annotate the genes in the ROH island. Through a large number of accurate literature searches, the biological function of each annotated gene in the ROH island was inferred.

Estimation of Genomic Inbreeding Coefficient

In this study, two types of genomic inbreeding coefficients were calculated, one based on ROH (F_{ROH}) and the other based on excess of homozygosity (F_{HOM}). Genomic inbreeding coefficients (F_{ROH}) were computed for all individuals by the following formula, as proposed by McQuillan et al. (2008):

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{auto}},$$

where ΣL_{ROH} is the length of ROHs, and L_{auto} is the total length of the genome covered by the SNPs included in this chip. F_{ROH} was also calculated based on three length classes: 1–5, 5–10, >10, and total (>1) Mb. Genomic inbreeding coefficients (F_{HOM}) were calculated as $F_{HOM} = \frac{O-E}{L-E}$, where O is the number of observed homozygous genotypes, E is the number of expected homozygous genotypes by chance, and L is the total number of genotyped autosomal SNPs. Pearson's correlation was used to compare the inbreeding coefficients estimated by these methods using R.

Extent of Linkage Disequilibrium

The LD was measured using the r^2, which was calculated for each pair of SNPs per chromosome according to Hill and Robertson (1968). The pairwise LD (r^2) were calculated using the parameters "--ld-window 99,999 --ld-window-kb 1,000 --ld-window-r2 0" in PLINK v1.90. To visualize the decline of LD, the physical distances between SNPs were divided into 100-Kb intervals, and the average of r^2 in each group was then estimated.

RESULTS AND DISCUSSION

After quality control, 262 pigs and 38,275 SNPs were retained. The average observed (H_o) and expected (H_e) heterozygosity estimates were 0.37 and 0.35, respectively, and the average MAF was 0.26. We observed that the H_o was somewhat higher than the H_e.

Distribution of ROH

In this study, a total of 7,248 ROHs were identified in the 262 animals with an average of 27.66 ROH per animal. The average ROH length was 6.32 Mb, and the longest fragment in *Sus scrofa* chromosome 8 (SSC8) was 57.84 Mb (1,000 SNPs). **Table 1** summarizes the descriptive statistics of the ROH number and length by classes. The total ROH number of the composite Xidu Black pigs was mainly composed of shorter segments (1–5 Mb), which accounted for about 56.16% of all the detected ROH. The genome coverage of long segments (>10 Mb) accounted for 37.11% of the total ROH. Short ROH reflects ancestral inbreeding history, while long ROH segments are usually formed by recent inbreeding. This indicated that both ancient and recent inbreeding events might have affected this population, but recent inbreeding or selection pressures have mainly influenced the genome of the Xidu Black pig population.

The relationship between the total ROH number and the total length of the genome covered by ROH in each individual is shown in **Figure 1**, and is greatly different among animals. In this population, the most extreme animal with long ROHs had a length of 777.06 Mb (34.41% of the pig genome). The variability of the total number and length of ROH among individuals was high. Similar distributions were also observed in other pigs (Xu et al., 2019; Shi et al., 2020) and livestock species, such as sheep (Mastrangelo et al., 2017) and cattle (Peripolli et al., 2018).

For chromosomes, the distribution of total number of ROHs in each chromosome and percentage coverage per chromosome are presented in **Figure 2**. The number of ROHs per chromosome was greater on SSC1 (710 segments), while the smallest number of ROHs was on SSC17 (118 segments). Previous studies on pigs (Xu et al., 2019; Zhan et al., 2020) have also reported the highest number of ROH on SSC1, possibly because SSC1 is the largest chromosome in the pig genome, and has more markers than other chromosomes. The highest ROH coverage was observed on SSC8 (3.26%), whereas the lowest was on SSC10 (1.19%). Our result suggests that the chromosomes with high ROH coverage might have been influenced by positive selection, which consequently increases the accumulation of advantageous alleles on the chromosome. According to our results, some genomic regions, with the highest ROH coverage, on the SSC4 require more consideration.

TABLE 1 | Descriptive statistics of runs of homozygosity (ROH) number and length (in Mb) by ROH length class (ROH 1–5 Mb, ROH 5–10 Mb, ROH > 10 Mb, and total).

ROH length (Mb)	ROH number	Percent (%)	Mean length (Mb)	Genome coverage (%)
1–5	4,072	56.18	3.49	31.02
5–10	2,134	29.44	6.84	31.88
>10	1,042	14.38	16.32	37.11
Total (>1)	7,248	100	6.32	100

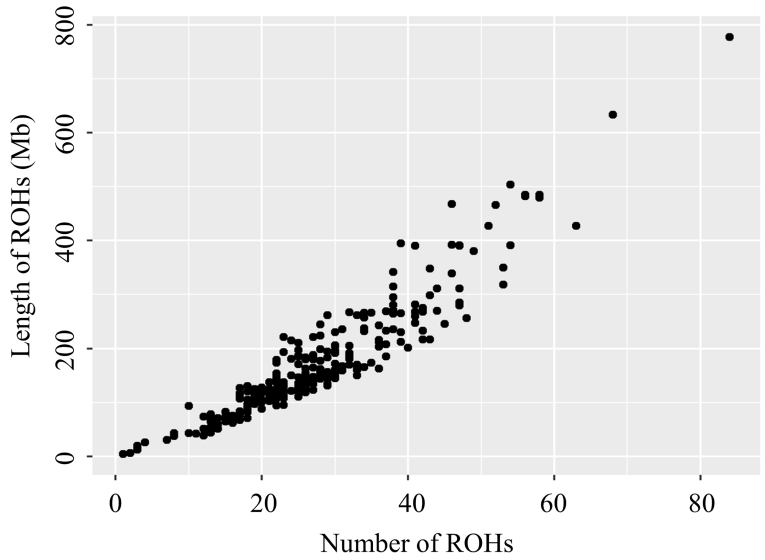

FIGURE 1 | Relationship between the total number of ROH segments (*x*-axis) and the total length (Mb) of the genome in ROH (*y*-axis) for all individuals. Each dot represents an individual.

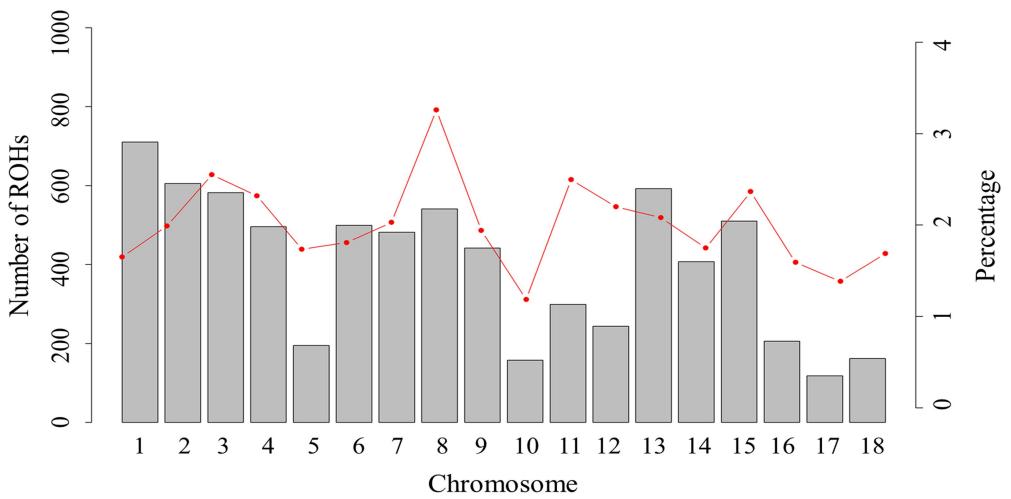

FIGURE 2 | The number of ROHs and percentage coverage per chromosome in Xiduhei pig population. The vertical bars show the total number of ROHs per chromosome and the line shows the percentage of chromosome covered with ROH.

Genomic Inbreeding Coefficients

Here, we calculated the genomic inbreeding coefficients (F_{ROH} and F_{HOM}) using the genotype data of 262 individuals. The mean value of the inbreeding coefficient based on total observed ROHs (F_{ROH_total}) was 0.077 and ranged from 0.002 to 0.344. The estimated F_{HOM} inbreeding coefficients was −0.054 with a range from −0.199 to 0.251 in this population (negative values correspond to individuals with lower-than-average homozygosity). Traditionally, the inbreeding coefficient was estimated based on pedigree data. However, when it comes to cross-bred individuals, things get complicated because the genealogical relationships between parental breeds cannot be established. Moreover, in reality, pedigree information might be incorrect and incomplete, and does not usually take into account the various stochastic events of recombination, which might have occurred during meiosis (Marras et al., 2015). Thus, the inbreeding coefficient value estimated based on pedigree data could not totally show the actual relatedness among individuals within a population. In this study, we used genomic data to estimate inbreeding coefficient.

To further investigate the inbreeding coefficients which were obtained by different estimation methods, we conducted a pairwise comparisons between F_{HOM} and F_{ROH}. The pairwise correlations among five types of inbreeding coefficients were shown in **Figure 3**.

Among all pairwise correlations, the highest correlation was 0.94 between F_{ROH_total} and $F_{ROH>10Mb}$. This result showed that long ROH segments (>10 Mb) were the main source of F_{ROH_total}. The inbreeding coefficients obtained by different categories of ROHs with F_{HOM} ranged from 0.63 to 0.83, with the highest correlation found between F_{ROH_total} and $F_{HOM.}$ These results are in line with previous research in other pig populations (Xu et al., 2019; Shi et al., 2020) and cattle (Mastrangelo et al., 2016; Biscarini et al., 2020). Furthermore, we found that many individuals had a negative F_{HOM} value, which might be because F_{HOM} was sensitive to allele frequency for populations with a higher level of heterozygosity compared to F_{ROH} estimators (Zhang et al., 2015). Zhang et al. (2015) found a negative F_{HOM} value for Danish Red Cattle (RDC), a composite breed, which is likely due to the admixture present in RDC. A similar result was obtained for a crossbred cattle in Vrindavani (Chhotaray et al., 2021). Therefore, the results of this study suggest that the inbreeding level based on total ROH may be a useful method, especially for crossbreed or composed populations.

ROH Islands

Furthermore, we plotted the percentage of SNPs in ROHs against their respective positions along the chromosomes (**Figure 4**). The result shows a non-uniformity in the frequency of different SNPs within the ROH regions across the genome. The most frequent SNP in ROH (121 occurrences, 46.18%) was mapped at ∼35 Mb on SSC11, and the closest gene to this SNP was the *U6* gene. Regions of the genome with high homozygosity around the ROH islands may contain positively selected targets and might be under strong selection pressure (Pemberton et al., 2012). To identify the genomic regions that were mostly associated with ROH in all individuals, we considered the top 1% of SNPs with the highest occurrences (over 30.15% of the samples) in the ROH as candidate SNPs (**Figure 4**). We identified a total of eight ROH island regions (**Table 2**), with length ranging from 0.804 Mb on SSC2 to 3.188 Mb on SSC11.

Chromosome position, the start and end position of ROH, ROH length, number of SNPs, and the number of genes within the ROH islands were reported in **Table 2**. To evaluate the potential functional importance of the detected ROH islands, we analyzed the gene content of the identified regions. In summary, we annotated a total of 199 genes that were detected within these ROH islands. The chromosome position, start and end, gene name and Ensembl Gene ID were provided in **Supplementary Table S1**.

Candidate Genes Within Runs of Homozygosity Islands

In this study, we focus on the genes related to some specific livestock-traits which are also important in breeding. We identified numerous candidate genes associated with muscular development and fat deposition. Among these genes is the *TRAF7* gene, a MyoD1 transcriptional target that can regulate NF-κB activity during myogenesis. Studies have shown that missense mutations in *TRAF7* causes developmental delay or skeletal dysplasia (Tsikitis et al., 2010; Tokita et al., 2018). *IGFBP7* can promote lipid accumulation and triglyceride production in mature adipocytes and plays an important regulatory role in the differentiation of preadipocyte cells that can affect fat deposition (Hu et al., 2021). *PRSS33* was related to lipid transport and metabolism, and was also detected in reported selection signature regions in Enshi black pigs, one of the founder breeds. Since Xidu black pigs are characterized for meat quality, we also detected some important genes that are associated with specific meat quality traits: *TRAP1*

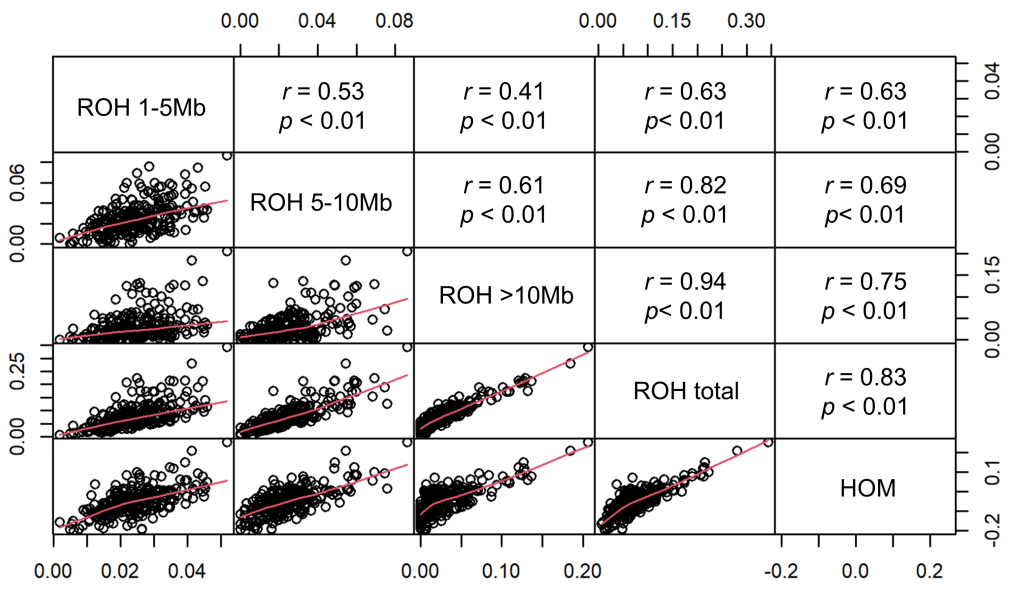

FIGURE 3 | Scatterplots (lower panel) and Pearson's correlations (upper panel) of the genomic inbreeding coefficients F_{ROH} ($F_{ROH\ 1\sim5\ Mb}$, $F_{ROH5\sim10\ Mb}$, $F_{ROH>10\ Mb}$, and F_{ROH_total}) and $F_{HOM.}$

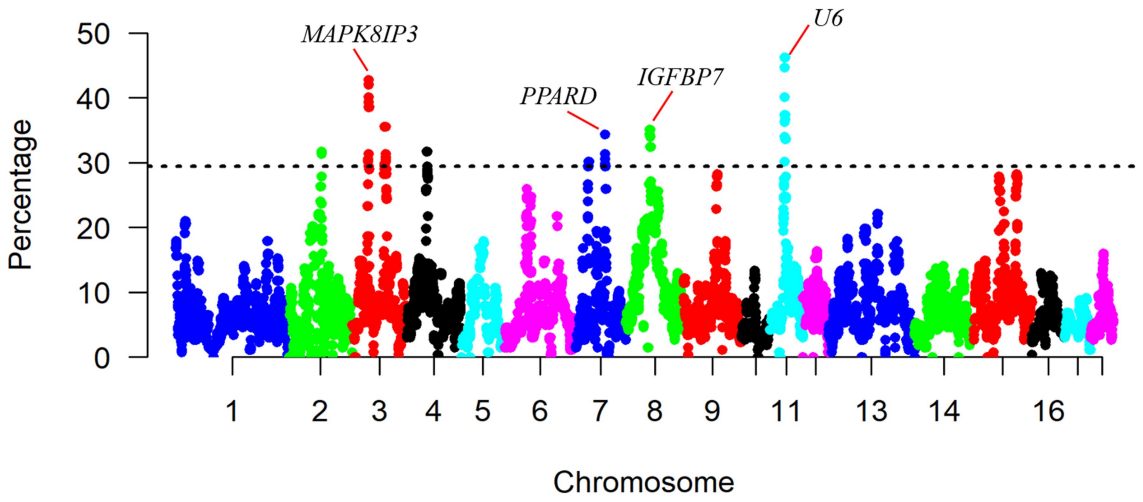

FIGURE 4 | Manhattan plot of occurrences (%) of a SNP in ROHs across individuals.

TABLE 2 | List of genomic regions of extended homozygosity detected in Xidu black pigs.

CHR	Start (bp)	End (bp)	Length (bp)	SNPs	Genes
2	77,541,872	78,345,614	803,743	19	22
3	38,431,596	41,463,322	3,031,727	82	123
3	78,653,188	81,780,864	3,127,677	62	18
4	47,685,504	48,592,012	906,509	26	2
7	31,217,540	32,755,132	1,537,593	30	26
7	70,752,052	71,626,349	874,298	20	1
8	56,064,681	58,992,457	2,927,777	77	3
11	34,189,314	37,377,736	3,188,423	66	2

and *CREBBP* are associated with pork meat pH in Finnish Yorkshire pigs (Verardo et al., 2017); *SLC9A3R2* gene has been shown to be differentially expressed in longissimus muscle tissues of Meishan and Large White pigs (Jiugang et al., 2011), and was considered to be the candidate genes for meat quality (Wu et al., 2020). *OTX1* is a novel regulator of proliferation, migration, invasion, and apoptosis in lung adenocarcinoma (Yang et al., 2020), and were involved in the battle between foot-and-mouth disease virus and the host (Zhang et al., 2018a). We identified several candidate genes related to reproduction traits: *XPO1*, a nuclear transport receptor, plays an essential role in meiotic resumption in porcine full-grown and growing oocytes (Onuma et al., 2018); *SLC26A8*, also known as testis anion transporter 1, is required for sperm terminal differentiation and male fertility in the mouse (Toure et al., 2007).

OR1F1, an olfactory receptor, was demonstrated to function in odor perception activation (Li et al., 2013). *CLDN9* played an essential role in maintaining barrier function in airway epithelial cells (Gon et al., 2017), and *E4F1* is essential for skin homeostasis (Lacroix et al., 2010). Notably, the Xidu black pigs reside in a subtropical region, which is characterized by high temperature and humidity. Therefore, we considered that *OR1F1*, *CLDN9*, and *E4F1* as key factors for the environmental adaptability of Xidu black pigs. The most interesting candidate gene in this population seems to be *PPARD* gene, which was shown to affect the shape of the external ear and fat deposition in pigs (Meidtner et al., 2009; Ren et al., 2011). Due to the hybrid parents (Hubei white pig, Meishan pig, and Enshi black pig) have different ear shapes, the ear shapes of the base population had erected, forward sloping and drooping phenotype. Therefore, during the breeding process of Xidu black pigs, the forward sloping and slightly drooping ear shape was constantly selected. This gene may play a key role in ear shape in this population and needs more attention.

Linkage Disequilibrium

A total of 38,275 SNPs were used to calculate the average LD between all adjacent SNPs, with a distance less than 100 Kb. The average and SD of estimated r^2 was 0.289 ± 0.316, which was similar to the result that (Joaquim et al., 2019) estimated in a crossbred Landrace pig population. It was observed that the LD value decreased with the increase of the distance between markers (**Figure 5**). When the distance was greater than 1,000 Kb, the average r^2 was about 0.15. The LD extent ($r^2 = 0.3$) in Xidu black pigs was about 25 Kb. This value was less than 79.54 Kb ($r^2 = 0.3$) for Meishan pigs (one of the original parental breed) obtained by the same method (Liu et al., 2020). The effectiveness of GWAS and genomic selection (GS) relies on the LD between the markers. According to the literature, a mean r^2 value above 0.30 is considered as a strong LD sufficient for QTL mapping (Shifman et al., 2003). However, an average r^2 of 0.20 is considered enough to achieve an accuracy of 0.85 for genomic estimated breeding value (GEBV; Meuwissen et al., 2001). In this study, we found that moderate LD ($r^2 \geq 0.2$) extended up to 150–200 Kb. Assuming that the total length of the pig genome is 2.5 Gb, this suggests that at least 12,500–16,667 SNPs would be required for effective GWAS in this breed. This result could be particularly useful when designing breed-specific SNP array panels in future genomic study and selection programs for this composite pig breed.

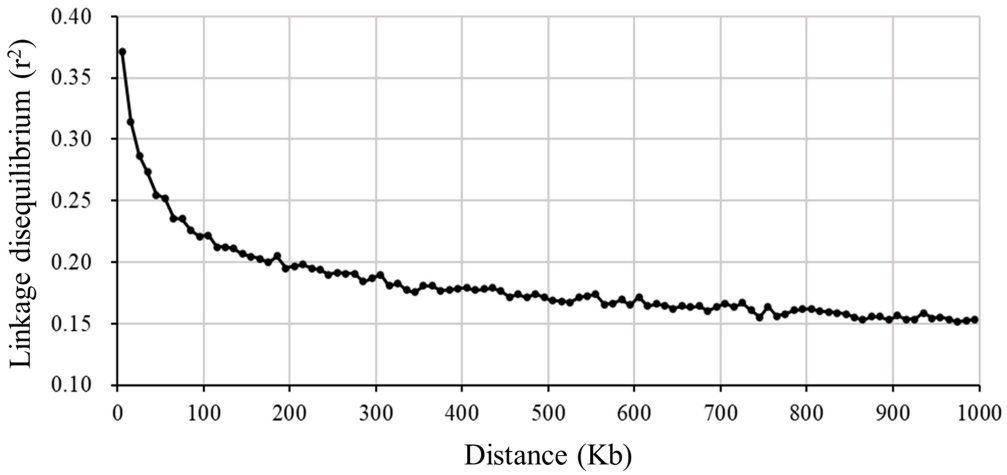

FIGURE 5 | Decay of average linkage disequilibrium (r^2) over distance (Kb) between markers for a composite pig breed.

CONCLUSION

In summary, in this study, we investigated the patterns of ROH, inbreeding coefficients and LD in the Xidu black pigs. To our knowledge, this is the first study about ROH patterns and autozygosity islands in a composite pig breed. The results of this study suggest that inbreeding based on total ROH may be a useful method, especially for crossbred or composite populations. The detected ROH patterns in this population suggests recent inbreeding events, agreeing with the newest developments in this composite pig breed. Besides, the reported genes within the identified ROH islands point to phenotypic characteristics related to reproduction, fat deposition, ear shape, and environmental adaptation. We believe that these findings will further assist in genome-wide association studies, GS, as well as the design and implementation of breed improvement and conservation programs.

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 at: https://figshare.com/articles/dataset/Genome-wide_assessment_of_runs_of_homozygosity_and_estimates_of_genomic_inbreeding_in_a_Chinese_composite_pig_breed/14904567.

ETHICS STATEMENT

All experimental procedures were approved by the Institutional Animal Care and Use Committee of the Hubei Academy of Agriculture Sciences, and all methods involved pigs were in accordance with the agreement of Institutional Animal Care and Use Committee of the Hubei Academy of Agriculture Sciences (Permit number: 36/2016).

AUTHOR CONTRIBUTIONS

SM and XP: data curation and supervision. SM: funding acquisition. JW: project administration. ZX: writing – original draft. SM, XP, and FO: writing – review and editing. JZ, YZ, MQ, HS, ZL, LL, and BD: data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the Hubei Natural Science Foundation (2020BBB069), China Agriculture Research System of MOF and MARA, the Innovation Team of the Hubei Agricultural Science and Technology Innovation Center (2019-620-000-001-18), and Hubei Academy of Agricultural Sciences Youth Science Fund (2020NKYJJ10).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.720081/full#supplementary-material

REFERENCES

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., and Ciampolini, R. (2020). Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. *Animals* 10:2285. doi: 10.3390/ani10122285

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, 220–234. doi: 10.1038/nrg.2017.109

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Chhotaray, S., Panigrahi, M., Pal, D., Ahmad, S. F., Bhanuprakash, V., Kumar, H., et al. (2021). Genome-wide estimation of inbreeding coefficient, effective population size and haplotype blocks in Vrindavani crossbred cattle strain of India. *Biol. Rhythm. Res.* 52, 666–679. doi: 10.1080/09291016.2019.1600266

China National Commission of Animal Genetic Resources (2011). *Animal Genetic Resources in China Pigs*. Beijing: Chinese Agriculture Press.

Dixit, S. P., Singh, S., Ganguly, I., Bhatia, A. K., Sharma, A., Kumar, N. A., et al. (2020). Genome-wide runs of homozygosity revealed selection signatures in bos indicus. *Front. Genet.* 11:92. doi: 10.3389/fgene.2020.00092

Ganteil, A., Rodriguez-Ramilo, S. T., Ligonesche, B., and Larzul, C. (2020). Characterization of autozygosity in pigs in three-way crossbreeding. *Front. Genet.* 11:584556. doi: 10.3389/fgene.2020.584556

Gon, Y., Maruoka, S., Kishi, H., Kozu, Y., Kazumichi, K., Nomura, Y., et al. (2017). NDRG1 is important to maintain the integrity of airway epithelial barrier through claudin-9 expression. *Cell Biol. Int.* 41, 716–725. doi: 10.1002/cbin.10741

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622

Howard, J. T., Tiezzi, F., Huang, Y., Gray, K. A., and Maltecca, C. (2016). Characterization and management of long runs of homozygosity in parental nucleus lines and their associated crossbred progeny. *Genet. Sel. Evol.* 48:91. doi: 10.1186/s12711-016-0269-y

Hu, Z., Wu, J., Qin, L., Jin, H., Cao, Y., and Zhao, Y. (2021). IGFBP7 downregulation or overexpression effect on bovine preadipocyte differentiation. *Anim. Biotechnol.* 32, 21–30. doi: 10.1080/10495398.2019.1642906

Jiugang, Z., Jing, L., and Yonggang, L. (2011). A novel porcine gene--SLC9A3R2, differentially expressed in the longissimus muscle tissues from meishan and large white pigs. *Folia Biol.* 57, 20–25.

Joaquim, L. B., Chud, T. C. S., Marchesi, J. A. P., Savegnago, R. P., Buzanskas, M. E., Zanella, R., et al. (2019). Genomic structure of a crossbred landrace pig population. *PLoS One* 14:e0212266. doi: 10.1371/journal.pone.0212266

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

Lacroix, M., Caramel, J., Goguet-Rubio, P., Linares, L. K., Estrach, S., Hatchi, E., et al. (2010). Transcription factor E4F1 is essential for epidermal stem cell maintenance and skin homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21076–21081. doi: 10.1073/pnas.1010167107

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi: 10.1038/ng.2811

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

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., et al. (2015). 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

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.* 49:84. doi: 10.1186/s12711-017-0360-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, 359–372. doi: 10.1016/j.ajhg.2008.08.007

Meidtner, K., Schwarzenbacher, H., Scharfe, M., Severitt, S., Blocker, H., and Fries, R. (2009). Haplotypes of the porcine peroxisome proliferator-activated receptor delta gene are associated with backfat thickness. *BMC Genet.* 10:76. doi: 10.1186/1471-2156-10-76

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Mokry, F. B., Buzanskas, M. E., de Alvarenga Mudadu, M., do Amaral Grossi, D., Higa, R. H., Ventura, R. V., et al. (2014). Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. *BMC Genomics* 15(Suppl. 7):S6. doi: 10.1186/1471-2164-15-S7-S6

Onuma, A., Fujioka, Y. A., Fujii, W., Sugiura, K., and Naito, K. (2018). Effects of exportin 1 on nuclear transport and meiotic resumption in porcine full-grown and growing oocytes. *Biol. Reprod.* 98, 501–509. doi: 10.1093/biolre/iox168

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., Munari, D. P., Silva, M., Lima, A. L. F., Irgang, R., and Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* 48, 255–271. doi: 10.1111/age.12526

Peripolli, E., Stafuzza, N. B., Amorim, S. T., de Lemos, M. V. A., Grigoletto, L., Kluska, S., et al. (2020). Genome-wide scan for runs of homozygosity in the composite Montana Tropical((R)) beef cattle. *J. Anim. Breed. Genet.* 137, 155–165. doi: 10.1111/jbg.12428

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

Ren, J., Duan, Y., Qiao, R., Yao, F., Zhang, Z., Yang, B., et al. (2011). A missense mutation in PPARD causes a major QTL effect on ear size in pigs. *PLoS Genet.* 7:e1002043. doi: 10.1371/journal.pgen.1002043

Saravanan, K. A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G. K., et al. (2020). Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in Tharparkar cattle breed of India. *Anim. Biotechnol.* 30, 1–15. doi: 10.1080/10495398.2020.1796696

Shi, L., Wang, L., Liu, J., Deng, T., Yan, H., Zhang, L., et al. (2020). Estimation of inbreeding and identification of regions under heavy selection based on runs of homozygosity in a large white pig population. *J. Anim. Sci. Biotechnol.* 11:46. doi: 10.1186/s40104-020-00447-0

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., and Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* 12, 771–776. doi: 10.1093/hmg/ddg088

Tokita, M. J., Chen, C. A., Chitayat, D., Macnamara, E., Rosenfeld, J. A., Hanchard, N., et al. (2018). De novo missense variants in TRAF7 cause developmental delay, congenital anomalies, and dysmorphic features. *Am. J. Hum. Genet.* 103, 154–162. doi: 10.1016/j.ajhg.2018.06.005

Toure, A., Lhuillier, P., Gossen, J. A., Kuil, C. W., Lhote, D., Jegou, B., et al. (2007). The testis anion transporter 1 (Slc26a8) is required for sperm terminal differentiation and male fertility in the mouse. *Hum. Mol. Genet.* 16, 1783–1793. doi: 10.1093/hmg/ddm117

Tsikitis, M., Acosta-Alvear, D., Blais, A., Campos, E. I., Lane, W. S., Sanchez, I., et al. (2010). Traf7, a MyoD1 transcriptional target, regulates nuclear factor-kappaB activity during myogenesis. *EMBO Rep.* 11, 969–976. doi: 10.1038/embor.2010.154

Verardo, L. L., Sevon-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.* 18:13. doi: 10.1186/s12863-017-0482-x

Wu, F., Sun, H., Lu, S., Gou, X., Yan, D., Xu, Z., et al. (2020). Genetic diversity and selection signatures within diannan small-ear pigs revealed by next-generation sequencing. *Front. Genet.* 11:733. doi: 10.3389/fgene.2020.00733

Xu, Z., Sun, H., Zhang, Z., Zhao, Q., Olasege, B. S., Li, Q., et al. (2019). Assessment of autozygosity derived from runs of homozygosity in Jinhua pigs disclosed by sequencing data. *Front. Genet.* 10:274. doi: 10.3389/fgene.2019.00274

Yang, C. Y., Wang, L., Mu, D. C., Li, F. F., Ran, P. Z., Shen, H., et al. (2020). OTX1 is a novel regulator of proliferation, migration, invasion and apoptosis

in lung adenocarcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 24, 9497–9510. doi: 10.26355/eurrev_202009_23035

Zhan, H., Zhang, S., Zhang, K., Peng, X., Xie, S., Li, X., et al. (2020). Genome-wide patterns of homozygosity and relevant characterizations on the population structure in Pietrain pigs. *Genes (Basel)* 11:577. doi: 10.3390/genes11050577

Zhang, Q., Calus, M. P., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16:88. doi: 10.1186/s12863-015-0227-7

Zhang, T., Chen, H., Qi, L., Zhang, J., Wu, R., Zhang, Y., et al. (2018a). Transcript profiling identifies early response genes against FMDV infection in PK-15 cells. *Viruses* 10:364. doi: 10.3390/v10070364

Zhang, Z., Zhang, Q., Xiao, Q., Sun, H., Gao, H., Yang, Y., et al. (2018b). Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data. *Anim. Genet.* 49, 579–591. doi: 10.1111/age.12730

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fgene.2021.719215

Assessing Genomic Diversity and Productivity Signatures in Dianzhong Cattle by Whole-Genome Scanning

Xianfu Zhang[1], Kaixing Qu[2], Peng Jia[3], Jicai Zhang[2], Jianyong Liu[2], Chuzhao Lei[3] and Bizhi Huang[2]**

[1]Key Laboratory of Applied Technology on Green-Eco-Healthy Animal Husbandry of Zhejiang Province, Zhejiang Provincial Engineering Laboratory for Animal Health Inspection and Internet Technology, College of Animal Science and Technology, College of Veterinary Medicine, Zhejiang A&F University, Hangzhou, China, [2]Yunnan Academy of Grassland and Animal Science, Kunming, China, [3]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Xianyang, China

Dianzhong cattle is a classic Chinese indigenous cattle breed with historical records dating back to 200 BC. But with its genomic differences having not been clearly elucidated, the quest for genomic characterization will be an essential step towards understanding the genomic basis of productivity and adaptation to survival under Chinese farming systems. Here we compared 10 Dianzhong cattle (four newly sequenced and six downloaded) with 29 published genomes of three underlying ancestral populations (Chinese zebu, Indian zebu, and Yanbian cattle) to characterize the genomic variations of Dianzhong cattle. Dianzhong cattle has a high nucleotide diversity (0.0034), second only to Chinese zebu. Together with analyses of linkage disequilibrium decay and runs of homozygosity, Dianzhong cattle displayed higher genomic diversity and weaker artificial selection compared with Yanbian cattle. From a selective sweep analysis by four methods (*F*st, π-ratio, XP-CLR, and XP-EHH), the positive selective signals were mainly manifested in candidate genes and pathways related to heat resistance, growth and development, fat deposition, and male reproduction. Missense mutations were detected in candidate genes, *SDS* (c.944C > A and p.Ala315Glu), *PDGFD* (c.473A > G and p.Lys158Arg), and *DDX4* (rs460251486, rs722912933, and rs517668236), which related to heat resistance, fat deposition, and spermatogenesis, respectively. Our findings unravel, at the genome-wide level, the unique diversity of Dianzhong cattle while emphasizing the opportunities for improvement of livestock productivity in further breeding programs.

Keywords: Dianzhong cattle, hybrid, genetic diversity, selection signatures, DDX4

OPEN ACCESS

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

Reviewed by:
Gábor Mészáros, University of Natural Resources and Life Sciences Vienna, Austria
Nina Moravčíková, Slovak University of Agriculture, Slovakia

****Correspondence:***
Xianfu Zhang
20060159@zafu.edu.cn
Bizhi Huang
hbz@ynbp.cn

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

Received: *02 June 2021*
Accepted: *10 September 2021*
Published: *05 October 2021*

Citation:
Zhang X, Qu K, Jia P, Zhang J, Liu J, Lei C and Huang B (2021) Assessing Genomic Diversity and Productivity Signatures in Dianzhong Cattle by Whole-Genome Scanning. Front. Genet. 12:719215. doi: 10.3389/fgene.2021.719215

INTRODUCTION

Domestic cattle generally refers to two subspecies, *Bos taurus* and *Bos taurus indicus*. They were domesticated in the Fertile Crescent (~10,000 years ago) and the Indus Valley (~8,000 years ago), respectively (Utsunomiya et al., 2019). These two subspecies can be interbred without barriers, unlike other bovine subspecies (buffalo, American bison, etc.) that have reproductive isolation (Wu et al., 2018). According to previous genomic study, the domestic cattle in the world could be divided into six major groups: European taurine, Eurasian taurine, East Asian taurine, African taurine, Indian indicine, and Chinese indicine (Chen et al., 2018). Besides that,

as one of the important routes for the migration of Indian indicine into China, there are many hybrid cattle with different lineages in the Yunnan region of China.

Dianzhong cattle is one of the most widely distributed indigenous breeds in Yunnan. It has a long history of breeding, with historical records dating back to 200 BC (China National Commission of Animal Genetic Resources, 2011). In addition to agricultural use, it also has social importance, including during marriage, birth, death, and sacrificial ceremonies, as well as being regarded as representatives of wealth, prestige, and status (China National Commission of Animal Genetic Resources, 2011). Due to the hot and humid climate of the original area, Dianzhong cattle display superior heat tolerance and resistance to parasites. Although small in size, its labor performance is excellent, owing to its long-term use as the main farming and transportation livestock (Wen et al., 2014). Moreover, as a classic indigenous cattle breed, it also shows advantageous characteristics of high intramuscular fat, strong disease resistance, and crude feed tolerance (Hao et al., 2017). In recent years, on account of its slow growth and low reproductive performance, which cannot meet the growing needs of beef, local people have blindly and massively introduced commercial cattle for hybrid improvement (Hao et al., 2017). However, blind hybridization as well as the lack of breed conservation planning caused the threat of breed degradation in Dianzhong cattle.

Based on whole-genome sequencing, many studies initially focused on the genetic architecture and economic traits under positive selection in commercial breeds (Bovine et al., 2009; Stothard et al., 2011), and then the focus gradually shifted to the adaptation of indigenous breeds, such as climate tolerance and disease resistance (Kim et al., 2017; Taye et al., 2017; Kim et al., 2020). With the development of next-generation sequencing technology and the enrichment of re-sequencing databases, genome-wide genetic analysis studies play an increasingly powerful role in the investigation of germplasm resources of landraces, such as cold tolerance of Yanbian cattle (Shen et al., 2020), excellent meat quality of Mongolian cattle (Chen et al., 2021), heat tolerance and parasite resistance of African cattle (Kim et al., 2017), and growth rate and feed conversion of Jiaxian red cattle (Xia et al., 2021). An earlier study on Dianzhong cattle using Illumina BovineHD BeadChip (777K) demonstrated that Dianzhong cattle contained mainly indicine ancestry mixed with taurine ancestry. Based on the estimate of observed heterozygosity and expected heterozygosity, Dianzhong cattle had the maximum level of inbreeding coefficients in Yunnan region (Li et al., 2019). In fact, single-nucleotide polymorphism (SNP) array data with only limited and known SNPs might make many important genetic information hard to detect. In order to further explore the genetic potential of Dianzhong cattle, we compared the genome re-sequencing data of 10 Dianzhong cattle (including four that were newly sequenced) with 29 reference cattle from Yanbian, Jiangxi, and India to search for the candidate signatures of positive selection by four methods (*F*st, π-ratio, XP-CLR, and XP-EHH).

MATERIALS AND METHODS

Sample Collection and Sequencing

We sampled four Dianzhong cattle from Chuxiong, Yunnan, China. Genomic DNA was extracted from the ear tissue samples as previously described (Jia et al., 2019). The pair-end libraries were constructed for each individual (500 bp insert size), and the DNA was subjected to Illumina NovaSeq sequencing using 2 × 150 bp model at Novogene Bioinformatics Institute (Beijing, China). Additionally, genomes of 35 publicly available representative groups, including Dianzhong cattle (n = 6), Chinese zebu (Jiangxi cattle, n = 10), Indian zebu (Brahman, Gir, and Nelore, n = 10), and Asian taurine (Yanbian cattle, n = 9), were used for the combined analysis (**Supplementary Table S1**). To obtain clean reads with high quality, the raw data was modified using Trimmomatic (LEADING:20 TRAILING:20 SLIDINGWINDOW:3:15 AVGQUAL:20 MINLEN:35 TOPHRED33) (Bolger et al., 2014).

Alignments and Variant Identification

Clean reads were aligned against the latest *B. taurus* reference genome (ARS-UCD1.2) using Burrows-Wheeler Aligner BWA-MEM (v0.7.13-r1126) with default parameters (Li et al., 2009). The Picard tool (http://broadinstitute.github.io/picard) was used to filter potential duplicate reads (REMOVE_DUPLICATES = true). SNP calling was performed by Genome Analysis Toolkit 3.8 (GATK) (Nekrutenko et al., 2012; Haplotype Caller, Genotype GVCFs and Select Variants module). After SNP calling, we used the module "Variant Filtration" of GATK to obtain high-quality SNPs with the parameters ("DP < 156 (1/3-fold total sequence depth for all individuals) || DP > 1404 (3-fold of total sequence depth for all individuals) || QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0 || SOR > 3.0"). Finally, we used ANNOVAR (Wang et al., 2010) to annotate the functions of the SNPs based on the *B. taurus* reference assembly ARS-UCD1.2.

Population Genetic Diversity

Nucleotide diversity for each group was investigated by VCFtools (Danecek et al., 2011) with a 50 kb non-overlapping window across all autosomes. The genetic relationship among four groups was performed by principal component analysis (PCA). Autosomal SNPs in high levels of pair-wise linkage disequilibrium (LD) were pruned using PLINK (Purcell et al., 2007) with the parameter (--indep-pair-wise 50 5 0.2) and were conducted using the smartPCA program in the EIGENSOFT v5.0 package (Patterson et al., 2006).

The LD decay for each group was measured using PopLDdecay (Zhang et al., 2019) with default parameters. Runs of homozygosity (ROHs) were identified using the Runs of Homozygosity program implemented in PLINK, which slides a window of 50 SNPs (-homozyg-window-snp 50) across the genome in estimating homozygosity (Xia et al., 2021). The following settings were performed for ROH identification: -homozyg-density 50 -homozyg-window-het 1 -homozyg-window-missing 2. The number and length of ROH for each group were estimated, and the length of ROH was divided into

three categories: 0.5–1, 1–2, and 2–4 Mb reflecting ancient, historical, and recent inbreeding, respectively (Bhati et al., 2020). The strong linkage disequilibrium between parental genomic loci formed ROH, with long haplotype segments derived from recent ancestors and short haplotype segments derived from earlier ancestors (Jane et al., 2006).

Selective Sweep Identification

To identify selective sweep regions, genetic differentiation (*F*st), genetic diversity (π-ratio), cross-population composite likelihood ratio test (XP-CLR), and cross-population extended haplotype homozygosity test (XP-EHH) were performed with 50 kb sliding window and 20 kb step between Dianzhong cattle and each other reference group separately. Selection scanning was performed using these four methods based on allele frequency, linkage imbalance, and population differentiation. The overlap of the top 1% windows in each method was considered as candidate signatures of selection. Genomic regions identified by at least two methods were considered to be candidate regions of positive selection (Sun et al., 2020). In addition, π and *F*st were computed with 5 kb sliding window and 2 kb step by using VCFtools for each candidate gene. To gain a better understanding of the gene functions and signaling pathways of the identified candidate genes, online Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analyses were conducted using DAVID 6.8 (Huang et al., 2009). The analyses for LD and construction of the haplotypes were performed with the online SHEsis software (http://analysis.bio-x.cn/myAnalysis.php) (Yong et al., 2005).

RESULTS

Genome Resequencing and SNP Identification

Individual genomes of four Dianzhong cattle were generated to an average of ×~9.8 coverage each and aligned to the *B. taurus* reference genome ARS-UCD1.2 with an average alignment rate of 99.72%. To reveal the diversity of Dianzhong cattle, 10 Dianzhong cattle (four new and six downloaded) were jointly genotyped with 29 publicly available genomes from three representative groups (10 Indian zebu, 10 Chinese zebu, and nine Yanbian cattle) (**Supplementary Table S1**). In total, 49,094,970 bi-allelic autosomal SNPs were detected, and the average alignment rate and sequencing depth of the final set reached 99.50% and ×~12, respectively. After functional annotation of the polymorphic sites, the largest number of SNPs in the exon region was Dianzhong cattle (616,110). That presented the richer genetic diversity of Dianzhong cattle (**Supplementary Table S2**). Meanwhile, the analyses of unique SNPs and non-synonymous SNPs displayed that the genetic diversity of Dianzhong cattle was second only to Chinese zebu (**Supplementary Figure S1**).

Population Genetic Diversity and Relationships

On a genome-wide window scale of 50 kb, Yanbian cattle (East Asian taurine) showed a reduced level of nucleotide diversity compared to other groups (**Figure 1A**). The nucleotide diversity of Dianzhong cattle (0.0034) was second only to Chinese zebu (0.0039) and close to Indian zebu (0.0030). To explore relatedness among Dianzhong cattle and other cattle groups, PCA was conducted through using autosomal SNPs. The analysis ignored breed membership; nevertheless, it revealed clear breed structures as samples from the same group cluster together. PC1 and PC2 accordingly explained 9.12 and 4.78% of the total variations and separated taurine from indicine and Chinese zebu from Indian zebu, respectively (**Figure 1B**).

Demographic inferences in cattle population were made based on the analyses of ROH and LD decay. For LD patterns, at short distances, Indian zebu showed lower LD level, and the highest LD level was found in Yanbian cattle, followed by Chinese zebu and Dianzhong cattle. Yanbian cattle continued to have a higher LD than all other breeds when the distances were larger, while a contrary trend was observed in indicine (Indian zebu and Chinese zebu) and hybrid (Dianzhong) (**Figure 1C**). To evaluate the ROH pattern of Dianzhong cattle and other cattle groups, we divided the length of ROH into three size classes: 0.5–1, 1–2, and 2–4 Mb (**Figure 1D**). The vast majority of ROH identified in all groups were between 0.5 and 1 Mb in length, but apparently Yanbian cattle had more medium (1–2 Mb) and long (2–4 Mb) ROHs, showing stronger inbreeding. As expected, the pattern of ROH profile was nearly consistent with the result of LD decay.

Candidate Regions and Genes Under Positive Selection

Due to the genetic separation between Dianzhong cattle and three reference groups, we compared Dianzhong cattle with the other three reference populations using four selective sweep methods respectively in three groups.

In the Dianzhong cattle vs. Yanbian cattle comparison, the top 1% of candidate genes in four selection methods were extracted at the intersection (**Supplementary Figure S2**): *MARS2* (Webb et al., 2015) and *ARL6IP6* (Abumansour et al., 2015) (growth and development), *RFTN2* (Wei et al., 2018) (DNA damage response), and *N6AMT1* (Wang et al., 2020) (feed efficiency). GO and KEGG analyses were performed on candidate genes scanned twice or more by four methods. From DAVID gene ontology, 12 significant ($p < 0.05$) GO biological process (BP) terms were enriched (**Supplementary Table S16**). Among them, "oxidation–reduction process, GO:0055114" ($n = 18$) contained more genes than other GO terms. The other candidate genes were mainly enriched in hot resistance ("cardiac muscle cell differentiation, GO:0055007," "lipid catabolic process, GO: 0016042," "positive regulation of ATPase activity, GO: 0032781," "mitochondrial DNA repair, GO:0043504," and "regulation of oxidative stress-induced intrinsic apoptotic signaling pathway, GO:1902175"). Moreover, we noticed a region of about 0.06 Mb scanned by XP-CLR on chromosome 17 (including TRNAG-CCC, *SDSL*, *SDS*, and *PLBD2*), showing a strong positive selection signal. Thereinto, a missense mutation (c.944C > A, p.Ala315Glu) was found at the *SDS* gene. This mutation presented a huge divergence between Dianzhong cattle (allele C frequency = 1) and Yanbian cattle (allele A frequency = 1).

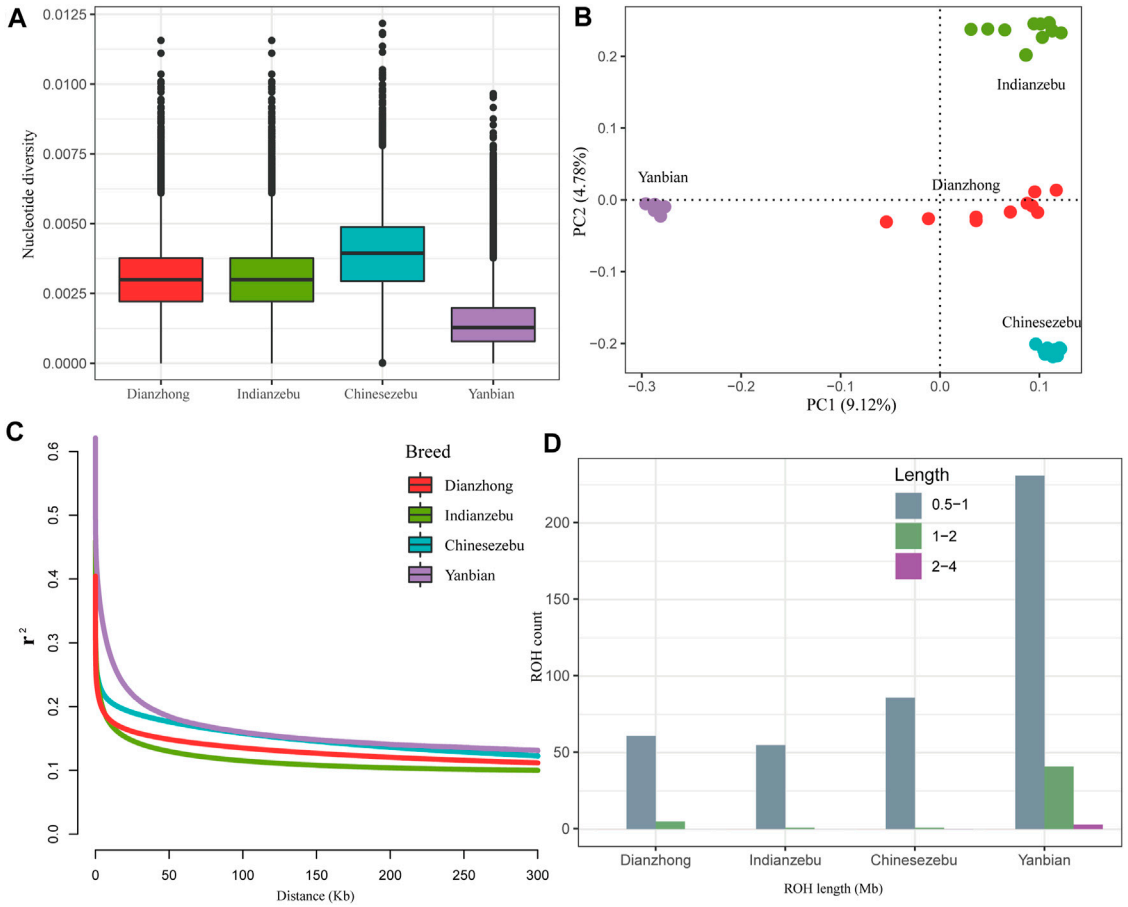

FIGURE 1 | Summary statistics for genomic diversity. **(A)** Genome-wide distribution of nucleotide diversity of each group. The horizontal line inside the box indicated the median of this distribution; box limits indicated the first and the third quartiles, points showed outliers. Data points outside the whiskers can be considered as outliers. **(B)** Principal component analysis of four cattle groups. **(C)** Genome-wide average LD decay estimated from each group. **(D)** The estimation of number and length of ROH for each group.

A total of 42 candidate genes were overlapped in four selection methods by comparing Dianzhong cattle with Indian zebu (**Supplementary Figure S2**), and three of them were significantly enriched in "kinase binding, GO: 0019900" (p-value = 0.02). In the functional prediction of all candidate genes, eight GO BP terms were significantly enriched ($p < 0.05$) (**Supplementary Table S17**). The candidate genes were mainly enriched in growth ("post-embryonic development, GO:0009791," "skeletal muscle tissue development, GO:0007519"; and "embryonic skeletal system morphogenesis, GO:0048704") and heat stress resistance ("DNA recombination, GO:0006310," "inflammatory response, GO:0006954," "response to hydrogen peroxide, GO:0042542," and "respiratory gaseous exchange, GO:0007585"). Most genes ($n = 11$) were included in "inflammatory response, GO:0006954." Furthermore, we detected a missense mutation (c.473A > G, p.Lys158Arg) at the fat deposition-related gene *PDGFD*. Allele A displayed a rare distribution (frequency 0.2) in Dianzhong cattle, whereas it showed an opposite pattern (frequency 0.9) in Indian zebu.

In the selection signal analysis between Dianzhong cattle and Chinese zebu, 36 common candidate genes were scanned by four selection methods. There were two genes enriched in the "response to retinoic acid, GO:0032526" (p-value = 0.04). The strongest signal was found in *KIT* gene, which was associated with sperm differentiation [33]. In KEGG analysis, four digestion-related pathways were enriched (**Supplementary Table S18**): "gastric acid secretion, bta04971," "pancreatic secretion, bta04972," "salivary secretion, bta04970," and "carbohydrate digestion and absorption, bta04973." In addition, three missense mutations (rs460251486, rs722912933, and rs517668236) with strong linkage effect were detected at candidate gene *DDX4*. According to the analyses for LD and haplotype construction of the three mutations, they exhibited a low recombination rate with r^2 values ranging from 0.31 to 0.61 (**Supplementary Figure S3**) (Ardlie et al., 2002). *DDX4* gene was the candidate gene selected by *F*st (**Figure 2A**), and through the calculation of *F*st and π with a smaller window (5 kb), a significant differentiation was observed between Dianzhong cattle and Chinese zebu (**Figure 2B**). Simultaneously, according to the

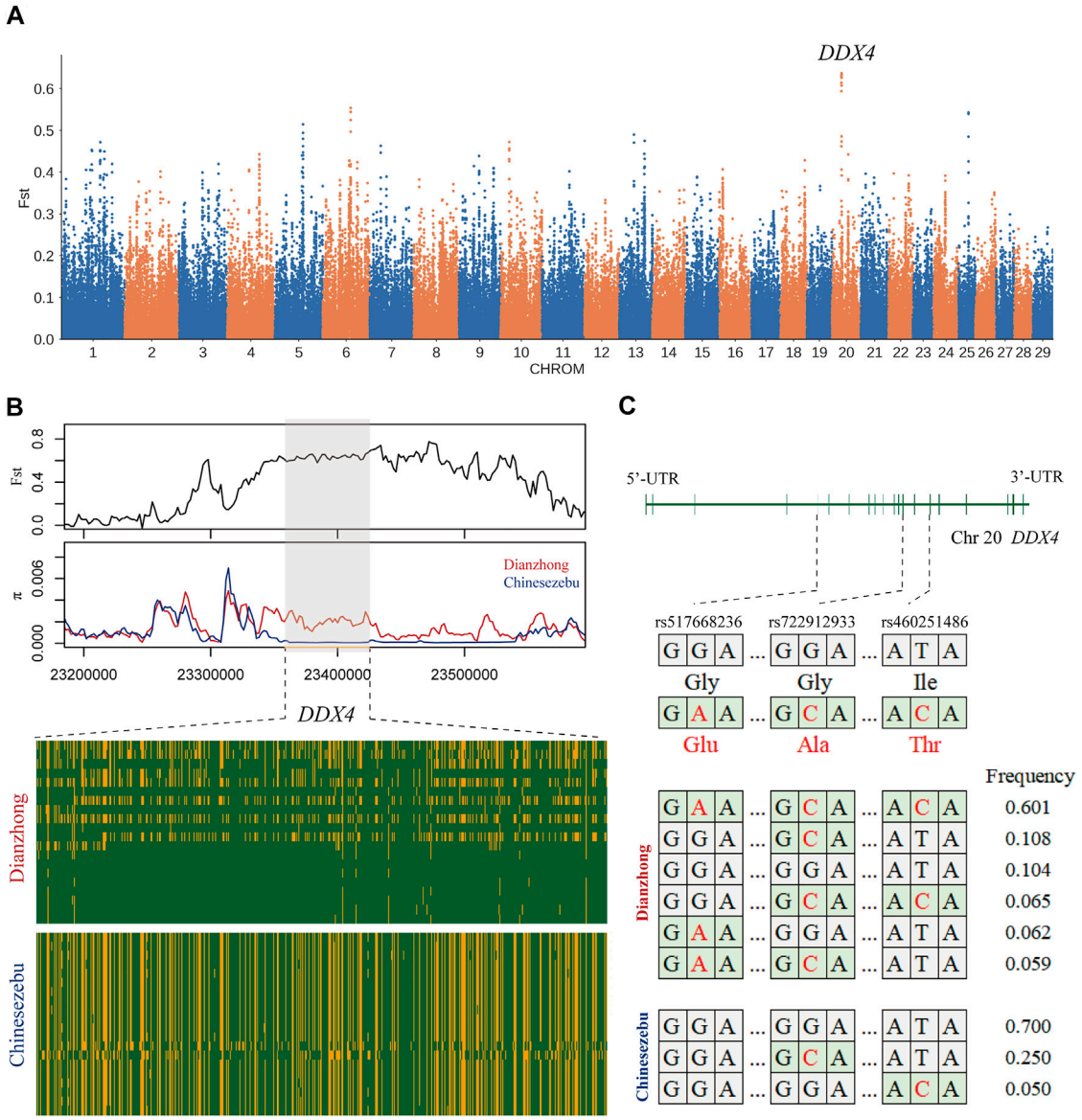

FIGURE 2 | Analysis of the signatures of positive selection between Dianzhong cattle and Chinese zebu. **(A)** Manhattan plot of selective sweeps. **(B)** Fst and π plots and haplotype patterns heatmap of the DDX4 gene region. **(C)** Structure of DDX4 with exons indicated by vertical bars. Three missense mutations and their haplotypes were highlighted in red.

calculation of the combined haplotype frequencies at the three loci, the haplotype with the most distribution in Dianzhong cattle was the homozygous mutational haplotype (frequency 0.601), while the homozygous wild haplotype (frequency 0.700) was the commonest one in Chinese zebu (**Figure 2C**).

DISCUSSION

Genomic features may reflect a variety of multifarious historical events, including climate change, species introduction, and artificial selection (Chen et al., 2021). The characteristics of population genetic diversity are essential for assessing the genetic potential of breeds as well as for the utilization and protection of cattle breed resources (Xia et al., 2021). In this study, three underlying ancestors of Dianzhong cattle were selected as reference groups to explore the genetic information of Dianzhong cattle. The nucleotide diversity ranking calculated in this study (Chinese zebu > Dianzhong cattle > Indian zebu > Yanbian cattle) was basically consistent with the earlier results (Chen et al., 2018). Dianzhong cattle, belonging to hybrid cattle, obtained more abundant genetic information from multiple genetic sources, while Chinese zebu had high nucleotide diversity which may be due to previous reception of introgression from Banteng (*Bos javanicus*)

(Chen et al., 2018). The monotonous environment and high degree of breeding may be the reasons why the nucleotide diversity of Yanbian cattle is lower than that of indicine and hybrid cattle (Singh et al., 2021; Xia et al., 2021). In addition, more 2–4 MB and 1–2 MB ROHs and more rapid LD decayed in Yanbian cattle confirmed this once again. The nucleotide diversity in our Indian zebu was higher than those in other groups, and the LD level was lower than others, suggesting that Indian zebu had a higher effective population size (Chen et al., 2021). In summary, the LD decay pattern and ROH distribution of each group were roughly consistent with the results of nucleotide diversity.

Owing to the humid and hot living environment, Dianzhong cattle evolved to be prominently heat resistant (China National Commission of Animal Genetic Resources, 2011). In the comparative analysis with Yanbian cattle living in a cold environment, functional enrichment analysis revealed that heat resistance-related pathways were significantly overrepresented in candidate genes under positive selection. Moreover, a putatively selected gene, *SDS*, encodes serine dehydratase, an enzyme that catalyzes the degradation of serine to pyruvate, which was functionally involved in heat resistance (Ogawa et al., 1989). *SDS* activated inflammasomes by mediating mitochondrial membrane potential, thereby affecting individual immune function (Çağdaş et al., 2020). Inflammasomes are important in the defense against microbial infections and have roles in shaping the adaptive immune responses (Joly et al., 2012; Yu et al., 2015).

The typical characteristic of Dianzhong cattle is small in body size, approximately 101 cm tall and weighing 170 kg in adulthood (Wen et al., 2014). Compared with the tall Indian zebu by selection analysis, there were many growth- and development-related items in GO enrichment. What is more, because of the preference of East Asians for good meat quality (high marbling score), Dianzhong cattle have high potential for accumulating intramuscular fat and producing highly marbled beef (Hao et al., 2017). The *PDGFD* gene selected by XP-CLR was related to perivascular adipose tissue generation (Dong et al., 2020). Members of the PDGF family have an anti-fat effect and inhibit the differentiation of preadipocytes (Ma et al., 2018; Pan et al., 2019). *PDGFD* played a pivotal role in inhibiting the differentiation of white adipocytes by regulating the expression of PPARγ2 and C/EBPα (Li et al., 2020). Early studies have reported that the *PDGFD* gene was hypothesized to be one of the candidate genes leading to fat tail formation in indigenous sheep (Zhao et al., 2020; Zhu et al., 2021). In addition to the possible involvement of intermuscular fat deposition, we ventured to speculate that the *PDGFD* gene may also contribute to the formation of indicine hump. However, this is just a conjecture, and more theoretical and experimental supports were required.

Poor fertility is one of the most important factors restricting the development of Dianzhong cattle (China National Commission of Animal Genetic Resources, 2011). The process of admixture among the *Bos* subspecies brought adaptability but caused a cost of reduced reproductive fitness due to genomic incompatibility at the same time (Kim et al., 2020). Jiangxi cattle (Chinese zebu) is a precocious puberty cattle breed that can breed for the first time at the age of 10 months (Shuanping et al., 2019), while Dianzhong bulls do so later than 30 months old (China National Commission of Animal Genetic Resources, 2011). In comparison between Dianzhong cattle and Chinese zebu, some candidate genes were significantly enriched in "response to retinoic acid, GO:0032526" ($p = 0.04$). Retinoic acid promoted the differentiation of male germ cells and induced the differentiation of mouse pluripotent cell bodies into outer male germ cells (Mahabadi et al., 2020). In addition, candidate gene *KIT*, as the highest-ranked in overlap, is a transmembrane protein receptor related to germ cell maturation (Rossi et al., 2000). It is also a sign of the loss of efficacy of spermatogonial stem cells (Schrans-Stassen et al., 1999). In general, the presence of multiple linkage loci associated with a specific phenotype in a gene indicated a highly probable connection between the gene and this phenotype (Jia et al., 2019; Cao et al., 2020). Hence, the haplotypes formed by the three strongly linked missense mutations in the *DDX4* gene were obviously different between Dianzhong cattle and Chinese zebu, implying that the spermatogenesis-related *DDX4* gene may be related to the weak reproductive performance of Dianzhong cattle. *DDX4* gene encodes an ATP-dependent RNA helicase (Hay et al., 1988). In *DDX4*-knockout mice, germ cell development was normal in female homozygous null mice, but male mice were sterile due to the failure of their germ cells to progress from leptotene to zygotene of meiotic prophase I, and the cells underwent apoptosis (Tanaka et al., 2000). The expression of *DDX4* in germ cells of male mice has been described in humans, dogs, cattle, pigs, and stallions (Lee et al., 2018a). Interestingly, *KIT* gene, together with *DDX4* gene, was considered to be a marker for undifferentiated spermatogonia before puberty and differentiated spermatogonia after puberty in porcine testis (Lee et al., 2018b).

In conclusion, our population genomic analyses of Dianzhong cattle and other three reference groups provide novel insights into their genetic diversity and selective sweep. This will point out the direction for genetic assessment and development of reasonable improvement of Dianzhong cattle. Moreover, we identified a series of candidate genes that may be important for the heat resistance, intermuscular fat deposition, and reproductive barriers of this breed. These results provide a basis for further research on the genome characteristics of other important indigenous beef cattle in the future (Gao et al., 2017).

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 Administration of Affairs Concerning Experimental Animals of China (Protocol number, WAFAC1008).

AUTHOR CONTRIBUTIONS

XZ and PJ contributed equally towards the construction and execution of this manuscript. KQ helped in sample collection. JZ and JL revised the manuscript and provided valuable suggestion. CL and BH contributed in the funding for the research.

FUNDING

This work was supported by the China Agriculture Research System of MOF and MARA (CARS-37), the Program of Yunling Scholar, the Young and Middle-Aged Academic Technology Leader Backup Talent Cultivation Program in Yunnan Province, China (no. 2018HB045), and the Yunnan Provincial Major S&T Project (no. 2019ZG007 and 2019ZG011).

ACKNOWLEDGMENTS

Thanks to the High-Performance Computing of Northwest A&F University for providing computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.719215/full#supplementary-material

REFERENCES

Abumansour, I. S., Hijazi, H., Alazmi, A., Alzahrani, F., Bashiri, F. A., Hassan, H., et al. (2015). ARL6IP6, a Susceptibility Locus for Ischemic Stroke, Is Mutated in a Patient with Syndromic Cutis Marmorata Telangiectatica Congenita. *Hum. Genet.* 134 (8), 815–822. doi:10.1007/s00439-015-1561-6

Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of Linkage Disequilibrium in the Human Genome. *Nat. Rev. Genet.* 3 (4), 299–309. doi:10.1038/nrg777

Bhati, M., Kadri, N. K., Crysnanto, D., and Pausch, H. (2020). Assessing Genomic Diversity and Signatures of Selection in Original Braunvieh Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 21 (1), 27. doi:10.1186/s12864-020-6446-y

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics (Oxford, England)* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Bovine, H. C., Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324 (5926), 528–532. doi:10.1126/science.1167936

Cao, Y., Jia, P., Wu, Z., Huang, M., Chen, S., Zhang, J., et al. (2020). A Novel SNP of MYO1A Gene Associated with Heat-Tolerance in Chinese Cattle. *Anim. Biotechnol.* 4, 1–6. doi:10.1080/10495398.2020.1837147

Çağdaş, D., Sürücü, N., Tan, Ç., Kayaoğlu, B., Özgül, R. K., Akkaya-Ulum, Y. Z., et al. (2020). Autoinflammation in Addition to Combined Immunodeficiency: SLC29A3 Gene Defect. *Mol. Immunol.* 121, 28–37. doi:10.1016/j.molimm.2020.02.014

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome Resequencing Reveals World-wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nat. Commun.* 9 (1), 2337. doi:10.1038/s41467-018-04737-0

Chen, Q., Shen, J., Hanif, Q., Chen, N., Huang, Y., Dang, R., et al. (2021). Whole Genome Analyses Revealed Genomic Difference between European Taurine and East Asian Taurine. *J. Anim. Breed. Genet.* 138 (1), 56–68. doi:10.1111/jbg.12501

China National Commission of Animal Genetic Resources (2011). *Animal Genetic Resources in China Bovines*. Beijing: Chinese Agricultural Press.

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

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. *BMC Genomics* 21 (1), 800. doi:10.1186/s12864-020-07210-9

Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., et al. (2017). Species Composition and Environmental Adaptation of Indigenous Chinese Cattle. *Sci. Rep.* 7 (1), 16196. doi:10.1038/s41598-017-16438-7

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended Tracts of Homozygosity in Outbred Human Populations. *Hum. Mol. Genet.* 15 (5), 789–795. doi:10.1093/hmg/ddi493

Hao, W., Zhixian, X., Peichang, Y., Chunfeng, L., Xiangsheng, S., Rongfa, L., et al. (2017). Protection, Exploitation and Utilization of Genetic Resources in Dianzhong Cattle. *Curr. Anim. husbandry* 403 (35), 6–7.

Hay, B., Jan, L. Y., and Jan, Y. N. (1988). A Protein Component of Drosophila Polar Granules Is Encoded by Vasa and Has Extensive Sequence Similarity to ATP-dependent Helicases. *Cell* 55 (4), 577–587. doi:10.1016/0092-8674(88)90216-4

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Jia, P., Cai, C., Qu, K., Chen, N., Jia, Y., Hanif, Q., et al. (2019). Four Novel SNPs of MYO1A Gene Associated with Heat-Tolerance in Chinese Cattle. *Animals* 9 (11), 964. doi:10.3390/ani9110964

Joly, S., Eisenbarth, S. C., Olivier, A. K., Williams, A., Kaplan, D. H., Cassel, S. L., et al. (2012).Cutting Edge: Nlrp10 Is Essential for Protective Antifungal Adaptive Immunity against Candida Albicans, *J Immunol*, 10, 189. , 2012 Baltimore, Md. *1950*, 4713–7. doi:10.4049/jimmunol.1201715

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The Genome Landscape of Indigenous African Cattle. *Genome Biol.* 18 (1), 34. doi:10.1186/s13059-017-1153-y

Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., et al. (2020). The Mosaic Genome of Indigenous African Cattle as a Unique Genetic Resource for African Pastoralism. *Nat. Genet.* 52 (10), 1099–1110. doi:10.1038/s41588-020-0694-2

Lee, R., Lee, W.-Y., Park, H.-J., Ha, W.-T., Woo, J.-S., Chung, H.-J., et al. (2018a). Stage-specific Expression of DDX4 and C-Kit at Different Developmental Stages of the Porcine Testis. *Anim. Reprod. Sci.* 190, 18–26. doi:10.1016/j.anireprosci.2017.12.020

Lee, R., Lee, W.-Y., Park, H.-J., Ha, W.-T., Woo, J.-S., Chung, H.-J., Lee, J.-H., Hong, K., and Song, H. (2018b). Stage-specific expression of DDX4 and c-kit at different developmental stages of the porcine testis. *Animal Reproduction Science* 190, 18–26. doi:10.1016/j.anireprosci.2017.12.020

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, R., Li, C., Chen, H., Liu, X., Xiao, H., and Chen, S. (2019). Genomic Diversity and Admixture Patterns Among Six Chinese Indigenous Cattle Breeds in Yunnan. *Asian-australas J. Anim. Sci.* 32 (8), 1069–1076. doi:10.5713/ajas.18.0605

Li, X., Yang, J., Shen, M., Xie, X.-L., Liu, G.-J., Xu, Y.-X., et al. (2020). Whole-genome Resequencing of Wild and Domestic Sheep Identifies Genes Associated with Morphological and Agronomic Traits. *Nat. Commun.* 11 (1), 2815. doi:10.1038/s41467-020-16485-1

Ma, L., Li, Z., Cai, Y., Xu, H., Yang, R., and Lan, X. (2018). Genetic Variants in Fat- and Short-Tailed Sheep from High-Throughput RNA-Sequencing Data. *Anim. Genet.* 49 (5), 483–487. doi:10.1111/age.12699

Mahabadi, J. A., Tameh, A. A., Talaei, S. A., Karimian, M., Rahiminia, T., Enderami, S. E., et al. (2020). Retinoic Acid And/or Progesterone Differentiate Mouse Induced Pluripotent Stem Cells into Male Germ Cells *In Vitro*. *J. Cel. Biochem.* 121 (3), 2159–2169. doi:10.1002/jcb.29439

Nekrutenko, A., and Taylor, J. (2012). Next-generation Sequencing Data Interpretation: Enhancing Reproducibility and Accessibility. *Nat. Rev. Genet.* 13 (9), 667–672. doi:10.1038/nrg3305

Ogawa, H., Gomi, T., Konishi, K., Date, T., Nakashima, H., Nose, K., et al. (1989). Human Liver Serine Dehydratase. *J. Biol. Chem.* 264 (27), 15818–15823. doi:10.1016/S0021-9258(18)71550-0

Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., et al. (2019). Rapid Evolution of a Retro-Transposable Hotspot of Ovine Genome Underlies the Alteration of BMP2 Expression and Development of Fat Tails. *BMC Genomics* 20 (1), 261. doi:10.1186/s12864-019-5620-6

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *Plos Genet.* 2, e190–e2093. doi:10.1371/journal.pgen.0020190

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 (3), 559–575. doi:10.1086/519795

Rossi, P., Sette, C., Dolci, S., and Geremia, R. (2000). Role of C-Kit in Mammalian Spermatogenesis. *J. Endocrinol. Invest.* 23 (9), 609–615. doi:10.1007/BF03343784

Schrans-Stassen, B. H. G. J., van de Kant, H. J. G., de Rooij, D. G., and van Pelt, A. M. M. (1999). Differential Expression of C-Kit in Mouse Undifferentiated and Differentiating Type a Spermatogonia. *Endocrinology* 140 (12), 5894–5900. doi:10.1210/endo.140.12.7172

Shen, J., Hanif, Q., Cao, Y., Yu, Y., Lei, C., Zhang, G., et al. (2020). Whole Genome Scan and Selection Signatures for Climate Adaption in Yanbian Cattle. *Front. Genet.* 11, 94. doi:10.3389/fgene.2020.00094

Shuanping, Z., Hai, J., Lei, X., Juan, W., Mo, L., and Yutang, J. (2019). Analysis on the Characteristics of Wannan Cattle Germplasm Resources and Hybridization Improvement Effect. *Chin. Cattle Sci.* 45 (05), 21–23.

Singh, A. K., Liu, W., Zakari, S., Wu, J., Yang, B., Jiang, X. J., et al. (2021). A Global Review of Rubber Plantations: Impacts on Ecosystem Functions, Mitigations, Future Directions, and Policies for Sustainable Cultivation. *Sci. Total Environ.* 796, 148948. doi:10.1016/j.scitotenv.2021.148948

Stothard, P., Choi, J.-W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., et al. (2011). Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery. *BMC Genomics* 12, 559. doi:10.1186/1471-2164-12-559

Sun, T., Huang, G.-y., Wang, Z.-h., Teng, S.-h., Cao, Y.-h., Sun, J.-l., et al. (2020). Selection Signatures of Fuzhong Buffalo Based on Whole-Genome Sequences. *BMC Genomics* 21 (1), 674. doi:10.1186/s12864-020-07095-8

Tanaka, S. S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., et al. (2000). The Mouse Homolog of Drosophila Vasa Is Required for the Development of Male Germ Cells. *Genes Dev.* 14 (7), 841–853. doi:10.1101/gad.14.7.841

Taye, M., Kim, J., Yoon, S. H., Lee, W., Hanotte, O., Dessie, T., et al. (2017). Whole Genome Scan Reveals the Genetic Signature of African Ankole Cattle Breed and Potential for Higher Quality Beef. *BMC Genet.* 18 (1), 11. doi:10.1186/s12863-016-0467-1

Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., et al. (2019). Genomic Clues of the Evolutionary History of *Bos indicus* Cattle. *Anim. Genet.* 50 (6), 557–568. doi:10.1111/age.12836

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Wang, X., and Kadarmideen, H. N. (2020). Metabolite Genome-wide Association Study (mGWAS) and Gene-Metabolite Interaction Network Analysis Reveal Potential Biomarkers for Feed Efficiency in Pigs. *Metabolites* 10 (5), 201. doi:10.3390/metabo10050201

Webb, B. D., Wheeler, P. G., Hagen, J. J., Cohen, N., Linderman, M. D., Diaz, G. A., et al. (2015). Novel, Compound Heterozygous, Single-Nucleotide Variants inMARS2Associated with Developmental Delay, Poor Growth, and Sensorineural Hearing Loss. *Hum. Mutat.* 36 (6), 587–592. doi:10.1002/humu.22781

Wei, F., Hao, P., Zhang, X., Hu, H., Jiang, D., Yin, A., et al. (2018). Etoposide-induced DNA Damage Affects Multiple Cellular Pathways in Addition to DNA Damage Response. *Oncotarget* 9 (35), 24122–24139. doi:10.18632/oncotarget.24517

Wen, L., Deliang, C., Xiaofang, O., Jincheng, Z., Xiangguang, M., and Jiacai, Z. (2014). Determination and Analysis of Growth Performance of Hybrid Progeny of Dianzhong Cattle. *Shanghai Anim. husbandry Vet. Newsl.* 01, 32–35.

Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y., Tokarska, M., et al. (2018). Pervasive Introgression Facilitated Domestication and Adaptation in the Bos Species Complex. *Nat. Ecol. Evol.* 2 (7), 1139–1145. doi:10.1038/s41559-018-0562-y

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22 (1), 43. doi:10.1186/s12864-020-07340-0

Yong, Y., and He, L. (2005). SHEsis, a Powerful Software Platform for Analyses of Linkage Disequilibrium, Haplotype Construction, and Genetic Association at Polymorphism Loci. *Cell Res* 15 (2), 97–98. doi:10.1038/sj.cr.7290272

Yu, S.-X., Du, C.-T., Chen, W., Lei, Q.-Q., Li, N., Qi, S., et al. (2015). Genipin Inhibits NLRP3 and NLRC4 Inflammasome Activation via Autophagy Suppression. *Sci. Rep.* 5, 17935. doi:10.1038/srep17935

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhao, F., Deng, T., Shi, L., Wang, W., Zhang, Q., Du, L., et al. (2020). Genomic Scan for Selection Signature Reveals Fat Deposition in Chinese Indigenous Sheep with Extreme Tail Types. *Animals* 10 (5), 773. doi:10.3390/ani10050773

Zhu, C., Li, N., Cheng, H., and Ma, Y. (2021). Genome Wide Association Study for the Identification of Genes Associated with Tail Fat Deposition in Chinese Sheep Breeds. *Biol. Open* 10 (5). doi:10.1242/bio.054932

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fgene.2021.758394

Taurine and Indicine Haplotype Representation in Advanced Generation Individuals From Three American Breeds

Tamar E. Crum[1*†], *Robert D. Schnabel*[1,2], *Jared E. Decker*[1,2] *and Jeremy F. Taylor*[1*]

[1]*Division of Animal Sciences, University of Missouri, Columbia, MO, United States,* [2]*Informatics Institute, University of Missouri, Columbia, MO, United States*

OPEN ACCESS

Edited by:
Tiago do Prado Paim, Goiano Federal Institute (IFGOIANO), Brazil

Reviewed by:
Marina Rufino Salinas Fortes, University of Queensland, Australia
Ron Lewis, University of Nebraska-Lincoln, United States

***Correspondence:**
Tamar E. Crum
Tamar.Crum@acuityswine.com
Jeremy F. Taylor
taylorjerr@missouri.edu

†Present Address:
Acuity Genetics, 7475 State Route 127, Carlyle, IL, United States

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

Received: *13 August 2021*
Accepted: *27 September 2021*
Published: *18 October 2021*

Citation:
Crum TE, Schnabel RD, Decker JE and Taylor JF (2021) Taurine and Indicine Haplotype Representation in Advanced Generation Individuals From Three American Breeds. Front. Genet. 12:758394. doi: 10.3389/fgene.2021.758394

Development of the American Breeds of beef cattle began in the 1920s as breeders and U. S. Experiment Station researchers began to create *Bos taurus taurus* × *Bos taurus indicus* hybrids using Brahman as the *B. t. indicus* source. By 1954, U.S. Breed Associations had been formed for Brangus (5/8 Angus × 3/8 Brahman), Beefmaster (½ Brahman × ¼ Shorthorn × ¼ Hereford), and Santa Gertrudis (5/8 Shorthorn × 3/8 Brahman). While these breeds were developed using mating designs expected to create base generation animals with the required genome contributions from progenitor breeds, each association has now registered advanced generation animals in which selection or drift may have caused the realized genome compositions to differ from initial expected proportions. The availability of high-density SNP genotypes for 9,161 Brangus, 3,762 Beefmaster, and 1,942 Santa Gertrudis animals allowed us to compare the realized genomic architectures of breed members to the base generation expectations. We used RFMix to estimate local ancestry and identify genomic regions in which the proportion of Brahman ancestry differed significantly from *a priori* expectations. For all three breeds, lower than expected levels of Brahman composition were found genome-wide, particularly in early-generation animals where we demonstrate that selection on beef production traits was likely responsible for the taurine enrichment. Using a proxy for generation number, we also contrasted the genomes of early- and advanced-generation animals and found that the indicine composition of the genome has increased with generation number likely due to selection on adaptive traits. Many of the most-highly differentiated genomic regions were breed specific, suggesting that differences in breeding objectives and selection intensities exist between the breeds. Global ancestry estimation is commonly performed in admixed animals to control for stratification in association studies. However, local ancestry estimation provides the opportunity to investigate the evolution of specific chromosomal segments and estimate haplotype effects on trait variation in admixed individuals. Investigating the genomic architecture of the American Breeds not only allows the estimation of indicine and taurine genome proportions genome-wide, but also the locations within the genome where either taurine or indicine alleles confer a selective advantage.

Keywords: indicine, taurine, hybrids, American breeds, selection, haplotype

INTRODUCTION

Indicine cattle were first imported into the United States from India in 1906 and then from Brazil in the 1920's and were used via crossbreeding with taurine cattle and backcrossing to develop the *Bos taurus indicus* Brahman (Sanders, 1980) which has very little residual *Bos taurus taurus* within its genome (Chan et al., 2010). The American Breeds of beef cattle are populations that were developed in the United States beginning shortly after the introduction of the *B. t. indicus* cattle to capitalize on breed complementarity and heterosis for production and adaptation to heat stress and the nutritional limitations, parasites, and disease-causing pathogens prevalent in the southern tier of the country (Cartwright, 1970; Dickerson, 1970). Indicine × taurine crossbred individuals have been widely produced throughout subtropical and tropical regions of the world (Porto-Neto et al., 2014; Goszczynski et al., 2018) and the use of systematic crossbreeding programs world-wide has resulted in the development of at least 46 recognized indicine × taurine breeds (https://en.wikipedia.org/wiki/List_of_cattle_breeds). Breed Associations for the American Breeds began to be formed in the 1940's and advanced generation composite animals now exist for the older Brangus, Beefmaster, and Santa Gertrudis breeds.

Brangus cattle were derived from animals created in public and private breeding experiments involving crosses between Angus (*B. t. taurus*) and Brahman cattle in Oklahoma, Mississippi, Texas, and Louisiana in the 1930's and have been stabilized at an expected genome content of ⅜ Brahman and ⅝ Angus (http://afs.okstate.edu/breeds/cattle/brangus/index.html/). The American Brangus Breeders Association was formed in 1949 but was later renamed the International Brangus Breeders Association (https://gobrangus.com/jan-17-bj-con-lilley/). Santa Gertrudis cattle were initially developed on the King Ranch in Kingsville, Texas, where experimental crossbreeding between Shorthorn (*B. t. taurus*), and Brahman cattle between 1910 and 1920 led to the birth of the bull "Monkey" from which all registered Santa Gertrudis cattle descend (http://afs.okstate.edu/breeds/cattle/santagertrudis/index.html). However, the utilized Brahman bulls ranged in composition from ¾ to ⅞ *B. t. indicus* and, consequently, the Santa Gertrudis breed is considered to have a composition of ⅜ Brahman and ⅝ Shorthorn (Rhoad, 1949; Warwick, 1958). Santa Gertrudis was recognized as a breed by the United States Department of Agriculture in 1940. The foundation animals for the Beefmaster breed were developed beginning in 1908 as cross between Brahman, Shorthorn, and Hereford (*B. t. taurus*) on the Lasater Ranch in Falfurrias, Texas and are now maintained at an expected pedigree proportion of ½ Brahman, ¼ Hereford, and ¼ Shorthorn (Warwick, 1958). Beefmaster was recognized as a beef breed by the United States Department of Agriculture in 1954. These American Breeds of cattle now provide an interesting opportunity to study the genomic architectures of advanced generation composites with *a priori* known expected genomic breed proportions based on pedigree that have been exposed to natural selection for adaptation and artificial selection for beef performance traits.

Several approaches have been developed for the estimation of local ancestry (breed of origin of the two alleles present at specific loci) in admixed individuals, however, these applications have primarily been focused on recently admixed populations. Individuals from admixed populations have chromosomes that comprise mosaics of chromosomal segments originating from each of the ancestral populations (Thornton and Bermejo, 2014). On the other hand, global ancestry estimates predict the relative proportions of the ancestral genomes present in an admixed individual, which is an average of the local ancestry estimates, and ignores information pertaining to the variability among locus-specific ancestries (Tang et al., 2005). Drift and strong selection can lead to regions of the genome with ancestries that differ significantly from breed expectation and examination of these regions may identify candidate genes that are under selection and suggest the nature of the selected phenotype. We estimated local ancestry for registered Brangus, Santa Gertrudis, and Beefmaster animals that had been genotyped with the BovineSNP50, or derivative assays, and examined the average ancestries at specific chromosomal locations to identify regions of the genome that differ from expected global proportions both within and across breeds. Using the total number of haplotypes detected in each animal's genome as a proxy for its generation number, we also contrasted the genomes of early- and advanced-generation animals to ascertain those genomic regions which had been exposed to recurrent selection within each of the breeds.

MATERIALS AND METHODS

Genotype Data

Genotype data were obtained for deidentified individuals from the International Brangus Breeders Association, Beefmaster Breeders United, and Santa Gertrudis Breeders International Breed Associations (American Breed Associations) (**Table 1**). These individuals had been genotyped using at least one of 8 commonly used assays including the GeneSeek (Lincoln, NE) BOVG50v1, GGP-90KT, GGP-HDV3, GGP-LDV3 and GGP-LDV4, the Illumina (San Diego, CA) BovineHD and BovineSNP50, and the Zoetis (Kalamazoo, MI) i50K. PLINK1.9 (Purcell et al., 2007) was used to filter variants and individuals. SNP positions were based on the ARS-UCD1.2 bovine reference genome assembly (Rosen et al., 2020). Non-autosomal variants were removed from the data. Variants and individuals with genotype call rates <0.90 were also removed. Genotypes were phased using Eagle 2.4 (Loh et al., 2016) with a reference panel of haplotypes for 9,937 individuals genotyped with the BovineHD (HD) assay. Phased haplotypes were then imputed to the SNP content represented in the union of the HD and GeneSeek GGP-F250 (F250) assays using Minimac3 (Das et al., 2016). The multi-breed reference set created by Rowan et al. (2019) was used for genotype imputation. The reference panel contained 2,719 animals that had been genotyped with both the F250 and the HD assays, 25,772 animals genotyped with only the F250, and 7,218 animals genotyped with only the

TABLE 1 | Genotyped samples for the American Breeds.

Assay[a]	No. Brangus	No. Beefmaster	No. Santa Gertrudis
BOVG50v1	0	836	264
GGP-90KT	688	0	6
GGP-HDV3	1,003	1,199	0
GGP-LDV3	0	36	756
GGP-LDV4	5,597	304	897
BovineHD	982	0	23
BovineSNP50	1,174	65	0
Zoetis i50K	14	1,332	0
Total[b]	9,458	3,772	1,946
Total[c]	9,161	3,762	1,942

[a]*Assays used to genotype the samples. Genotypes were imputed to 836,118 variants.*
[b]*Number of individuals passing quality control after imputation and phasing.*
[c]*Number of individuals passing filtering for breed composition.*

HD assay. Following imputation, each sample had genotypes for 836,118 variants.

Reference Panels

Local ancestry estimation requires a reference panel of genotypes for representatives of each ancestral population. We developed two reference panels for each American Breed. The first panel was identical for all three breeds and comprised animals registered by the American Angus Association, American Hereford Association, American Shorthorn Association, and American Brahman Breeders Association for which CRUMBLER (Crum et al., 2019) breed composition estimates were ≥85% to the respective breed (**Table 2**). The second reference panel was created using only individuals from the respective ancestral breeds (ANCESTRAL reference) (**Table 2**). Since some ancestry estimation software is influenced by unequal reference panel sample sizes (Maples et al., 2013; Crum et al., 2019), the breed possessing the smallest number of available individuals determined the approximate random sample size for the remaining breed(s). Reference panel individuals had been genotyped using one of 8 commercially available assays including the GeneSeek BOVG50v1, GGP-F250, GGP-HDV3, and GGP-LDV3, the BovineHD, BovineSNP50, i50K, and Irish Cattle Breeding Federation (Cork, Ireland) IDBv3, and were phased and imputed following the same procedures as for the American Breed individuals.

Local Ancestry Estimation

Gobena et al. (2018) used ADMIXTURE to estimate ancestry in an Angus × Brahman population, however, we have previously found the software to be sensitive to the order of animals within the input files (Crum et al., 2019) and consequently we used RFMix v2.03, for local ancestry and admixture estimation (Maples et al., 2013). RFMix partitions chromosomes into non-overlapping windows and infers ancestry for each window. If contiguous windows are assigned the same ancestry, RFMix concatenates the windows resulting in a variable number of haplotypes of varying length predicted for each individual. We first inferred local ancestry with a window size of 100 SNPs (spanning an average of ~300 kb) using two reference panels for comparison. We next examined a window size of 25 SNPs (spanning on average ~75 kb) for the ANCESTRAL reference panel. RFMix allows the specification of the number of generations separating the query samples from the initial reference population admixture event. However, because pedigree and generation information were not provided by the American Breed Associations, we used a generation interval of 5 years and the dates of formation of each Breed Association to arrive at an estimate of a maximum of 16 generations, with an average of about 8 generations for these individuals. This reflects the fact that first generation American Breed animals are continuously being generated and registered by breeders using superior animals from the requisite foundation breeds to capture the benefits of on-going selection within the numerically larger foundational breeds.

Many of the American Breed Associations allow the registration of the purebred, F_1, and back-cross animals used to create first generation registered animals to enable a complete pedigree based on consistent registration numbers within their herdbook. To determine if some of these animals may have been genotyped and provided by the Breed Associations, we removed samples assigned by CRUMBLER to be ≥ 50% Brahman or ≥ 90% Angus or Shorthorn ancestry from the Brangus and Santa Gertrudis data, respectively (Brangus, n = 297; Santa Gertrudis, n = 4). For the Beefmaster, we removed samples with ≥ 90% assignment to Shorthorn, Hereford, or Brahman to remove potential purebred founders and any samples with ≤ 5% assignment to any one of these breeds to remove potential F_1 individuals (n = 10).

Generation Proxy

Genotypes used in this study were provided by the American Breed Associations with each animal's identity anonymized and without

TABLE 2 | Genotype data for registered individuals from 4 breeds used to generate reference panels.

Breed	No. Registered[a]	No. Individuals >85%[b]	CRUMBLER Reference Panel[c]	ANCESTRAL Reference: Brangus[d]	ANCESTRAL Reference: Santa Gertrudis[d]	ANCESTRAL Reference: Beefmaster[d]
Angus	6,699	252	200	997[e]	—	-
Hereford	3,651	227	200	—	—	500
Shorthorn	487	183	183	—	487	487
Brahman	954	361	200	954	500	500
Total	11,791	1,422	783	1,954	987	1,487

[a]*Total number of available registered animals with genotypes.*
[b]*Number of registered animals identified with ≥85% CRUMBLER assignment probability to the respective breed.*
[c]*Random sample of ≤200 animals/breed from individuals with ≥85% CRUMBLER assignment probability to their respective breed.*
[d]*Ancestral breed sample sizes were determined by the breed with the fewest available registered animals.*
[e]*A random sample of 1,000 Angus resulted in 997 animals remaining following quality control for imputation and phasing.*

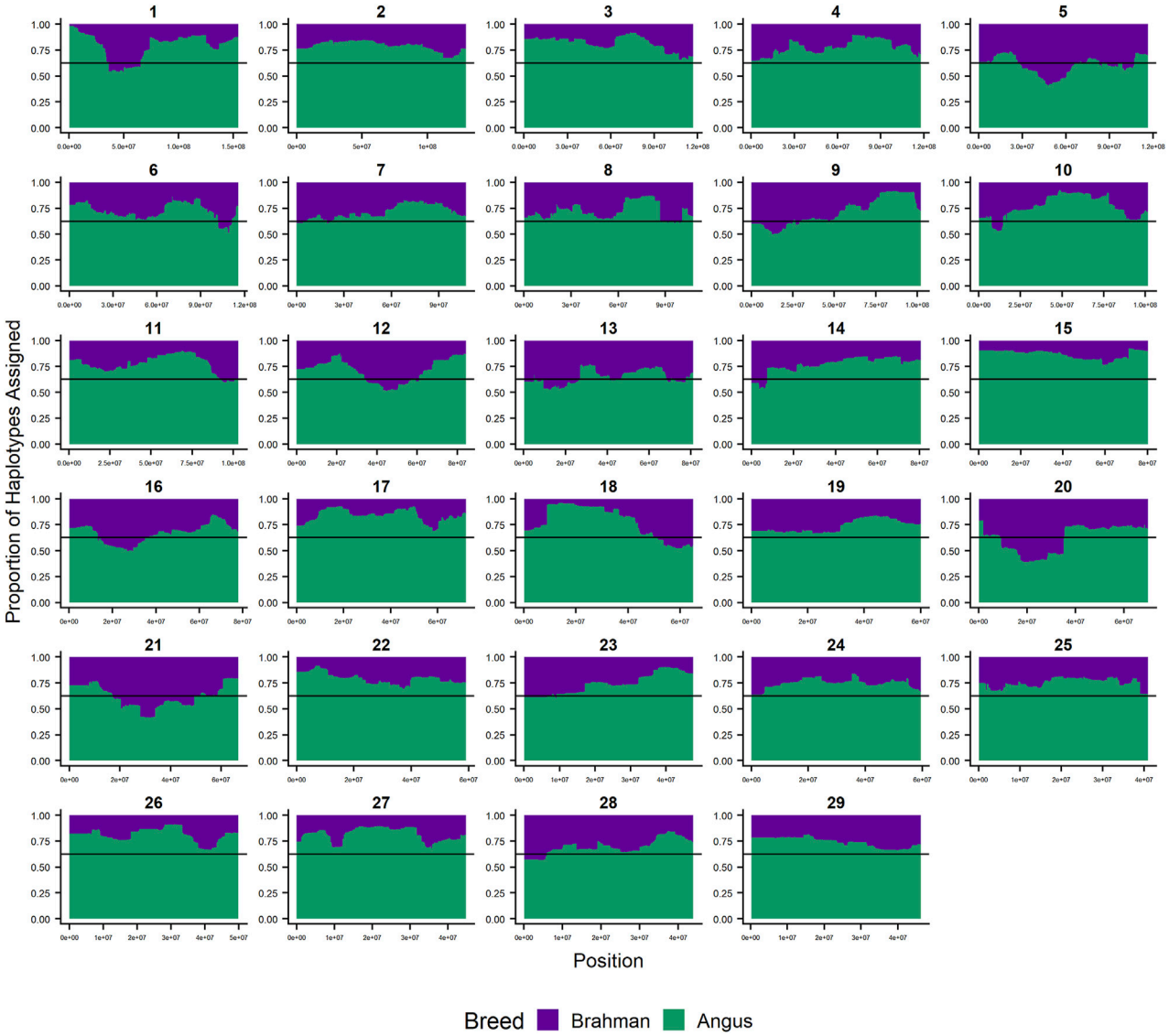

FIGURE 1 | RFMix most likely assignment by chromosome for the ANCESTRAL reference panel for each 25 SNP window for Brangus. Brahman (purple) and Angus (green). Horizontal line indicates 5/8 expected Angus proportion.

pedigree or generation information. Since selection operates each generation, the cumulative effects of selection on a composite genome should increase with generation number. To enable the stratification of the animals within each American Breed based on generation number, we utilized the RFMix output to compute the total number of taurine and indicine haplotypes present within the diploid autosomal genome ($2N = 58$) of each animal and the average length of all haplotypes within the diploid genome. Under the assumption of selective neutrality, the number of haplotypes within the genome should increase each generation due to recombination and correspondingly, the average length of the haplotypes should decrease each generation. Purebred and F_1 animals have 58 haplotypes represented as full length chromosomes and assuming an average of one crossover per chromosome pair each meiosis, back-cross (Purebred × F_1) animals have, on average, 87 haplotypes which average ¾ of the chromosome length and F_2 ($F_1 \times F_1$) animals have, on average, 116 haplotypes which average ½ of the chromosome length.

Genomic Divergence From Breed Expectation

Within each of the American Breeds, for the *i*th window within the genome we tested the null hypothesis that Ho: $\theta_i = \theta$ against the alternate hypothesis Ha: $\theta_i \neq \theta$ using the Z-statistic:

$$Z_i = \frac{p_i - \theta}{\sqrt{Var(\theta_i)}}$$

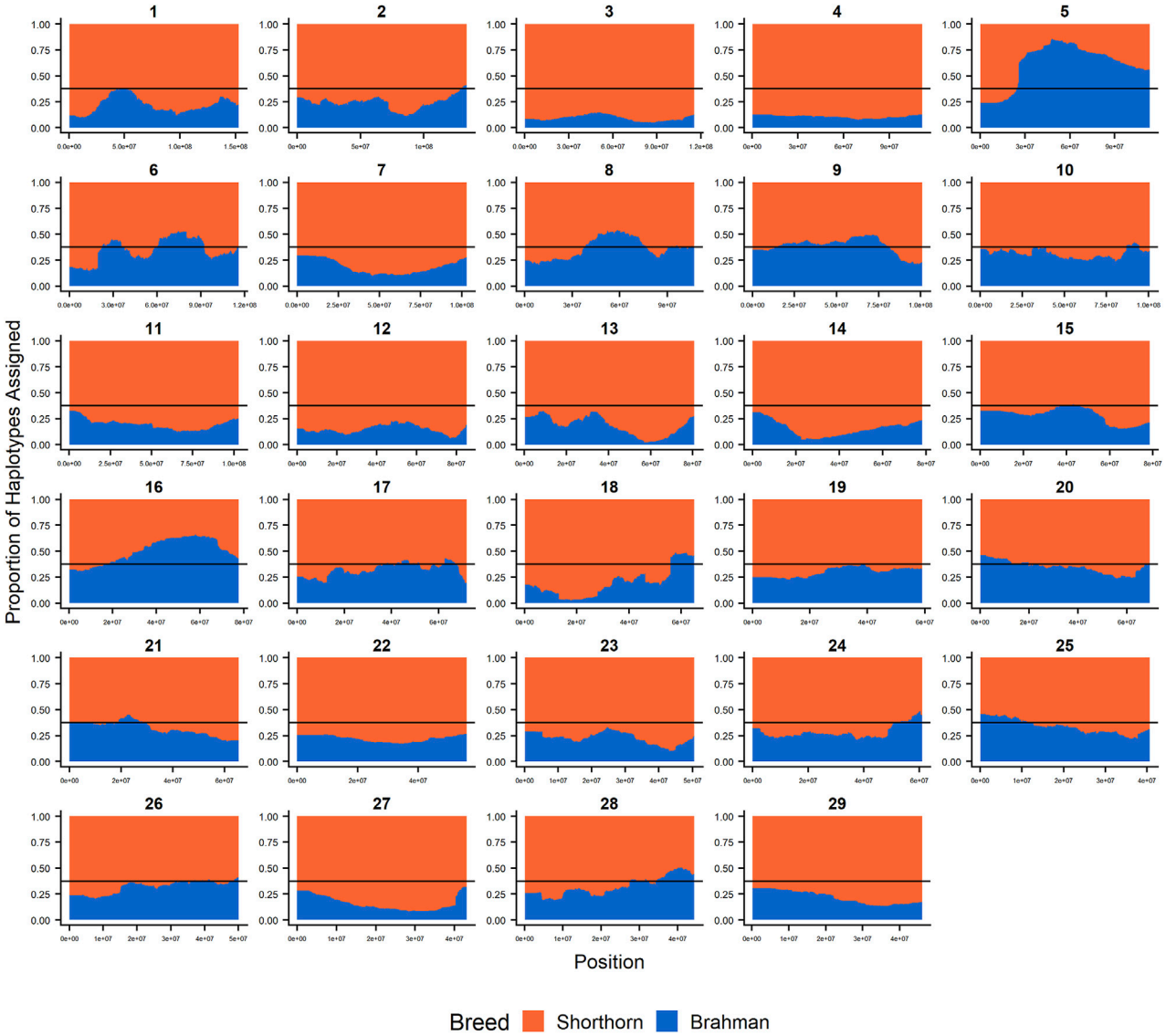

FIGURE 2 | RFMix most likely assignment by chromosome for the ANCESTRAL reference panel for each 25 SNP window for Santa Gertrudis. Brahman (blue) and Shorthorn (orange). Horizontal line indicates 3/8 expected Brahman proportion.

Here, θ_i is the breed Brahman proportion and p_i is the sample Brahman proportion within the *i*th window. The parameter θ represents the American Breed's genome average expected Brahman proportion under selective neutrality in the absence of drift and was set to $\theta = 0.375$ for Brangus and Santa Gertrudis, and 0.5 for Beefmaster. $\mathrm{Var}(\theta_i)$ is the variation in Brahman proportion across windows throughout the genome under selective neutrality and in the absence of drift. $\mathrm{Var}(\theta_i)$ cannot be estimated from the sample unless the null hypothesis is true and so to obtain an estimate of this parameter for each breed composition, we conducted a simulation using 1,000 animals per generation with each animal genotyped at the 836,118 loci with ARS-UCD1.2 reference genome coordinates. The number of recombination events per chromosome was assumed to follow a Poisson distribution with mean (λ) determined by chromosome length × recombination rate (e.g., λ = 158.532931 Mb × 0.01 recombination events per Mb = 1.58532931 for chromosome 1). The location of recombination events within each chromosome was simulated by sampling from a uniform distribution. For the Brangus and Santa Gertrudis simulation, 1,000 first generation ⅜ × ⅝ genomes were simulated by first creating 1,000 F_1 × Purebred crosses (¾ taurine × ¼ indicine) which were then randomly mated to independently sampled F_1 individuals. The first-generation animals were then randomly mated to produce 1,000 s generation individuals and so on for 8 generations of random mating. Ten replicate simulations were performed and within each replicate, θ was estimated as the average Brahman proportion across all genotyped loci and $\sqrt{Var(\theta_i)}$ was estimated as the square root of the variance of the Brahman proportion across all loci in the 8th generation individuals. Estimates of θ $\left(\hat{\theta} \pm SD\left(\hat{\theta}\right);\ 0.3722 \pm 0.0013\right)$ and

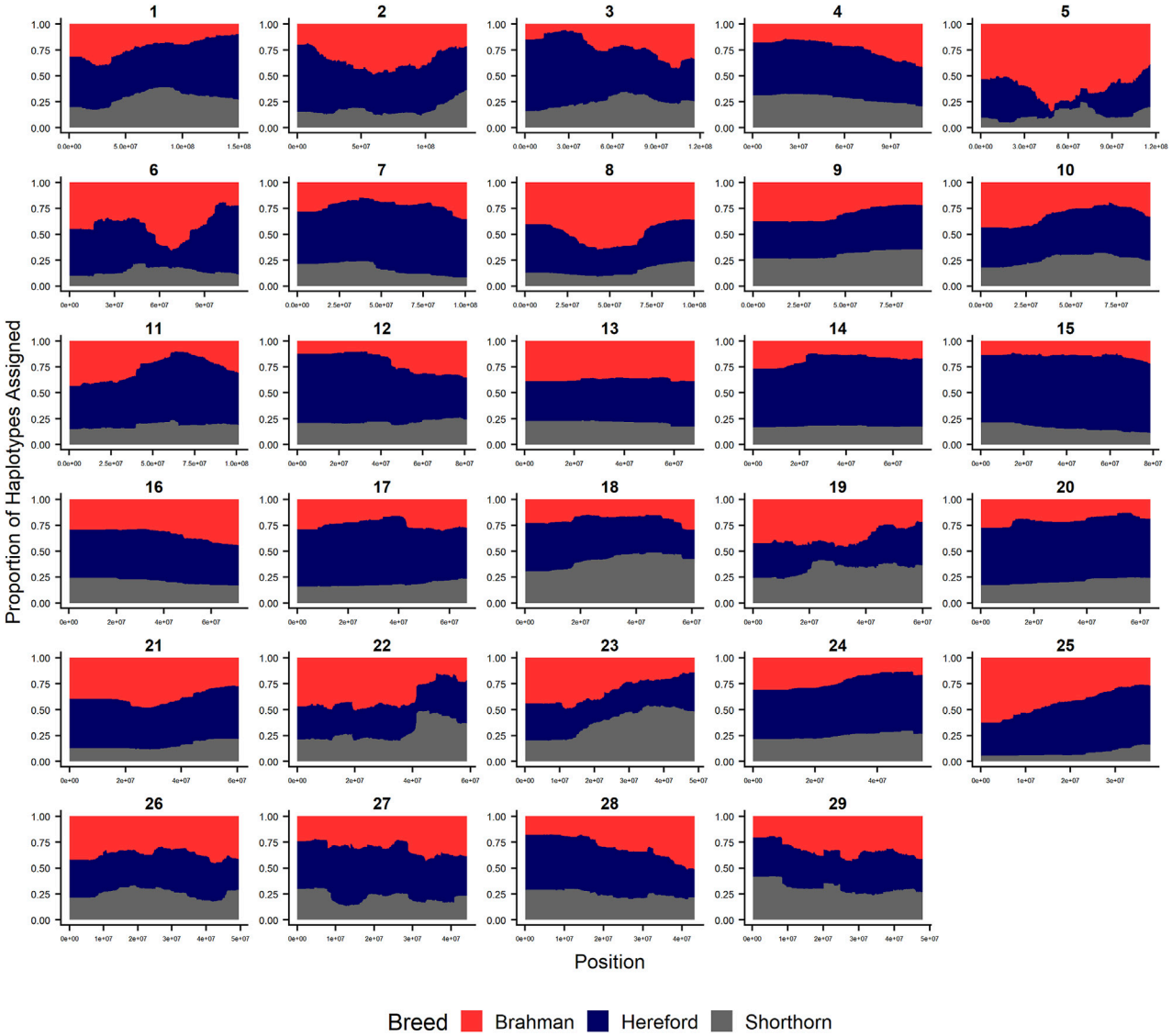

FIGURE 3 | RFMix most likely assignment by chromosome for the ANCESTRAL reference panel for each 25 SNP window for Beefmaster. Brahman (red), Hereford (navy) and Shorthorn (gray). Beefmaster is expected to be ½ Brahman, ¼ Hereford and ¼ Shorthorn.

$\sqrt{Var(\theta_i)}$ $(\sqrt{\widehat{\text{Var}(\theta_i)}} \pm SD(\sqrt{\widehat{\text{Var}(\theta_i)}}); 0.0106 \pm 0.0002)$ were obtained by averaging estimates across replicates. The Beefmaster simulation was identical to that for the Brangus and Santa Gertrudis except the 1,000 first generation ½ × ¼ × ¼ genomes were simulated by randomly sampling and mating 1,000 F_1 genomes (½ taurine × ½ indicine). Estimates of θ and $\sqrt{Var(\theta_i)}$ were 0.4976 ± 0.0013 and 0.0111 ± 0.0002, respectively. These estimates of $\sqrt{\widehat{Var(\theta_i)}}$ were used in place of $\sqrt{Var(\theta_i)}$ in the calculation of the Z-statistics for each of the American Breeds. Controlling for the error rate in multiple testing of windows within each breed was achieved using adjusted *p*-values proposed by Benjamini and Hochberg. (1995) at a false discovery rate (FDR) $\leq$ 0.001. Significant regions were queried for QTL reported in the CattleQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/BT/search; Hu et al., 2019).

Genomic Divergence Between Early- and Advanced-Generations

Within each of the American Breeds, we sampled ~10% of individuals with the smallest and largest number of haplotypes with their genomes to characterize early- and advanced-generation individuals, respectively. Sampling was such that all animals within a haplotype number class were included resulting in slightly unequal sample sizes. There were 918, 207, and 400 animals in the early- and 955, 213, and 423 Brangus, Santa Gertrudis, and Beefmaster animals in the advanced-generation groups, respectively.

For the *i*th window within the genome, we tested the null hypothesis that Ho: $\theta_{ei} = \theta_{ai}$ against the alternate hypothesis Ha: $\theta_{ei} \neq \theta_{ai}$ using the Z-statistic:

TABLE 3 | Average ancestry to reference populations by chromosome for each of the American breeds[a].

Chr	Brangus Av. Brahman % (SD)	Santa Gertrudis Av. Brahman % (SD)	Beefmaster Av. Brahman % (SD)	Beefmaster Av. Hereford % (SD)
1	21.69 (12.25)	23.03 (7.83)	22.51 (9.03)	48.00 (6.42)
2	21.07 (5.03)	24.70 (6.18)	34.79 (9.10)	47.57 (7.25)
3	19.18 (6.94)	***9.51 (2.83)***	22.67 (10.97)	52.48 (13.60)
4	22.06 (6.91)	***10.45 (1.21)***	24.74 (8.59)	47.69 (4.92)
5	**39.32 (8.54)[b]**	**63.28 (16.19)**	**60.85 (9.93)**	26.73 (11.78)
6	28.33 (7.86)	34.63 (10.73)	40.97 (12.79)	44.45 (14.61)
7	28.56 (6.36)	18.33 (6.26)	21.72 (4.82)	61.74 (5.08)
8	27.81 (7.90)	**36.88 (10.00)**	**51.00 (10.14)**	34.81 (6.59)
9	27.62 (13.19)	38.45 (8.42)	29.20 (6.34)	39.57 (2.97)
10	24.62 (10.82)	31.65 (4.43)	31.19 (8.18)	42.67 (4.29)
11	23.00 (8.83)	19.78 (4.82)	26.72 (10.87)	55.11 (9.10)
12	27.82 (10.95)	***14.46 (4.22)***	22.43 (9.65)	55.76 (11.09)
13	35.18 (6.93)	19.16 (8.68)	36.60 (1.17)	42.06 (1.56)
14	23.43 (7.82)	16.28 (6.98)	***17.20 (4.40)***	65.29 (3.95)
15	***13.78 (4.15)[c]***	28.19 (7.46)	***14.62 (2.11)***	70.17 (2.00)
16	33.27 (9.25)	**51.21 (11.00)**	34.93 (5.57)	45.16 (3.30)
17	***16.03 (5.90)***	32.78 (6.45)	24.40 (4.24)	57.11 (5.98)
18	25.00 (15.38)	21.66 (13.73)	***18.93 (3.75)***	38.68 (5.51)
19	26.11 (6.10)	30.69 (4.41)	36.49 (7.79)	29.70 (7.87)
20	**38.38 (13.24)**	34.82 (5.83)	19.22 (3.37)	59.62 (1.90)
21	**38.80 (10.37)**	31.52 (7.06)	38.74 (7.03)	45.54 (3.38)
22	20.87 (5.11)	21.67 (2.97)	37.35 (12.35)	32.77 (4.53)
23	24.40 (8.75)	22.80 (5.72)	31.03 (10.94)	29.10 (4.16)
24	24.80 (3.89)	28.83 (6.86)	21.05 (6.25)	53.63 (3.73)
25	24.87 (4.66)	33.40 (6.82)	**41.11 (10.78)**	50.42 (7.89)
26	19.11 (6.62)	32.93 (5.63)	36.43 (4.43)	37.57 (3.80)
27	***18.06 (6.56)***	14.82 (6.16)	31.44 (5.67)	47.51 (4.95)
28	28.82 (7.27)	34.05 (9.75)	32.30 (10.74)	43.56 (8.18)
29	26.63 (4.61)	20.33 (5.77)	33.86 (5.92)	36.25 (3.58)
Genome Wide	25.81 (8.01)	27.60 (7.05)	30.84 (7.48)	46.23 (6.00)

[a]*RFMix local ancestry estimates averaged across all chromosome windows and throughout the genome. SD = standard deviation.*
[b]*Three chromosomes with largest Brahman content within each breed are indicated in bold.*
[c]*Three chromosomes with smallest Brahman content within each breed are indicated in bold and italics.*

$$Z_i = \frac{p_{ei} - p_{ai}}{\sqrt{p_i(1 - p_i)\left(\frac{1}{n_e} + \frac{1}{n_a}\right)}}.$$

Here, θ_{ei} and θ_{ai} are the *i*th window Brahman proportions within the early- and advanced-generation individuals within the breed, which are estimated by the sample proportions p_{ei} and p_{ai}, respectively. The statistic $p_i = \frac{n_e p_{ei} + n_a p_{ai}}{n_e + n_a}$ is the pooled estimate of the Brahman proportion when the null hypothesis is true and n_e and n_a are the numbers of haplotypes within the early- and advanced-generation individuals (twice the sample sizes), respectively. Multiple testing error rate was again controlled using Benjamini and Hochberg. (1995) adjusted *p*-values at a FDR $\leq$ 0.001. Significant regions were queried for QTL reported in the CattleQTLdb. This test was also applied genome-wide to test for differences in the global Brahman genome composition of early- and advanced-generation individuals.

RESULTS

Reference Panels

RFMix generates three output files for each analysis: 1) the most likely reference population assignments for each haplotype found in a window, 2) the marginal probabilities of each reference population being the ancestral population of origin for the haplotypes found in the window, and 3) global diploid ancestry estimates (Maples et al., 2013). We utilized the most likely assignment files to define the breed of origin of haplotypes found within each window and then computed the relative frequencies of haplotype origin from each representative breed in each of the windows.

Following the strategy employed by Browning et al. (2016), we created two reference panels to evaluate the effects of panel size on the estimation of local ancestry. We first selected the Brahman, Angus, Hereford, and Shorthorn animals present in the CRUMBLER (Crum et al., 2019) reference panel which had been extensively developed and evaluated for use in global ancestry estimation for breeds commonly found in North America. Alternatively, for each of the American Breeds, an ANCESTRAL reference panel was created from all available registered animals. To avoid potential issues caused by unequal sample sizes, we randomly down-sampled the larger populations to the sample size for the breed with the fewest available samples (**Table 2**). For each of the American Breeds, the ANCESTRAL reference panel contained at least twice the number of animals from the ancestral populations than the CRUMBLER reference panel and ancestry estimates produced using the two panels were similar (**Supplementary Figures**

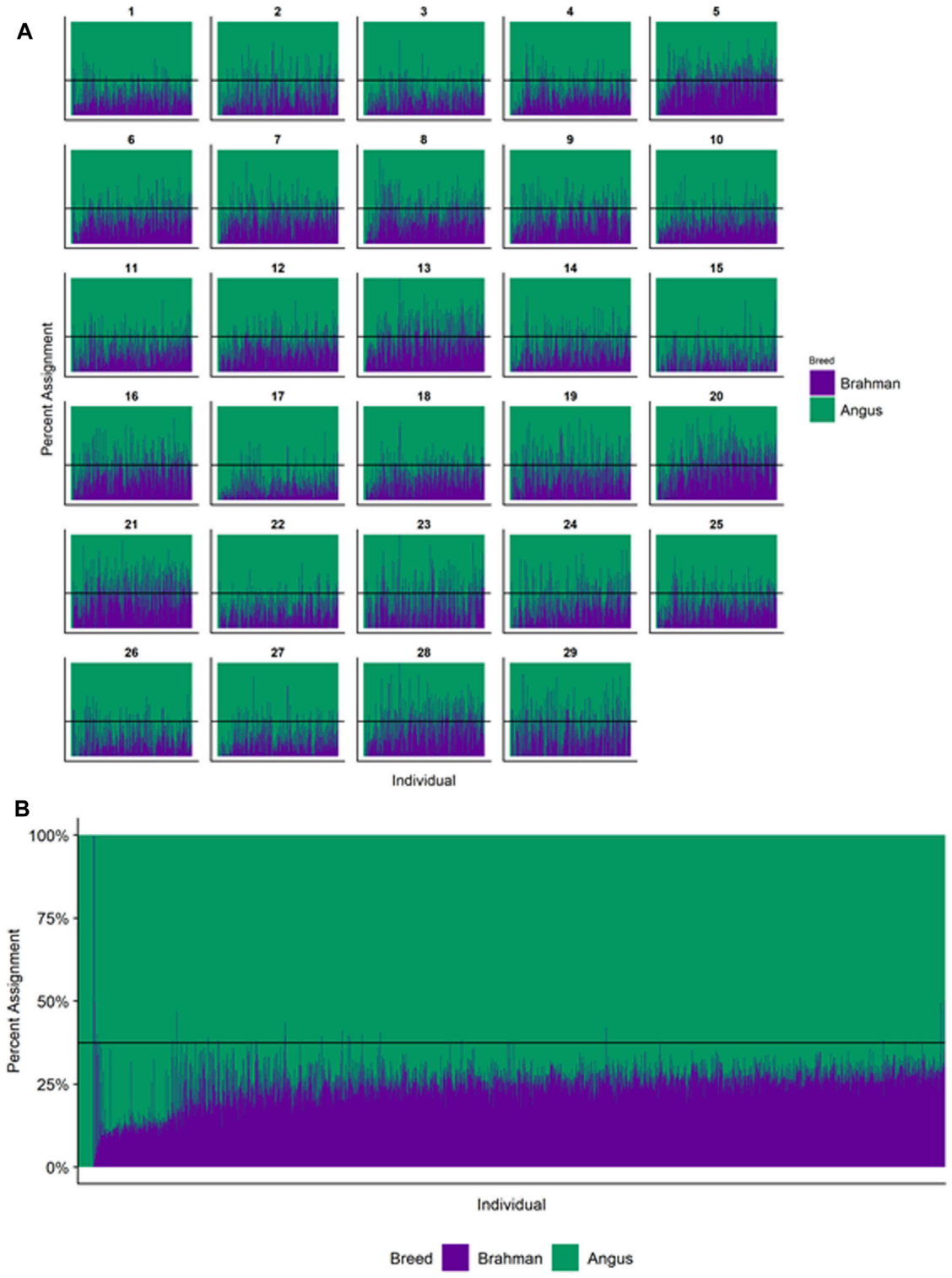

FIGURE 4 | Ancestry estimated by RFMix for each Brangus individual plotted with animals sorted by total number of haplotypes. **(A)** Ancestry assignment for each chromosome, and **(B)** Genome-wide ancestry. Individuals are represented by vertical bars within each plot with Brahman proportion in purple and Angus proportion in green. Animals to the left on the *X*-axes have the fewest number of haplotypes predicted within their genomes. The black line represents the expected proportion of 3/8 Brahman.

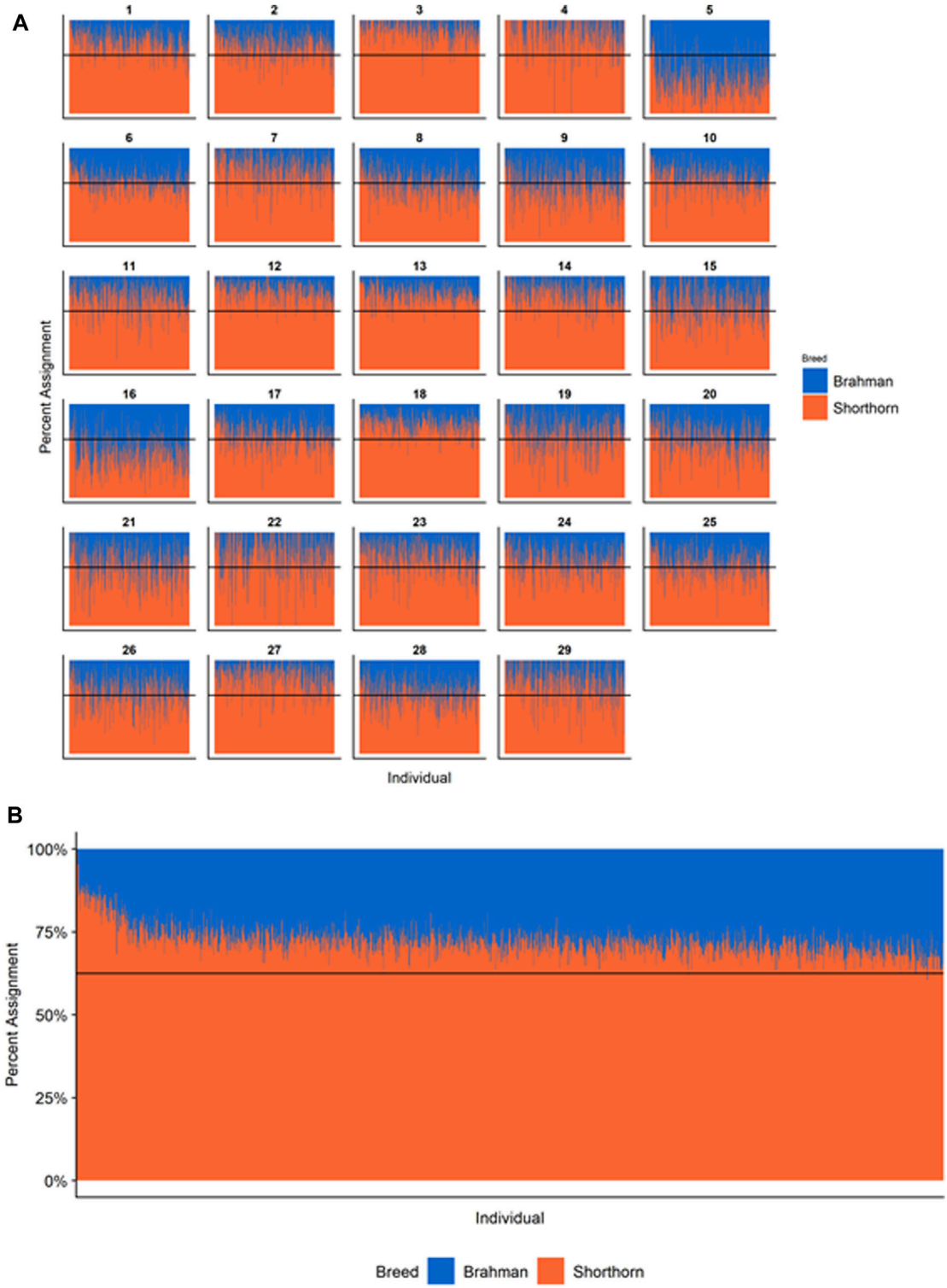

FIGURE 5 | Ancestry estimated by RFMix for each Santa Gertrudis individual plotted with animals sorted by total number of haplotypes. **(A)** Ancestry assignment for each chromosome, and **(B)** Genome-wide ancestry. Individuals are represented by vertical bars within each plot with Brahman proportion in blue and Shorthorn proportion in orange. Animals to the left on the *X*-axes have the fewest number of haplotypes predicted within their genomes. The black line represents the expected proportion of 5/8 Shorthorn.

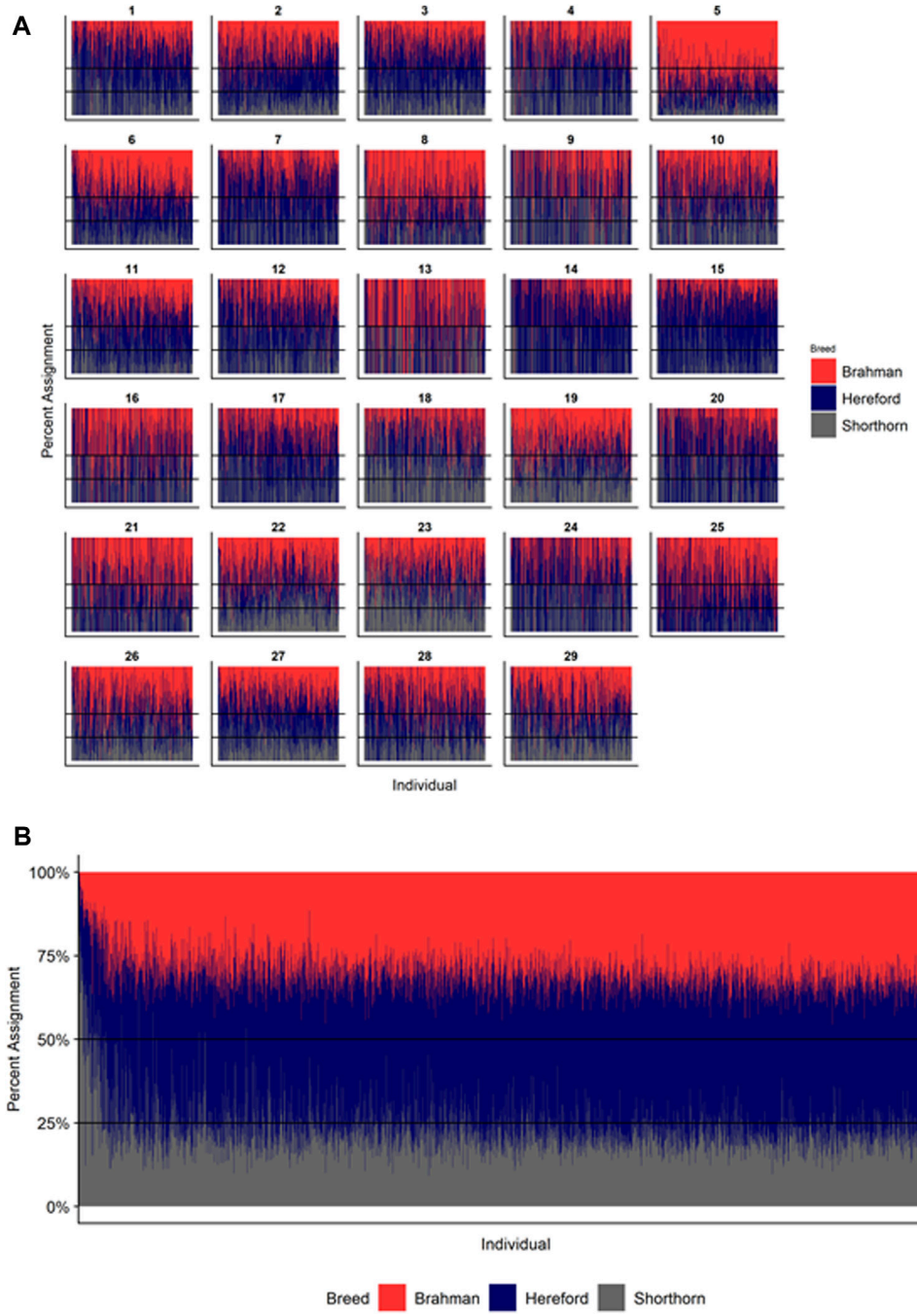

FIGURE 6 | Ancestry estimated by RFMix for each Beefmaster individual plotted with animals sorted by total number of haplotypes. **(A)** Ancestry assignment for each chromosome, and **(B)** Genome-wide ancestry. Individuals are represented by vertical bars within each plot with Brahman proportion in red, Hereford in navy, and Shorthorn proportion in gray. Animals to the left on the *X*-axes have the fewest number of haplotypes predicted within their genomes.

S1–S3). However, the proportion of Hereford ancestry appeared to be underestimated and the proportion of Shorthorn ancestry overestimated when the CRUMBLER reference panel was used for local ancestry estimation in Beefmaster (**Supplementary Figure S3**). Consequently, we decided to utilize the ANCESTRAL reference panels for all further analyses.

Window Size

Since RFMix concatenates contiguous windows with the same ancestry to form extended haplotypes within individuals and considering the large number of SNPs used in this study, we were interested in whether a window size of 25 SNPs would impact the analysis in terms of resolution near recombination breakpoints or the ability to discriminate between haplotype breed of origin.

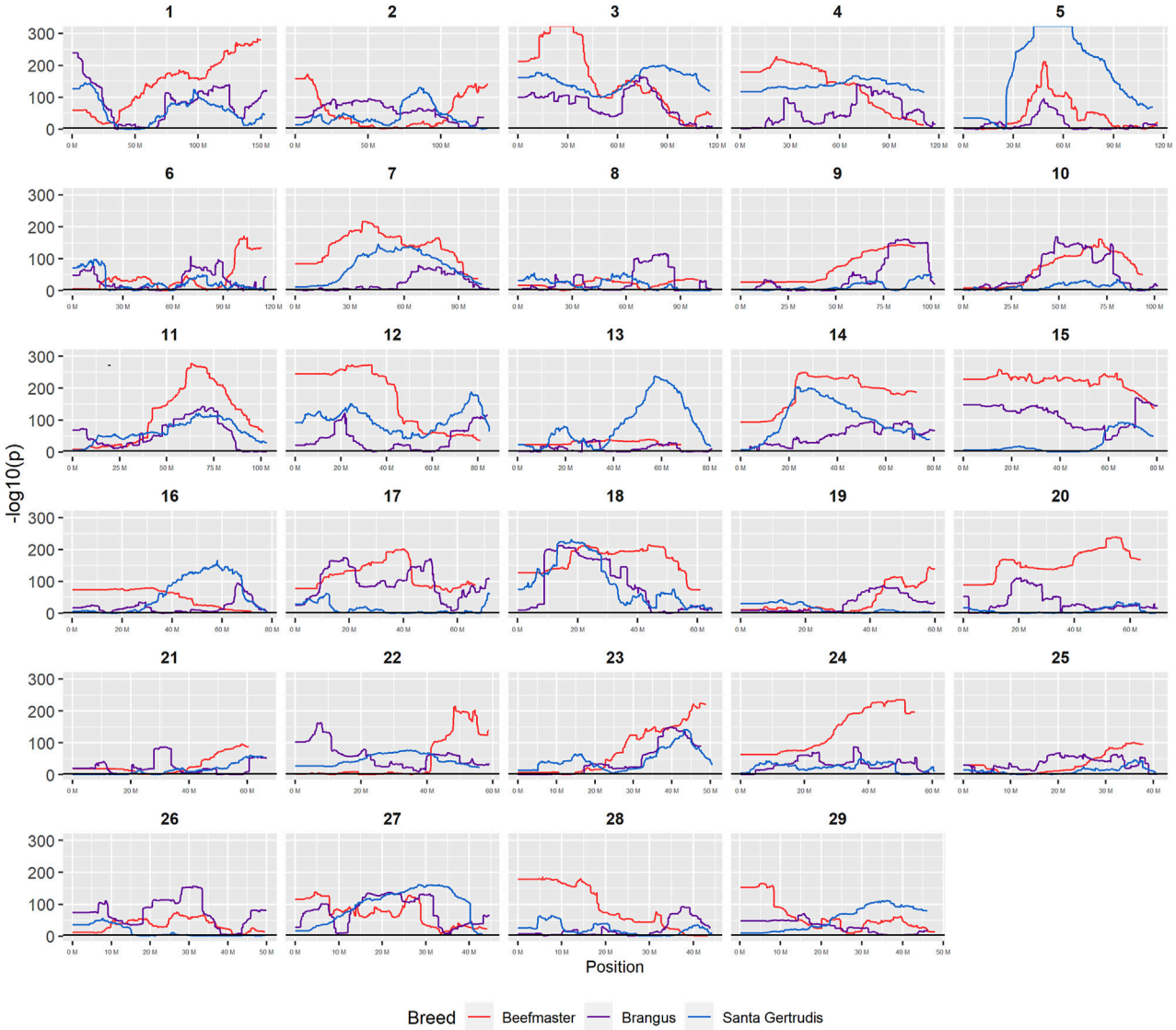

FIGURE 7 | Plots of $-\log_{10}P$ for tests of each genomic window for deviation from breed expected Brahman proportion by chromosome.

Figures 1–3 show the local ancestry by chromosomal location for each of the American Breeds and are strikingly similar to the results shown in **Supplementary Figures S1B, S2B, and S3B**. Consequently, we conducted all further analyses using a window size of 25 SNPs.

Generation Proxy

Using the RFMix output for all samples provided by the American Breed Associations (**Table 2**), we estimated the number of Angus, Shorthorn, Hereford, and Brahman haplotypes present within the genotyped individuals, and also the average length of these haplotypes by breed of origin. Haplotype metrics for all three of the American Breeds are presented in **Supplementary Table S1** and suggest that the genomic architectures of the Brangus and Santa Gertrudis breeds are quite similar. Each breed possesses, on average, 50% more taurine than indicine haplotypes and the taurine haplotypes are almost twice the length of the indicine haplotypes.

The correlation between the number of Angus and Brahman haplotypes in each Brangus individual was 0.80, while the correlation between the number of Shorthorn and Brahman haplotypes in each Santa Gertrudis individual was 0.73. This follows our expectation that the total number of haplotypes from each of the foundation breeds increases in advanced generation animals due to recombination. However, the correlation between the number of Brahman and taurine (Shorthorn and Hereford) haplotypes in each Beefmaster individual was only 0.40. The correlation between the number of Brahman and Hereford haplotypes was 0.50 while the correlation between the number of Brahman and Shorthorn haplotypes was −0.11 suggesting that other forces such as selection may be influencing the evolution of the Beefmaster genome.

TABLE 4 | Five most significantly diverged regions from expected Brahman breed proportion by breed.

Chr	Start Pos.[a] (bp)	Size[b] (bp)	P_i [c]	-log10(P)
		Brangus		
1	376,390	4,082,465[d]	0.02	240.08
10	48,956,514	89,910	0.08	168.77
17	17,920,239	62,461	0.07	175.01
17	50,105,014	68,938	0.07	171.87
18	13,773,453	123,002	0.04	212.80
		Santa Gertrudis		
3	86,676,910	332,543	0.05	201.02
5	42,298,343	184,428	0.79	>300[e]
13	57,508,582	82,009	0.02	237.25
14	23,570,574	96,661	0.05	204.54
18	17,876,786	67,551	0.03	232.08
		Beefmaster		
1	146,960,035	92,662	0.10	283.59
3	19,467,876	57,666	0.08	>300[e]
11	63,215,699	125,829	0.10	277.62
12	32,131,527	236,418	0.10	273.19
15	14,921,941	334,777	0.12	259.04

[a]*Window start coordinate.*
[b]*Window size in bp.*
[c]*Brahman proportion within the window.*
[d]*Contiguous windows concatenated by RFMix because full length haplotypes for all individuals were either Brahman or Angus.*
[e]*Due to rounding error* p = *0.00 and* $-log_{10}P$ = *infinity.*

Breed Ancestry

The average Brahman genome content was 25.81 ± 8.01% (±standard deviation among sampled individuals) for Brangus and 27.60 ± 7.05% for Santa Gertrudis (**Table 3**), far less than the 37.5% based upon breed expectations. Likewise, the Beefmaster had considerably fewer haplotypes of Brahman origin of smaller than average length (**Supplementary Table S1**) resulting in an average Brahman genome content of 30.84 ± 7.48%, far less than the breed expectation of 50%. The average Shorthorn genome content in Beefmaster individuals was 22.93 ± 8.01% and the average Hereford genome content was 46.23 ± 6.00%.

Table 3 shows that chromosome 5 consistently had the highest Brahman ancestry in all three breeds, but there was not a strong concordance between chromosomes with the highest or lowest Brahman ancestry across the breeds. Despite this, the correlations between estimated Brahman ancestry percentages across all 29 autosomes were 0.49 for Brangus and Santa Gertrudis, 0.43 between Brangus and Beefmaster, and 0.61 between Santa Gertrudis and Beefmaster which also share a Shorthorn ancestry. These correlations certainly indicate that, on a chromosomal basis, there is a tendency for the breeds to share elevated or reduced Brahman ancestry.

Individual Ancestry

Figures 4–6 show the proportions of ancestral breed inheritance for each individual by chromosome (panel A) and genome-wide (panel B). Individuals within each figure are sorted by the total number of haplotypes predicted within their genomes (*X*-axes and see **Supplementary Table S1**). For all three breeds, the proportion of Brahman content within the diploid genome increases with haplotype number, which we use as a proxy for the generation number of these individuals (see also **Supplementary Figure S4**). The results for individual chromosomes (**Figures 4A**, **5A**, **6A**) are not quite as obvious, however, the evolution of chromosome 5 clearly differs from most of the other chromosomes in all three breeds (see also **Table 3**).

Genomic Divergence From Breed Expectation

Nominal significance thresholds to achieve a FDR <0.001 using the Benjamini and Hochberg (1995) procedure were $-\log_{10} p$ = 3.0527 for Brangus, 3.0506 for Santa Gertrudis, and 3.0193 for Beefmaster. Using these values, 88.6% of tests performed for Brangus, 85.2% for Santa Gertrudis, and 95.2% for Beefmaster were significant. Values of the $-\log_{10}P_i$ values for each American breed are plotted according to their chromosomal coordinates in **Figure 7**. These results reveal that most of the genomes of each of the American Breeds differ in Brahman composition from the expected breed proportion based upon pedigree when assuming selective neutrality and the absence of drift.

Regions With the Greatest Brahman Divergence From Breed Expectation

The 5 most differentiated windows for Brahman genome content within each of the American breeds are in **Table 4**. These windows vary in size due to variation in the distribution of SNP locations throughout the genome, but also because RFMix concatenates contiguous windows where all individuals have the same ancestral population origin of haplotypes. Except for the locus on chromosome 5, all regions were enriched for taurine alleles. **Table 5** contains beef trait QTL from the CattleQTLdb within the 1 Mb region centered on the regions enriched for taurine alleles in **Table 4**, except for the regions on chromosomes 1 and 18 in Brangus which are separately discussed.

Genomic Divergence Between Early- and Advanced-Generations

The average number of haplotypes was 112.10 ± 9.03, 125.21 ± 11.16, and 103.60 ± 4.25 within the early-generation and 177.06 ± 5.12, 172.28 ± 5.13, and 140.71 ± 3.73 within the advanced-generation Brangus, Santa Gertrudis, and Beefmaster individuals, respectively. The average Brahman genome content was 15.92 ± 6.28, 21.39 ± 11.27, and 26.86 ± 14.44 within the early-generation and 29.44 ± 12.52, 32.22 ± 13.41, and 33.77 ± 1 3.90 within the advanced-generation Brangus, Santa Gertrudis, and Beefmaster individuals, respectively. The genome-wide Brahman proportions were significantly greater in the advanced-generation Brangus (p < 3.13E-23), Santa Gertrudis (p < 0.0002), and Beefmaster (p < 0.0012) individuals, respectively.

Nominal significance thresholds to achieve a FDR <0.001 using the Benjamini and Hochberg. (1995) procedure were $-\log_{10}p$ = 3.0317 for Brangus, 3.2030 for Santa Gertrudis, and 3.2399 for Beefmaster. Using these values, 91.23% of tests

TABLE 5 | CattleQTLdb queries for QTL influencing traits selected in U. S. beef cattle for regions enriched for taurine alleles in American Breed cattle.

Chr	Region[a] (bp)	Records[b]	Traits[c]	Most Likely QTL[d]	References
				Brangus	
10	49,001,469	50	28	Calving Ease; Stillbirth	Kühn et al. (2003); Schnabel et al. (2005); Seidenspinner et al. (2009)
				Age At Puberty	Hawken et al. (2012)
				Yearling Weight; Yearling Height	McClure et al. (2010); Snelling et al. (2010)
				Tick Resistance	Machado et al. (2010)
17	17,951,470	38	29	Birth Weight; Calving Ease; Stillbirth	Ashwell et al. (2005); McClure et al. (2010); Cole et al. (2011)
				Female Fertility	Cole et al. (2011)
				Average Daily Gain; Metabolic Body Weight; Slaughter Weight; Yearling Weight	MacNeil and Grosz (2002); Peters et al. (2012); Seabury et al. (2017); Akanno et al. (2018)
				Carcass Fat Thickness; Marbling	Casas et al. (2003); McClure et al. (2010)
				Bovine Respiratory Disease Susceptibility	Neupane et al. (2018)
17	50,139,483	40	11	Female Fertility	Gaddis et al. (2016)
				Santa Gertrudis	
3	86,843,182	41	17	Birth Weight; Calf Size; Calving Ease; Stillbirth	Sahana et al. (2011); Sugimoto et al. (2012)
				Female Fertility	Gaddis et al. (2016)
13	57,549,587	99	47	Birth Weight; Calf Size; Calving Ease; Gestation Length; Stillbirth	Kühn et al. (2003); Cole et al. (2011); Sahana et al. (2011); Sikora et al. (2011); Höglund et al. (2012); Saatchi et al. (2014b); Akanno et al. (2018)
				Male Fertility; Female Fertility; Age At Puberty; Estrus Interval	Sahana et al. (2010); Liu et al. (2017); Cole et al. (2011); Fortes et al. (2013); Höglund et al. (2015); Galliou et al. (2020)
				Stature; Yearling Weight	Kim et al. (2003); Cole et al. (2011)
				Carcass Yield Grade; Marbling	McClure et al. (2010); Saatchi et al. (2014b)
14	23,618,905	95	36	Birth Weight; Calving Ease; Gestation Length; Stillbirth	Kneeland et al. (2004); Maltecca et al. (2008); Snelling et al. (2010); Pausch et al. (2011); Martínez et al. (2014); Saatchi et al. (2014b); Michenet et al. (2016); Akanno et al. (2018); Smith et al. (2020)
				Male Fertility; Female Fertility; Twinning	Cobanoglu et al. (2005); Schnabel et al. (2005); Fortes et al. (2013); Hawken et al. (2012); Irano et al. (2016); Mota et al. (2017); Oliveira et al. (2017); Kiser et al. (2019); Galliou et al. (2020); Oliveira et al. (2020)
				Feed Intake; Feed Efficiency; Growth Rate; Stature; Weaning Weight; Yearling Weight; Carcass Weight; *Longissimus Dorsi* Muscle Area	Spelman et al. (1999); Kneeland et al. (2004); McClure et al. (2010); Snelling et al. (2010); Saatchi et al. (2014a); Saatchi et al. (2014b); Sharma et al. (2014); Brunes et al. (2021); Naserkheil et al. (2020); Smith et al. (2020); Srikanth et al. (2020)
				Carcass Fat Thickness; Marbling	Casas et al. (2003); McClure et al. (2010); Bolormaa et al. (2011); Abo-Ismail et al. (2018)
				Tick Resistance	Gasparin et al. (2007)
18	17,910,562	43	25	Calving Ease; Stillbirth	Kühn et al. (2003); Seidenspinner et al. (2009); Höglund et al. (2012)
				Metabolic Body Weight; Mature Weight; Carcass Weight; Feed Efficiency; Yield Grade; *Longissimus Dorsi* Muscle Area	Casas et al. (2003); Sherman et al. (2009); McClure et al. (2010); Saatchi et al. (2014b); Seabury et al. (2017); Brunes et al. (2021)
				Immune function; M. Paratuberculosis Susceptibility; Bovine Tuberculosis Susceptibility	Leach et al. (2010); Küpper et al. (2014); Wang et al. (2015); Richardson et al. (2016)
				Beefmaster	
1	147,006,366	180	41	Calving Ease; Stillbirth	Cole et al. (2011)
				Female Fertility; Estrus Interval	Cole et al. (2011); Melo et al. (2019)
				Growth Rate	Snelling et al. (2010); Seabury et al. (2017)
				Tick Resistance; Bovine Respiratory Disease Susceptibility	Mapholi et al. (2016); Sollero et al. (2017); Neupane et al. (2018)
3	19,496,709	33	19	Birth Weight; Gestation Length	Casas et al. (2003); McClure et al. (2010); Maltecca et al. (2011)
				Female Fertility	Cochran et al. (2013)
				Weaning Weight; Carcass Weight; Maturity Rate; Meat Yield	McClure et al. (2010); Doran et al. (2014); Saatchi et al. (2014b); Crispim et al. (2015)
				Carcass Fat Thickness	Casas et al. (2003)
				Bovine Tuberculosis Susceptibility	González-Ruiz et al. (2019)
11	63,278,614	74	45	Birth Weight; Calving Ease; Stillbirth	Cole et al. (2011); Sun et al. (2011)

(Continued on following page)

TABLE 5 | (*Continued*) CattleQTLdb queries for QTL influencing traits selected in U. S. beef cattle for regions enriched for taurine alleles in American Breed cattle.

Chr	Region[a] (bp)	Records[b]	Traits[c]	Most Likely QTL[d]	References
				Male Fertility; Female Fertility; Estrus Interval	Höglund et al. (2009b); McClure et al. (2010); Kolbehdari et al. (2008)
				Body Size; Growth; Stature; Yearling Weight; Mature Weight; Carcass Weight	Snelling et al. (2010); McClure et al. (2010); Cole et al. (2011); Imumorin et al. (2011); Sun et al. (2011)
				Carcass Fatness; Marbling	Stone et al. (1999); McClure et al. (2010); Imumorin et al. (2011)
12	32,249,736	17	15	Calf Size; Calving Ease; Stillbirth	Sahana et al. (2011); Höglund et al. (2012)
				Age At Puberty	Stafuzza et al. (2019)
				Weaning Weight; Yearling Weight; Mature Weight; Carcass Weight; Mature Height	McClure et al. (2010); Saatchi et al. (2014b)
				Gastrointestinal Nematode Burden	Kim et al. (2013)
15	15,089,330	25	21	Birth Weight	McClure et al. (2010)
				Male Fertility	Druet et al. (2009); McClure et al. (2010)
				Growth; Yearling Weight; *Longissimus Dorsi* Muscle Area	McClure et al. (2010); Snelling et al. (2010); Li and Kim (2015); Li et al. (2017)
				Carcass Fatness	Casas et al. (2003); McClure et al. (2010)

[a] *1 Mb region centered on this coordinate.*
[b] *QTL/Association records returned from CattleQTLdb (August 10, 2021).*
[c] *Number of traits influenced by QTL/Associations.*
[d] *QTL for traits known to be selected in U. S. registered beef cattle for which taurine alleles are expected to have been selected.*

performed for Brangus, 62.61% for Santa Gertrudis, and 57.46% for Beefmaster were significant. The Brahman proportion was greater in the advanced-generation animals for all significant regions in Brangus and Santa Gertrudis and for 93.84% of the significant regions in Beefmaster. Values of the $-\log_{10}P_i$ values for each American breed are plotted according to their chromosomal coordinates in **Supplementary Figure S6**. These results reveal that most of the genomes of advanced-generation individuals from each of the American Breeds have a greater Brahman composition than early-generation individuals.

Regions With the Greatest Brahman Divergence Between Early- and Advanced-Generations

The 5 most differentiated windows for Brahman genome content between early and advanced-generation animals within each of the American Breeds are in **Supplementary Table S2**. For all 15 regions, the Brahman proportion was greater in the advanced- than in the early-generation animals. **Supplementary Table S3** contains beef trait QTL from the CattleQTLdb within the 1 Mb region centered on the differentiated region in **Supplementary Table S2**.

DISCUSSION

Reference Panels

Reference panel sample sizes have been shown to have significant effects on the accuracy of RFMix estimates (Maples et al., 2013). Reference panel sizes for analyses of local ancestry in human have ranged from as few as 19 samples to more than 500 samples per population (Maples et al., 2013; Browning et al., 2016) and human effective population size has been estimated to be 3,100 for Europeans and 7,500 for Yorubans (Tenesa et al., 2007). In cattle, the effective population size of most taurine breeds has been estimated to be about 100 (Bovine HapMap Consortium et al., 2009) and, consequently, we would expect that smaller reference panel sizes would be sufficient to capture the haplotypic diversity present within cattle breeds than for human. However, random samples of 200 individuals from each ancestral breed may not be sufficient to capture a large proportion of the haplotypic diversity within these breeds and larger sample sizes would certainly capture more of the rare haplotypes potentially leading to a greater accuracy of local ancestry estimation in admixed individuals (Maples et al., 2013). This is particularly important considering that the individuals in the reference panels and the founders of the American Breeds are separated by up to 16 generations and selection and drift may have caused large differences in haplotype frequencies between members of the groups. This appears to have impacted the estimation of local ancestry for the Beefmaster animals where proportions of Shorthorn and Hereford within the genome varied significantly depending on whether the CRUMBLER or ANCESTRAL reference panels were used. This may have been because the representation of horned Hereford and Line 1 Hereford animals was limited by sample size in the CRUMBLER reference panel and animals from these Hereford populations would have been prevalent at the time that the Beef master was initially formed.

Generation Proxy, Breed, and Individual Ancestry

To examine genome evolution in these breeds, we utilized the total number of haplotypes present within the genomes of the animals as proxy for the generation number of the animal. Under selective neutrality we would expect the number of taurine and

indicine haplotypes present within individuals to increase with generation number due to recombination among homologous chromosomes at each meiosis. Indeed, we found quite strong correlations (70–80%) between the numbers of indicine and taurine haplotypes within Brangus and Santa Gertrudis, but less so (50%) in Beefmaster. Strong selection for alleles found predominantly in one of the breeds will reduce the strength of this correlation and reduce the accuracy of the proxy for the prediction of generation number. However, we did not have any animals with known generation numbers with which to directly validate the utility of the proxy. Moreover, the results for Beefmaster may have been affected by our use of the most likely breed of haplotype origin, an absolute assignment, rather than the marginal probabilities of breed assignment. If the marginal probabilities of haplotype assignment to Shorthorn and Hereford are similar, but tend to favor one breed throughout the genome, we could see large differences in predicted Hereford and Shorthorn proportions genome-wide when, in fact, this is an artifact, and the true differences genome-wide are small. This would also artificially impact the correlations between taurine and indicine haplotype numbers in Beefmaster.

However, the results in **Figures 4B**, **Figures 5B**, **Figures 6B** are both encouraging and interesting. In all three figures, the animals with the fewest numbers of haplotypes within their genomes have by far the lowest Brahman genome content. We suspect that these animals are not early generation Brangus, Santa Gertrudis, or Beefmaster animals but are intermediary crosses (1/8 × 7/8 and ¼ × ¾) produced by cattle breeders generating first generation members of each of the breeds who decided to genotype these animals and provide them to their respective Breed Associations (see e.g., https://gobrangus.com/tperkins-breeding-up/). Despite this, the two interesting features of these plots are that the proportions of Brahman within all three of these breeds are far less than would be expected based upon the mating schemes used to produce the animals, and the proportion of Brahman within the genomes of advanced generation animals is consistently increasing supporting our results from the direct comparison of early- and advanced-generation animals. Our interpretation of these results are that breeders of American Breed cattle select very strongly for performance and type (coat color, polled, sheath score, performance) characteristics, particularly in early generation animals, resulting in a selective advantage for taurine alleles which creates widespread linkage drag throughout the genome dramatically reducing the Brahman content. However, in advanced-generation cattle, selection emphases change towards adaptive alleles (nutrient utilization, temperature tolerance, pathogen resistance) favoring indicine alleles. In the Beefmaster, the increase in Brahman genome content in advanced-generation animals appears to be primarily at the expense of the Shorthorn content (**Supplementary Figure S5**).

Goszczynski et al. (2018) used STRUCTURE to estimate the Brahman proportion within the genomes of 100 Argentinean registered Brangus animals to be 34.7%, which is very close to the breed expectation of 0.375 based upon pedigree. We suspect that the difference between this result and our finding reflects a very different selection history within the U. S. and Argentinian registered Brangus populations. Paim et al. (2020a) used ADMIXTURE applied to genotypes for 59 registered U.S Brangus cattle and estimated the Brahman proportion of their genomes to be 29.6%, on average, slightly larger than found in this study, but below breed expectation. Their estimates of breed proportion by chromosomal location show considerable similarity to our results in **Figures 1–3**, and they also found that chromosomes 5 and 15 had the lowest and highest Angus contents, respectively (**Table 3**), despite a much smaller sample size (Paim et al., 2020a).

Genomic Divergence From Breed Expectation

To test genomic windows for divergence from breed expectations, we conducted a simulation to estimate the variance in the proportion of Brahman genome content that would be expected across loci in a large population of advanced generation individuals. The estimate of variance from the simulation was quite small ($\sqrt{\widehat{Var}(\theta_i)} \approx 0.01$) and because the overall Brahman genome content of individuals from each of the breeds was considerably smaller than that expected under neutrality, we found that very large proportions (>85%) of the genomes of each of the breeds were diverged from breed expectations. However, these values are very similar to the finding of Decker et al. (2012), who estimated that 82.41% of the genome of registered American Angus cattle was exposed to strong selection. The values of θ and $\mathrm{Var}(\theta_i)$ used to calculate the Z-statistic are important for determining the number of performed tests that are deemed significant, however, the produced Z-statistics rank identically to the sample Brahman proportion estimates p_i regardless of the values of θ and $\mathrm{Var}(\theta_i)$ that are used. Our simulations are probably not representative of the generational mix within the samples we received from each Breed Association and the sample size simulated may not be representative of the number of animals used to generate each generation and it is possible that $\mathrm{Var}(\theta_i)$ is considerably higher than the value we used. This would result in a smaller proportion of the genomes of each breed being found to have diverged from breed expectation. For example, if we had used $\sqrt{\widehat{\mathrm{Var}}(\theta_i)} = 0.02$, 0.03, 0.04, or 0.05 (almost a 5-fold increase in standard deviation and 25 fold increase in variance) the statistical threshold for a FDR <0.001 would have been 3.16, 3.22, 3.32, and 3.45 in Brangus and we would have estimated that 69.5, 60.17, 47.3, or 35.6% of the Brangus genome was diverged from breed expectation, respectively. Whatever value of $\sqrt{\mathrm{Var}(\theta_i)}$ is used, the most extreme test statistics are concordant with the most diverged genomic regions.

Figure 7 reveals that the patterns of divergence are complex but quite similar among the breeds for several chromosomes. The entirety of chromosome 25 possesses the least divergence, while chromosomes 1, 5, 11, 12, and 18 possess regions of very high divergence for all three breeds - characteristic of loci that have been strongly selected. Entire chromosomes are highly diverged for their

Brahman composition including 15 in Brangus, 3 and 4 in Santa Gertrudis, and 14, 15, and 20 in Beefmaster. Chromosomes 8, 9, 13, 17, 20, 22, 24, and 29 reveal loci that have been exposed to selection in only a subset of the breeds. Selection for favorable alleles at multiple polygenes of large effect on a chromosome will generate a complex signature of selection that is dependent on initial phase relationships among the selected alleles and the relative intensities of selection applied to each of the alleles. Consequently, it is not likely to be a fruitful exercise to attempt to identity the underlying selected loci and phenotypes except perhaps for the largest effect loci and for loci that are well known to have been exposed to selection in these breeds.

Goszczynski et al. (2018) found a significant enrichment for Brahman alleles within the BoLA region of chromosome 23 (chr23:28,720,113–28,724,739), however, we found no evidence for this (**Figure 7**) and once again speculate that this reflects a different selection history for the U.S and Argentinian populations.

Coat Color and Polled Loci

Breed Characteristics

Brangus cattle are required to be polled (absence of horns) and to have solid black or red coat colors. Santa Gertrudis cattle may be horned or polled, with a light or dark solid red coat color preferred; roan (red coat with white patches) or white outside the underline or other spotting in other parts of the body disqualifies an animal from registration. Beefmaster cattle may be horned or polled and while brownish-red is the most common color, the breed has no color standards. Brahman cattle are primarily horned and have complex coat colors that can range from solid gray to brindled. Angus are polled due to the autosomal dominant *Celtic polled* allele, a complex structural insertion, at *Polled* located near the centromere of chromosome 1 (Brenneman et al., 1996; Medugorac et al., 2012) and have either solid black (American Angus Association) or red (American Red Angus Association) coat colors due to variation within *MC1R* located at 14,704,918–14,707,018 on chromosome 18. American Shorthorn can be horned or polled and can have red, white, or roan coat color patterns caused by variation at *KITLG* (Seitz et al., 1999) located at 18,245,986–18,317,616 on chromosome 5. American Hereford cattle can be horned or polled and have dark red to red-yellow coat colors. They are piebald with white on the head, ventral areas, lower legs, and tail switch which is inherited as an autosomal dominant due to structural variation located proximal to *KIT* at 70,157,944–70,262,786 Mb on chromosome 6 (Grosz and MacNeil, 1999; Whitacre, 2014). QTL associated with variation in white spotting in cattle have also been identified on chromosome 6 near *KIT* (Reinsch et al., 1999; Liu et al., 2009).

Selection Signatures at Coat Color and Polled Loci

Figure 7 confirms the selection of Angus alleles (**Figure 1**) at *Polled* and *MC1R* in Brangus (windows with strongest signal Chr1:376,390–4,458,855; p_i = 0.02, -$\log_{10}p$ = 240.08; Chr18: 13,773,453–13,896,455; p_i = 0.04, $-\log_{10}p$ = 212.80) and Shorthorn alleles (**Figure 2**) at *MC1R* in Santa Gertrudis (Chr18:17,876,786–17,944,337; p_i = 0.03, $-\log_{10}p$ = 232.08), respectively. These windows were among the 5 most differentiated windows for Brahman genome content in both breeds (**Table 4**). Paim et al. (2020b) found evidence of strong enrichment of Angus alleles at *Polled* in 59 Brangus animals, but not at *MC1R*. Ancestral breed frequency signatures on chromosome 5 are impacted by a very strongly selected locus enriched in frequency for Brahman alleles in all three of the American Breeds towards the center of the chromosome (**Figures 1**–**3**). Despite this, **Figures 2**, **3** show a reduced proportion of Shorthorn ancestry in Beefmaster, but an increased proportion of Shorthorn ancestry in Santa Gertrudis animals in the vicinity of *KITLG* on chromosome 5 presumably due to selection for the allele conferring red coat color and against that conferring white in Santa Gertrudis. Regions on chromosome 6 near *KIT* were strongly enriched for Angus alleles in Brangus (Chr6:70,550,407–70,635,750; p_i = 0.14, $-\log_{10}p$ = 107.30), Brahman alleles in Santa Gertrudis (Chr6:71,137,845–71,233,574; p_i = 0.51, $-\log_{10}p$ = 38.49), and Brahman alleles in Beefmaster (Chr6:68,154,476–68,192,608; p_i = 0.66, $-\log_{10}p$ = 47.03). This suggests that selection has occurred for solid coat color patterns in all three breeds for both aesthetic and economic reasons due to losses caused by ocular squamous cell carcinoma, which is most common in Hereford animals that lack pigment around their eyes (Heeney and Valli, 1985).

Large Effect Locus Enriched for Ancestral Brahman Alleles

As the global proportion of indicine ancestry increases in *B. t. taurus* × *B. t. indicus* hybrids, coat color lightens, males tend to have more pendulous sheaths, body weight and condition scores increase, and tick and worm burdens decrease (Porto-Neto et al., 2014). *B. t. indicus* cattle tend to have a lower performance than *B. t. taurus* cattle under favorable conditions, but significantly outperform *B. t. taurus* cattle in extreme climates and environments where parasites, heat, and low inputs play important roles in the production system (Frisch and Vercoe, 1977; Frisch and Vercoe, 1984; Menjo et al., 2009; Porto-Neto et al., 2014). In almost all cases, the most significantly diverged regions are enriched for taurine alleles (**Figures 1**–**3** and **Figure 7**). However, Brahman alleles are strongly enriched in frequency in all three of the American breeds in the central region of chromosome 5 (**Figures 1**–**3**). The window with the greatest Brahman composition was Chr5: 48,183,014–48,275,554; p_i = 0.59, $-\log_{10}p$ = 96.10 in Brangus, Chr5:42,298,343–42,482,771; p_i = 0.79, $-\log_{10}p$ = >300 in Santa Gertrudis, and Chr5:48,090,162–48,159,989; p_i = 0.84, $-\log_{10}p$ = 212.56) in Beefmaster. The window detected in Santa Gertrudis was almost 6 Mb from those found in Brangus and Beefmaster, but **Figure 2** shows strong enrichment for Brahman alleles across the majority of chromosome 5 in Santa Gertrudis suggesting that selection may be acting on several loci on this chromosome in this breed. Querying the concatenated region Chr5:42,298,343–48,275,554 in the CattleQTLdb retrieved 458 QTL or association records for 87 different traits. Among these, loci responsible for variation in traits that are known to be under strong selection in beef

cattle that have previously been reported in studies of indicine or indicine × taurine crossbred cattle include Birth Weight (Peters et al., 2012; Saatchi et al., 2014b), Male and Female Reproduction (Hawken et al., 2012; McDaneld et al., 2014; Buzanskas et al., 2017), Stature, Yearling Weight, Carcass Weight and *Longissimus Dorsi* Muscle Area (Imumorin et al., 2011; Pryce et al., 2011; Saatchi et al., 2014b), and intramuscular fatness measured as Marbling Score (Peters et al., 2012; Leal-Gutiérrez et al., 2019). Also within this region are QTL that have been found to be associated with Immune Function (Leach et al., 2010), Susceptibility to Bovine Respiratory disease (Neupane et al., 2018), and Tick Resistance (Gasparin et al., 2007).

The genes closest to the windows with the greatest Brahman ancestry in Brangus and Beefmaster are *HMGA2* (Chr5: 47,819,475–47,966,336) and *MSRB3* (Chr5: 48,330,041–48,511,868). In human, *HMGA2* has been associated with 62 traits (https://www.ebi.ac.uk/gwas/genes/HMGA2; Accessed Sept. 18, 2021), including body height, birth weight, systolic blood pressure, brain cortical volume, intelligence, and insomnia. Brahman females are known to suppress the birth weight of their calves (Dillon et al., 2015) to increase calving ease and reduce calf mortality. *MSRB3* has been associated with 33 traits in human (https://www.ebi.ac.uk/gwas/genes/MSRB3; Accessed Sept. 18, 2021) including brain cortical volume, snoring, lung function, height, and temperament. The window with the greatest Brahman ancestry in Santa Gertrudis overlaps *CPNE8*, associated with 14 human traits (https://www.ebi.ac.uk/gwas/genes/CPNE8; Accessed Sept. 18, 2021) including chronotype and circadian rhythm, IgG glycosylation, heart rate, and bipolar disorder. Consequently, this broad genomic region includes genes which may explain several of the physiological differences between *B. t. taurus* and *B. t. indicus cattle*, including birth weight, height, blood vessel morphology, hair morphology, disease resistance, and temperament.

Large Effect Loci Enriched for Ancestral Taurine Alleles

We queried CattleQTLdb for the remaining 12 loci in **Table 4** to identify candidate QTL responsible for the enrichment of taurine alleles at these highly diverged loci and the results are in **Table 5**. We performed the query for 1 Mb regions centered on the window centers reported in **Table 4** and restricted our attention to loci responsible for variation in traits known to be under artificial selection or potentially under natural selection in U. S. beef breeds and discovered in taurine or *B. t. taurus* × *B. t. indicus* hybrid populations. We also collapsed database entries from the same author with separate QTL identities into a single QTL when they co-localized within the genome. For example, the query for chromosome 14 region retrieved 5 database entries for Birth Weight from Snelling et al. (2010) with QTL identifiers 68615, 68626, 68630, 68658, and 68672 with coordinates Chr14:23,893,200–23,893,240, Chr14: 23,946,416–23,946,456, Chr14:23,571,834–23,571,874, Chr14:23,853,791–23,853,831, and Chr14: 23,421,913–23,421,953, respectively. These clearly all represent the same Birth Weight QTL and were collapsed into a single entry in **Table 5**. However, this feature of the CattleQTLdb makes it difficult to perform enrichment analyses, since the numbers of trait associations detected in any study is dependent upon the number of markers used. The CattleQTLdb contains 161,781 QTL from 1,049 publications representing 680 different traits (https://www.animalgenome.org/cgi-bin/QTLdb/index; Accessed August 11, 2021). The bovine autosomal genome length is 2,489.39 Mb in the ARS-UCD1.2 bovine reference genome assembly and so if QTL are randomly located through the genome, we would expect to find an average of 64.98 QTL within each 1 Mb window. We retrieved an average of 61.25 ± 45.49 QTL or Association entries for 27.83 ± 12.07 traits within the 12 genomic windows in **Table 5** indicating that these genomic regions are typical for their QTL content.

Two features of **Table 5** are particularly interesting. First, 7 of the 12 differentiated regions harbor single reports of QTL for Tick or Disease Resistance. However, the CattleQTLdb contains 2,923 Disease, 210 Parasite, 254 Immune Capacity, and 124 Parasite Pest Resistance QTL or Association entries, prior to coalescence to remove redundant QTL reporting. These data suggest that we should have found, on average, 1.41 QTL or Association entries within the CattleQTLdb per 1 Mb autosomal region. From the Poisson distribution, the probability of 1 or more CattleQTLdb entries per region is 0.76 and from the Binomial distribution, the probability of 7 or more regions possessing CattleQTLdb entries is 0.95. The second feature is that 11 of the 12 differentiated regions harbor QTL for Birth Weight or Calving Ease for which 1,180 and 4,105 QTL or Association entries are reported, respectively, in the CattleQTLdb. This corresponds to an average of 2.12 database entries per 1 Mb window and a probability of 0.57 of 11 or more of the windows having CattleQTLdb entries for Birth Weight or Calving Ease. However, the distributions of database entries for Birth Weight and Calving Ease by chromosome are not random, with 511 (43.3%) entries for Birth Weight on chromosome 6 and 1,968 (47.9%) entries for Calving Ease on chromosome 21. Removing these outlier chromosomes from the analysis leads to the expectation of 1.22 database entries per 1 Mb window and a probability of only 0.09 of finding entries for Birth Weight or Calving Ease in at least 11 of 12 randomly sampled 1 Mb windows.

Ten of the regions in **Table 5** have CattleQTLdb entries for growth and weight traits and 7 of these are for Weaning Weight. We found 616 CattleQTLdb entries for Weaning Weight, again with an overrepresentation on chromosome 6 (146 or 24%). After removing chromosome 6 from the analysis, this corresponds to an average of only 0.20 database entries per 1 Mb window for the remainder of the autosomal genome and a probability of only 0.002 of finding at least 7 out of 12 randomly sampled 1 Mb genomic regions containing database entries for Weaning Weight. Four of the regions in **Table 5** have CattleQTLdb entries for Marbling Score for which there are 441 CattleQTLdb entries corresponding to an average of only 0.18 database entries per 1 Mb window for the autosomal

genome. This generates a probability of 0.12 of finding at least 4 out of 12 randomly sampled 1 Mb genomic regions containing database entries for Marbling Score. Finally, there are 19,572 entries in CattleQTLdb for Fertility indicating that we should have expected to find fertility trait QTL in all 12 of the genomic regions reported in **Table 5**.

These results suggest that artificial selection on production traits practiced by breeders of American Breed cattle is responsible for the surfeit of taurine alleles at the 12 loci in **Table 5**. Our analyses suggest that American Breed cattle breeders are placing the greatest selection emphases on growth, calving ease, and meat quality traits, which is consistent with the findings of Decker et al. (2012) in U.S registered Angus cattle.

Genomic Divergence Between Early- and Advanced-Generations

Although statistical power was limited when comparing the Brahman content of the genomes of early- and advanced-generation animals due to the use of only ~10% of the animals from each breed in each tail of the haplotype number distribution, these analyses clearly revealed that the Brahman proportion of the genome of these American breeds has increased with generation number and that this increase has occurred almost genome-wide.

There are several interesting features of **Supplementary Table S3**. First, we found two of the most differentiated 1 Mb regions between early- and advanced-generation American Breed animals to have no previously identified QTL within them. Considering an average of 64.98 QTL within each 1 Mb window for the CattleQTLdb, the probability of two such intervals in 15 randomly sampled 1 Mb genomic regions is small (p = 3.88E-55), indicating that the selected loci in these regions impact traits that have not been well studied in cattle, suggesting that they may have roles in environmental adaptation. Second, the locus on chromosome 6 at 71.5 Mb in Beefmaster supports our conjecture that Beefmaster breeders have selected for solid coat colors in advanced-generation animals, despite the fact that the breed does not have coat color standards. Finally, if we remove the locus on chromosome 6 at 71.5 Mb in Beefmaster, 11 of the 12 genomic regions with identified QTL harbor QTL that influence heat tolerance or animal health, immune response, parasite burden or disease susceptibility. The probability of 11 or all 12 regions harboring disease or parasite CattleQTLdb entries reduces to 0.18. These results suggest that these genomic regions have been selected to increase the proportion of Brahman alleles because they confer an adaptive advantage in American Breed cattle.

CONCLUSION

The American breeds are advanced generation *B. t. taurus* × *B. t. indicus* hybrid cattle that were developed to capitalize on breed complementarity and heterosis to produce cattle that were suited to harsh, subtropical climates as well as disease and parasite threats, while still maintaining acceptable levels of growth and productivity. The breeds employed mating systems designed to produce cattle that were either ⅝ taurine and ⅜ indicine (Brangus and Santa Gertrudis) or ½ taurine and ½ indicine (Beefmaster). However, we found strong evidence that selection for polledness, coat color phenotypes, growth, calving ease, and intra-muscular fat content produced early-generation cattle with much smaller than expected indicine composition within the genomes of all three breeds. At least 85% of the genomes within each of the breeds possess less Brahman ancestry than expected in the absence of selection and drift. We utilized the total number of haplotypes predicted by RFMix in the diploid genome of each animal as a proxy for the generation number for each animal and when we ranked animals based on this proxy, we detected an increase in the Brahman genome content in advanced-generation cattle of all three breeds. By comparing the genomes of early- and advanced-generation animals within each breed, we found evidence for strong selection for indicine alleles which likely confer an adaptive advantage for heat tolerance and healthfulness in advance-generation animals.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study belong to the respective Breed Associations. For access to the data, contact: Darrell Wilkes, Executive Vice President, International Brangus Breeders Association: dwilkes@gobrangus.com. Collin Osbourn, Executive Vice President, Beefmaster Breeders United: cosbourn@beefmasters.org, Webb Fields, Executive Director, Santa Gertrudis Breeders International: wfields@santagertrudis.com.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because all animals for which samples were obtained for genotyping in this study were either sampled via cryopreserved semen or tissue samples (e.g., hair, blood) obtained by the owners of the cattle according to standard industry practices. Written informed consent for participation was not obtained from the owners because genotype data are owned and were provided to us by the three Breed Associations.

AUTHOR CONTRIBUTIONS

All authors contributed to the overall design and approaches used to conduct the research and read and edited the final manuscript. TEC and JFT conducted the analyses and wrote the manuscript. RDS managed the genotypes and prepared files for the analyses. JED oversaw the development of the genotype imputation pipeline and generated the imputed data set.

FUNDING

JT was supported by grant number 2017-67015-26760 from the United States Department for Agriculture National Institute for Food and Agriculture.

ACKNOWLEDGMENTS

The authors appreciate data sharing by the International Brangus Breeders Association, Beefmaster Breeders United, and Santa Gertrudis Breeders International Breed Associations. We also appreciate the efforts of two reviewers which prompted us to more formally examine the increasing Brahman content within advanced generation animals.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.758394/full#supplementary-material

REFERENCES

Abo-Ismail, M. K., Lansink, N., Akanno, E., Karisa, B. K., Crowley, J. J., Moore, S. S., et al. (2018). Development and Validation of a Small SNP Panel for Feed Efficiency in Beef Cattle1. *J. Anim. Sci.* 96, 375–397. doi:10.1093/jas/sky020

Akanno, E. C., Chen, L., Abo-Ismail, M. K., Crowley, J. J., Wang, Z., Li, C., et al. (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

Ashwell, M. S., Heyen, D. W., Weller, J. I., Ron, M., Sonstegard, T. S., Van Tassell, C. P., et al. (2005). Detection of Quantitative Trait Loci Influencing Conformation Traits and Calving Ease in Holstein-Friesian Cattle. *J. Dairy Sci.* 88, 4111–4119. doi:10.3168/jds.s0022-0302(05)73095-2

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bovine HapMap ConsortiumGibbs, R. A., Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324, 528–532. doi:10.1126/science.1167936

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., et al. (2003). Detection of Genes Influencing Economic Traits in Three French Dairy Cattle Breeds. *Genet. Sel. Evol.* 35, 77–101. doi:10.1186/1297-9686-35-1-77

Bolormaa, S., Neto, L. R. P., Zhang, Y. D., Bunch, R. J., Harrison, B. E., Goddard, M. E., et al. (2011). A Genome-wide Association Study of Meat and Carcass Traits in Australian Cattle1. *J. Anim. Sci.* 89, 2297–2309. doi:10.2527/jas.2010-3138

Brenneman, R. A., Davis, S. K., Sanders, J. O., Burns, B. M., Wheeler, T. C., Turner, J. W., et al. (1996). The Polled Locus Maps to BTA1 in a *Bos indicus* X *Bos taurus* Cross. *J. Hered.* 87, 156–161. doi:10.1093/oxfordjournals.jhered.a022975

Browning, S. R., Grinde, K., Plantinga, A., Gogarten, S. M., Stilp, A. M., Kaplan, R. C., et al. (2016). Local Ancestry Inference in a Large US-Based Hispanic/Latino Study: Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *G* 6, 1525–1534. doi:10.1534/g3.116.028779

Brunes, L. C., Baldi, F., Lopes, F. B., Lôbo, R. B., Espigolan, R., Costa, M. F. O., et al. (2021). Weighted Single-step Genome-wide Association Study and Pathway Analyses for Feed Efficiency Traits in Nellore Cattle. *J. Anim. Breed. Genet.* 138, 23–44. doi:10.1111/jbg.12496

Buzanskas, M. E., Grossi, D. D. A., Ventura, R. V., Schenkel, F. S., Chud, T. C. S., Stafuzza, N. B., et al. (2017). Candidate Genes for Male and Female Reproductive Traits in Canchim Beef Cattle. *J. Anim. Sci Biotechnol* 8, 67. doi:10.1186/s40104-017-0199-8

Cai, Z., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2019). Prioritizing Candidate Genes for Fertility in Dairy Cows Using Gene-Based Analysis, Functional Annotation and Differential Gene Expression. *BMC Genomics* 20, 255. doi:10.1186/s12864-019-5638-9

Carignano, H. A., Roldan, D. L., Beribe, M. J., Raschia, M. A., Amadio, A., Nani, J. P., et al. (2018). Genome-wide Scan for Commons SNPs Affecting Bovine Leukemia Virus Infection Level in Dairy Cattle. *BMC Genomics* 19, 142. doi:10.1186/s12864-018-4523-2

Cartwright, T. C. (1970). Selection Criteria for Beef Cattle for the Future. *J. Anim. Sci.* 30, 706–711. doi:10.2527/jas1970.305706x

Casas, E., Hessman, B. E., Keele, J. W., and Ridpath, J. F. (2015). A Genome-wide Association Study for the Incidence of Persistent Bovine Viral Diarrhea Virus Infection in Cattle. *Anim. Genet.* 46, 8–15. doi:10.1111/age.12239

Casas, E., Shackelford, S. D., Keele, J. W., Koohmaraie, M., Smith, T. P. L., and Stone, R. T. (2003). Detection of quantitative trait loci for growth and carcass composition in cattle. *J. Anim. Sci.* 81, 2976–2983.

Casas, E., Shackelford, S. D., Keele, J. W., Stone, R. T., Kappes, S. M., and Koohmaraie, M. (2000). Quantitative Trait Loci Affecting Growth and Carcass Composition of Cattle Segregating Alternate Forms of Myostatin. *J. Anim. Sci.* 78, 560–569. doi:10.2527/2000.783560x

Casas, E., and Snowder, G. D. (2008). A Putative Quantitative Trait Locus on Chromosome 20 Associated with Bovine Pathogenic Disease Incidence1,2. *J. Anim. Sci.* 86, 2455–2460. doi:10.2527/jas.2008-0933

Casas, E., and Stone, R. T. (2006). Putative Quantitative Trait Loci Associated with the Probability of Contracting Infectious Bovine Keratoconjunctivitis1,2. *J. Anim. Sci.* 84, 3180–3184. doi:10.2527/jas.2006-200

Chan, E. K. F., Nagaraj, S. H., and Reverter, A. (2010). The Evolution of Tropical Adaptation: Comparing Taurine and Zebu Cattle. *Anim. Genet.* 41, 467–477. doi:10.1111/j.1365-2052.2010.02053.x

Cobanoglu, O., Berger, P. J., and Kirkpatrick, B. W. (2005). Genome Screen for Twinning Rate QTL in Four North American Holstein Families. *Anim. Genet.* 36, 303–308. doi:10.1111/j.1365-2052.2005.01299.x

Cochran, S. D., Cole, J. B., Null, D. J., and Hansen, P. J. (2013). Discovery of Single Nucleotide Polymorphisms in Candidate Genes Associated with Fertility and Production Traits in Holstein Cattle. *BMC Genet.* 14, 49. doi:10.1186/1471-2156-14-49

Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Jr, Crooker, B. A., et al. (2011). Genome-wide Association Analysis of Thirty One Production, Health, Reproduction and Body Conformation Traits in Contemporary U.S. Holstein Cows. *BMC Genomics* 12, 408. doi:10.1186/1471-2164-12-408

Coppieters, W., Mes, T. H., Druet, T., Farnir, F., Tamma, N., Schrooten, C., et al. (2009). Mapping QTL Influencing Gastrointestinal Nematode burden in Dutch Holstein-Friesian Dairy Cattle. *BMC Genomics* 10, 96. doi:10.1186/1471-2164-10-96

Crispim, A. C., Kelly, M. J., Guimarães, S. E. F., e Silva, F. F., Fortes, M. R. S., Wenceslau, R. R., et al. (2015). Multi-trait GWAS and New Candidate Genes Annotation for Growth Curve Parameters in Brahman Cattle. *PLoS One* 10, e0139906. doi:10.1371/journal.pone.0139906

Crum, T. E., Schnabel, R. D., Decker, J. E., Regitano, L. C. A., and Taylor, J. F. (2019). CRUMBLER: A Tool for the Prediction of Ancestry in Cattle. *PLoS One* 14, e0221471. doi:10.1371/journal.pone.0221471

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation Genotype Imputation Service and Methods. *Nat. Genet.* 48, 1284–1287. doi:10.1038/ng.3656

de Oliveira, P. S., Cesar, A. S., do Nascimento, M. L., Chaves, A. S., Tizioto, P. C., Tullio, R. R., et al. (2014). Identification of Genomic Regions Associated with Feed Efficiency in Nelore Cattle. *BMC Genet.* 15, 100. doi:10.1186/s12863-014-0100-0

Decker, J. E., Vasco, D. A., McKay, S. D., McClure, M. C., Rolf, M. M., Kim, J., et al. (2012). A Novel Analytical Method, Birth Date Selection Mapping, Detects

Response of the Angus (*Bos taurus*) Genome to Selection on Complex Traits. *BMC Genomics* 13, 606. doi:10.1186/1471-2164-13-606

Dickerson, G. (1970). Efficiency of Animal Production-Molding the Biological Components. *J. Anim. Sci.* 30, 849–859. doi:10.2527/jas1970.306849x

Dikmen, S., Cole, J. B., Null, D. J., and Hansen, P. J. (2013). Genome-wide Association Mapping for Identification of Quantitative Trait Loci for Rectal Temperature during Heat Stress in Holstein Cattle. *PLoS One* 8, e69202. doi:10.1371/journal.pone.0069202

Dillon, J. A., Riley, D. G., Herring, A. D., Sanders, J. O., and Thallman, R. M. (2015). Genetic Effects on Birth Weight in Reciprocal Brahman-Simmental Crossbred Calves1. *J. Anim. Sci.* 93, 553–561. doi:10.2527/jas.2014-8525

Doran, A. G., Berry, D. P., and Creevey, C. J. (2014). Whole Genome Association Study Identifies Regions of the Bovine Genome and Biological Pathways Involved in Carcass Trait Performance in Holstein-Friesian Cattle. *BMC Genomics* 15, 837. doi:10.1186/1471-2164-15-837

Druet, T., Fritz, S., Sellem, E., Basso, B., Gérard, O., Salas-Cortes, L., et al. (2009). Estimation of Genetic Parameters and Genome Scan for 15 Semen Characteristics Traits of Holstein Bulls. *J. Anim. Breed. Genet.* 126, 269–277. doi:10.1111/j.1439-0388.2008.00788.x

Duarte, D. A. S., Newbold, C. J., Detmann, E., Silva, F. F., Freitas, P. H. F., Veroneze, R., et al. (2019). Genome-wide Association Studies Pathway-based Meta-analysis for Residual Feed Intake in Beef Cattle. *Anim. Genet.* 50, 150–153. doi:10.1111/age.12761

Elo, K. T., Vilkki, J., Koning, D.-J. d., Velmala, R. J., and Mäki-Tanila, A. V. (1999). A Quantitative Trait Locus for Live Weight Maps to Bovine Chromosome 23. *Mamm. Genome* 10, 831–835. doi:10.1007/s003359901098

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R. P., Fernandes Júnior, G. A., Gordo, D. G. M., et al. (2015). Associations between Single Nucleotide Polymorphisms and Carcass Traits in Nellore Cattle Using High-Density Panels. *Genet. Mol. Res.* 14, 11133–11144. doi:10.4238/2015.september.22.7

Fortes, M. R. S., Reverter, A., Kelly, M., McCulloch, R., and Lehnert, S. A. (2013). Genome-wide Association Study for Inhibin, Luteinizing Hormone, Insulin-like Growth Factor 1, Testicular Size and Semen Traits in Bovine Species. *Andrology* 1, 644–650. doi:10.1111/j.2047-2927.2013.00101.x

Frisch, J. E., and Vercoe, J. E. (1977). Food Intake, Eating Rate, Weight Gains, Metabolic Rate and Efficiency of Feed Utilization in *Bos taurus* and *Bos indicus* Crossbred Cattle. *Anim. Sci.* 25, 343–358. doi:10.1017/s0003356100016755

Frisch, J. E., and Vercoe, T. E. (1984). An Analysis of Growth of Different Cattle Genotypes Reared in Different Environments. *J. Agric. Sci.* 103, 137–153. doi:10.1017/s0021859600043409

Gaddis, K. L., Null, D. J., and Cole, J. B. (2016). Explorations in Genome-wide Association Studies and Network Analyses with Dairy Cattle Fertility Traits. *J. Dairy Sci.* 99, 6420–6435. doi:10.3168/jds.2015-10444

Galliou, J. M., Kiser, J. N., Oliver, K. F., Seabury, C. M., Moraes, J. G. N., Burns, G. W., et al. (2020). Identification of Loci and Pathways Associated with Heifer Conception Rate in U.S. Holsteins. *Genes* 11, 767. doi:10.3390/genes11070767

Garcia, M. D., Matukumalli, L., Wheeler, T. L., Shackelford, S. D., Smith, T. P. L., and Casas, E. (2010). Markers on Bovine Chromosome 20 Associated with Carcass Quality and Composition Traits and Incidence of Contracting Infectious Bovine Keratoconjunctivitis. *Anim. Biotechnol.* 21, 188–202. doi:10.1080/10495398.2010.495012

Gasparin, G., Miyata, M., Coutinho, L. L., Martinez, M. L., Teodoro, R. L., Furlong, J., et al. (2007). Mapping of Quantitative Trait Loci Controlling Tick [Riphicephalus (Boophilus) Microplus] Resistance on Bovine Chromosomes 5, 7 and 14. *Anim. Genet.* 38, 453–459. doi:10.1111/j.1365-2052.2007.01634.x

Gobena, M., Elzo, M. A., and Mateescu, R. G. (2018). Population Structure and Genomic Breed Composition in an Angus-Brahman Crossbred Cattle Population. *Front. Genet.* 9, 90. doi:10.3389/fgene.2018.00090

González-Ruiz, S., Strillacci, M. G., Durán-Aguilar, M., Cantó-Alarcón, G. J., Herrera-Rodríguez, S. 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* 9, 636. doi:10.3390/ani9090636

Goszczynski, D. E., Corbi-Botto, C. M., Durand, H. M., Rogberg-Muñoz, A., Munilla, S., Peral-Garcia, P., et al. (2018). Evidence of Positive Selection towards Zebuine Haplotypes in the BoLA Region of Brangus Cattle. *Animal* 12, 215–223. doi:10.1017/s1751731117001380

Grosz, M. D., and MacNeil, M. D. (1999). Brief Communication. The 'spotted' Locus Maps to Bovine Chromosome 6 in Hereford-cross Population. *J. Hered.* 90, 233–236. doi:10.1093/jhered/90.1.233

Gutierrez-Gil, B., Ball, N., Burton, D., Haskell, M., Williams, J. L., and Wiener, P. (2008). Identification of Quantitative Trait Loci Affecting Cattle Temperament. *J. Hered.* 99, 629–638. doi:10.1093/jhered/esn060

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of Quantitative Trait Loci Controlling Trypanotolerance in a Cross of Tolerant West African N'Dama and Susceptible East African Boran Cattle. *Proc. Natl. Acad. Sci.* 100, 7443–7448. doi:10.1073/pnas.1232392100

Hawken, R. J., Zhang, Y. D., Fortes, M. R. S., Collis, E., Barris, W. C., Corbet, N. J., et al. (2012). Genome-wide Association Studies of Female Reproduction in Tropically Adapted Beef Cattle1. *J. Anim. Sci.* 90, 1398–1410. doi:10.2527/jas.2011-4410

Heeney, J. L., and Valli, V. E. (1985). Bovine Ocular Squamous Cell Carcinoma: an Epidemiological Perspective. *Can. J. Comp. Med.* 49, 21–26.

Hiendleder, S., Thomsen, H., Reinsch, N., Bennewitz, J., Leythe-Horn, B., Looft, C., et al. (2003). Mapping of QTL for Body Conformation and Behavior in Cattle. *J. Hered.* 94, 496–506. doi:10.1093/jhered/esg090

Hiltpold, M., Niu, G., Kadri, N. K., Crysnanto, D., Fang, Z.-H., Spengeler, M., et al. (2020). Activation of Cryptic Splicing in Bovine WDR19 Is Associated with Reduced Semen Quality and Male Fertility. *Plos Genet.* 16, e1008804. doi:10.1371/journal.pgen.1008804

Höglund, J. K., Buitenhuis, A. J., Guldbrandtsen, B., Su, G., Thomsen, B., and Lund, M. S. (2009a). Overlapping Chromosomal Regions for Fertility Traits and Production Traits in the Danish Holstein Population. *J. Dairy Sci.* 92, 5712–5719. doi:10.3168/jds.2008-1964

Höglund, J. K., Buitenhuis, B., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015). Genome-wide Association Study for Female Fertility in Nordic Red Cattle. *BMC Genet.* 16, 110. doi:10.1186/s12863-015-0269-x

Höglund, J. K., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2012). Analyzes of Genome-wide Association Follow-Up Study for Calving Traits in Dairy Cattle. *BMC Genet.* 13, 71. doi:10.1186/1471-2156-13-71

Höglund, J. K., Guldbrandtsen, B., Su, G., Thomsen, B., and Lund, M. S. (2009b). Genome Scan Detects Quantitative Trait Loci Affecting Female Fertility Traits in Danish and Swedish Holstein Cattle. *J. Dairy Sci.* 92, 2136–2143. doi:10.3168/jds.2008-1104

Holmberg, M., and Andersson-Eklund, L. (2006). Quantitative Trait Loci Affecting Fertility and Calving Traits in Swedish Dairy Cattle. *J. Dairy Sci.* 89, 3664–3671. doi:10.3168/jds.s0022-0302(06)72406-7

Howard, J. T., Kachman, S. D., Snelling, W. M., Pollak, E. J., Ciobanu, D. C., Kuehn, L. A., et al. (2014). Beef Cattle Body Temperature during Climatic Stress: a Genome-wide Association Study. *Int. J. Biometeorol.* 58, 1665–1672. doi:10.1007/s00484-013-0773-5

Hu, Z.-L., Park, C. A., and Reecy, J. M. (2019). Building a Livestock Genetic and Genomic Information Knowledgebase through Integrative Developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–D710. doi:10.1093/nar/gky1084

Imumorin, I. G., Kim, E.-H., Lee, Y.-M., De Koning, D.-J., van Arendonk, J. A., De Donato, M., et al. (2011). Genome Scan for Parent-Of-Origin QTL Effects on Bovine Growth and Carcass Traits. *Front. Gene* 2, 44. doi:10.3389/fgene.2011.00044

Irano, N., de Camargo, G. M. F., Costa, R. B., Terakado, A. P. N., Magalhães, A. F. B., Silva, R. M. d. O., et al. (2016). Genome-wide Association Study for Indicator Traits of Sexual Precocity in Nellore Cattle. *PLoS One* 11, e0159502. doi:10.1371/journal.pone.0159502

Jivanji, S., Worth, G., Lopdell, T. J., Yeates, A., Couldrey, C., Reynolds, E., et al. (2019). Genome-wide Association Analysis Reveals QTL and Candidate Mutations Involved in white Spotting in Cattle. *Genet. Sel. Evol.* 51, 62. doi:10.1186/s12711-019-0506-2

Júnior, G. A. F., Costa, R. B., de Camargo, G. M. F., Carvalheiro, R., Rosa, G. J. M., Baldi, F., et al. (2016). Genome Scan for Postmortem Carcass Traits in Nellore Cattle1. *J. Anim. Sci.* 94, 4087–4095. doi:10.2527/jas.2016-0632

Karisa, B. K., Thomson, J., Wang, Z., Stothard, P., Moore, S. S., and Plastow, G. S. (2013). Candidate Genes and Single Nucleotide Polymorphisms Associated with Variation in Residual Feed Intake in Beef Cattle1. *J. Anim. Sci.* 91, 3502–3513. doi:10.2527/jas.2012-6170

Kaupe, B., Brandt, H., Prinzenberg, E.-M., and Erhardt, G. (2007). Joint Analysis of the Influence of CYP11B1 and DGAT1 Genetic Variation on Milk Production, Somatic Cell Score, Conformation, Reproduction, and Productive Lifespan in German Holstein Cattle1. *J. Anim. Sci.* 85, 11–21. doi:10.2527/jas.2005-753

Keele, J. W., Kuehn, L. A., McDaneld, T. G., Tait, R. G., Jones, S. A., Keel, B. N., et al. (2016). Genomewide Association Study of Liver Abscess in Beef Cattle1,2. *J. Anim. Sci.* 94, 490–499. doi:10.2527/jas.2015-9887

Kim, E.-S., Sonstegard, T. S., Silva, M. V. G. B., Gasbarre, L. C., and Van Tassell, C. P. (2013). Identification of Quantitative Trait Loci Affecting Gastrointestinal Parasite Resistance in an Experimental Angus Population. *Anim. Genet.* 45, 117–121. doi:10.1111/age.12101

Kim, J.-J., Farnir, F., Savell, J., and Taylor, J. F. (2003). Detection of Quantitative Trait Loci for Growth and Beef Carcass Fatness Traits in a Cross between *Bos taurus* (Angus) and *Bos indicus* (Brahman) Cattle1. *J. Anim. Sci.* 81, 1933–1942. doi:10.2527/2003.8181933x

Kiser, J. N., Clancey, E., Moraes, J. G. N., Dalton, J., Burns, G. W., Spencer, T. E., et al. (2019). Identification of Loci Associated with conception Rate in Primiparous Holstein Cows. *BMC Genomics* 20, 840. doi:10.1186/s12864-019-6203-2

Kneeland, J., Li, C., Basarab, J., Snelling, W. M., Benkel, B., Murdoch, B., et al. (2004). Identification and fine Mapping of Quantitative Trait Loci for Growth Traits on Bovine Chromosomes 2, 6, 14, 19, 21, and 23 within One Commercial Line of *Bos taurus*1. *J. Anim. Sci.* 82, 3405–3414. doi:10.2527/2004.82123405x

Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., et al. (2008). A Whole-Genome Scan to Map Quantitative Trait Loci for Conformation and Functional Traits in Canadian Holstein Bulls. *J. Dairy Sci.* 91, 2844–2856. doi:10.3168/jds.2007-0585

Kühn, C., Bennewitz, J., Reinsch, N., Xu, N., Thomsen, H., Looft, C., et al. (2003). Quantitative Trait Loci Mapping of Functional Traits in the German Holstein Cattle Population. *J. Dairy Sci.* 86, 360–368. doi:10.3168/jds.s0022-0302(03)73614-5

Küpper, J. D., Brandt, H. R., and Erhardt, G. (2014). Genetic Association betweenNOD2polymorphism and Infection Status byMycobacterium aviumssp.Paratuberculosisin German Holstein Cattle. *Anim. Genet.* 45, 114–116. doi:10.1111/age.12097

Leach, R. J., Craigmile, S. C., Knott, S. A., Williams, J. L., and Glass, E. J. (2010). Quantitative Trait Loci for Variation in Immune Response to a Foot-And-Mouth Disease Virus Peptide. *BMC Genet.* 11, 107. doi:10.1186/1471-2156-11-107

Leach, R. J., O'Neill, R. G., Fitzpatrick, J. L., Williams, J. L., and Glass, E. J. (2012). Quantitative Trait Loci Associated with the Immune Response to a Bovine Respiratory Syncytial Virus Vaccine. *PLoS One* 7, e33526. doi:10.1371/journal.pone.0033526

Leal-Gutiérrez, J. D., Elzo, M. A., Johnson, D. D., Hamblen, H., and Mateescu, R. G. (2019). Genome Wide Association and Gene Enrichment Analysis Reveal Membrane Anchoring and Structural Proteins Associated with Meat Quality in Beef. *BMC Genomics* 20, 151. doi:10.1186/s12864-019-5518-3

Li, G., Peñagaricano, F., Weigel, K. A., Zhang, Y., Rosa, G., and Khatib, H. (2012). Comparative Genomics between Fly, Mouse, and Cattle Identifies Genes Associated with Sire conception Rate. *J. Dairy Sci.* 95, 6122–6129. doi:10.3168/jds.2012-5591

Li, Y., Gao, Y., Kim, Y. S., Iqbal, A., and Kim, J. J. (2017). A Whole Genome Association Study to Detect Additive and Dominant Single Nucleotide Polymorphisms for Growth and Carcass Traits in Korean Native Cattle, Hanwoo. *Asian-australas J. Anim. Sci.* 30, 8–19. doi:10.5713/ajas.16.0170

Li, Y., and Kim, J.-J. (2015). Multiple Linkage Disequilibrium Mapping Methods to Validate Additive Quantitative Trait Loci in Korean Native Cattle (Hanwoo). *Asian Australas. J. Anim. Sci.* 28, 926–935. doi:10.5713/ajas.15.0077

Liu, A., Wang, Y., Sahana, G., Zhang, Q., Liu, L., Lund, M. S., et al. (2017). Genome-wide Association Studies for Female Fertility Traits in Chinese and Nordic Holsteins. *Sci. Rep.* 7, 8487. doi:10.1038/s41598-017-09170-9

Liu, L., Harris, B., Keehan, M., and Zhang, Y. (2009). Genome Scan for the Degree of white Spotting in Dairy Cattle. *Anim. Genet.* 40, 975–977. doi:10.1111/j.1365-2052.2009.01936.x

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., et al. (2016). Reference-based Phasing Using the Haplotype Reference Consortium Panel. *Nat. Genet.* 48, 1443–1448. doi:10.1038/ng.3679

Lu, D., Miller, S., Sargolzaei, M., Kelly, M., Vander Voort, G., Caldwell, T., et al. (2013). Genome-wide Association Analyses for Growth and Feed Efficiency Traits in Beef Cattle1. *J. Anim. Sci.* 91, 3612–3633. doi:10.2527/jas.2012-5716

Lund, M. S., Guldbrandtsen, B., Buitenhuis, A. J., Thomsen, B., and Bendixen, C. (2008). Detection of Quantitative Trait Loci in Danish Holstein Cattle Affecting Clinical Mastitis, Somatic Cell Score, Udder Conformation Traits, and Assessment of Associated Effects on Milk Yield. *J. Dairy Sci.* 91, 4028–4036. doi:10.3168/jds.2007-0290

Machado, M., S Azevedo, A., Teodoro, R. L., Pires, M. A., Cd Peixoto, M., de Freitas, C., et al. (2010). Genome Wide Scan for Quantitative Trait Loci Affecting Tick Resistance in Cattle (*Bos taurus* × *Bos indicus*). *BMC Genomics* 11, 280. doi:10.1186/1471-2164-11-280

MacNeil, M. D., and Grosz, M. D. (2002). Genome-wide Scans for QTL Affecting Carcass Traits in Hereford × Composite Double Backcross Populations. *J. Anim. Sci.* 80, 2316–2324. doi:10.2527/2002.8092316x

Maltecca, C., Weigel, K. A., Khatib, H., Cowan, M., and Bagnato, A. (2008). Whole-genome Scan for Quantitative Trait Loci Associated with Birth Weight, Gestation Length and Passive Immune Transfer in a Holstein X Jersey Crossbred Population. *Anim. Genet.* 40, 27–34. doi:10.1111/j.1365-2052.2008.01793.x

Maltecca, C., Gray, K. A., Weigel, K. A., Cassady, J. P., and Ashwell, M. (2011). A Genome-wide Association Study of Direct Gestation Length in US Holstein and Italian Brown Populations. *Anim. Genet.* 42, 585–591. doi:10.1111/j.1365-2052.2011.02188.x

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., et al. (2016). Genome-wide Association Study of Tick Resistance in South African Nguni Cattle. *Ticks Tick-borne Dis.* 7, 487–497. doi:10.1016/j.ttbdis.2016.02.005

Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix: a Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *Am. J. Hum. Genet.* 93, 278–288. doi:10.1016/j.ajhg.2013.06.020

Martínez, R., Gómez, Y., and Rocha, J. F. M. (2014). Genome-wide Association Study on Growth Traits in Colombian Creole Breeds and Crossbreeds with Zebu Cattle. *Genet. Mol. Res.* 13, 6420–6432. doi:10.4238/2014.august.25.5

McClure, M. C., Morsci, N. S., Schnabel, R. D., Kim, J. W., Yao, P., Rolf, M. M., et al. (2010). A Genome Scan for Quantitative Trait Loci Influencing Carcass, post-natal Growth and Reproductive Traits in Commercial Angus Cattle. *Anim. Genet.* 41, 597–607. doi:10.1111/j.1365-2052.2010.02063.x

McDaneld, T. G., Kuehn, L. A., Thomas, M. G., Snelling, W. M., Smith, T. P. L., Pollak, E. J., et al. (2014). Genomewide Association Study of Reproductive Efficiency in Female Cattle1,2,3,4. *J. Anim. Sci.* 92, 1945–1957. doi:10.2527/jas.2012-6807

Medugorac, I., Seichter, D., Graf, A., Russ, I., Blum, H., Göpel, K. H., et al. (2012). Bovine Polledness - an Autosomal Dominant Trait with Allelic Heterogeneity. *PLoS One* 7, e39477. doi:10.1371/journal.pone.0039477

Melo, T. P., Fortes, M. R. S., Fernandes Junior, G. A., Albuquerque, L. G., and Carvalheiro, R. (2019). RAPID COMMUNICATION: Multi-Breed Validation Study Unraveled Genomic Regions Associated with Puberty Traits Segregating across Tropically Adapted Breeds1. *J. Anim. Sci.* 97, 3027–3033. doi:10.1093/jas/skz121

Menjo, D. K., Bebe, B. O., Okeyo, A. M., and Ojango, J. M. K. (2009). Analysis of Early Survival of Holstein-Friesian Heifers of Diverse Sire Origins on Commercial Dairy Farms in Kenya. *Trop. Anim. Health Prod.* 41, 171–181. doi:10.1007/s11250-008-9172-z

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. doi:10.1111/age.12249

Michenet, A., Barbat, M., Saintilan, R., Venot, E., and Phocas, F. (2016). Detection of Quantitative Trait Loci for Maternal Traits Using High-Density Genotypes of Blonde d'Aquitaine Beef Cattle. *BMC Genet.* 17, 88. doi:10.1186/s12863-016-0397-y

Mizoshita, K., Watanabe, T., Hayashi, H., Kubota, C., Yamakuchi, H., Todoroki, J., et al. (2004). Quantitative Trait Loci Analysis for Growth and Carcass Traits in a Half-Sib Family of Purebred Japanese Black (Wagyu) Cattle1. *J. Anim. Sci.* 82, 3415–3420. doi:10.2527/2004.82123415x

Morris, C. A., Bottema, C. D. K., Cullen, N. G., Hickey, S. M., Esmailizadeh, A. K., Siebert, B. D., et al. (2010). Quantitative Trait Loci for Organ Weights and Adipose Fat Composition in Jersey and Limousin Back-Cross Cattle Finished on Pasture or Feedlot. *Anim. Genet.* 41, 589–596. doi:10.1111/j.1365-2052.2010.02058.x

Mota, R. R., Guimarães, S. E. F., Fortes, M. R. S., Hayes, B., Silva, F. F., Verardo, L. L., et al. (2017). Genome-wide Association Study and Annotating Candidate Gene Networks Affecting Age at First Calving in Nellore Cattle. *J. Anim. Breed. Genet.* 134, 484–492. doi:10.1111/jbg.12299

Naserkheil, M., Bahrami, A., Lee, D., and Mehrban, H. (2020). Integrating Single-step GWAS and Bipartite Networks Reconstruction Provides Novel Insights into Yearling Weight and Carcass Traits in Hanwoo Beef Cattle. *Animals* 10, 1836. doi:10.3390/ani10101836

Nayeri, S., Schenkel, F., Fleming, A., Kroezen, V., Sargolzaei, M., Baes, C., et al. (2019). Genome-wide Association Analysis for β-hydroxybutyrate Concentration in Milk in Holstein Dairy Cattle. *BMC Genet.* 20, 58. doi:10.1186/s12863-019-0761-9

Neupane, M., Kiser, J. N., and Neibergs, H. L. (2018). Gene Set Enrichment Analysis of SNP Data in Dairy and Beef Cattle with Bovine Respiratory Disease. *Anim. Genet.* 49, 527–538. doi:10.1111/age.12718

Oliveira, J. G. A., Perez, B. C., Cole, J. B., Santana, M. H. A., Silveira, J., Mazzoni, G., et al. (2017). Genomic Study and Medical Subject Headings Enrichment Analysis of Early Pregnancy Rate and Antral Follicle Numbers in Nelore Heifers. *J. Anim. Sci.* 95, 4796–4812. doi:10.2527/jas2017.1752

Oliveira, J. G. A., Santos, D. J. A., Cesar, A. S. M., Boison, S. A., Ventura, R. V., Perez, B. C., et al. (2020). Fine Mapping of Genomic Regions Associated with Female Fertility in Nellore Beef Cattle Based on Sequence Variants from Segregating Sires. *J. Anim. Sci. Biotechnol.* 10, 97. doi:10.1186/s40104-019-0403-0

Paim, T. D. P., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., Paiva, S. R., et al. (2020a). Dynamics of Genomic Architecture during Composite Breed Development in Cattle. *Anim. Genet.* 51, 224–234. doi:10.1111/age.12907

Paim, T. D. P., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., Paiva, S. R., et al. (2020b). Genomic Breed Composition of Selection Signatures in Brangus Beef Cattle. *Front. Genet.* 11, 710. doi:10.3389/fgene.2020.00710

Pausch, H., Flisikowski, K., Jung, S., Emmerling, R., Edel, C., Götz, K. U., et al. (2011). Genome-wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related Traits in Cattle. *Genetics* 187, 289–297. doi:10.1534/genetics.110.124057

Peters, S. O., Kizilkaya, K., Garrick, D. J., Fernando, R. L., Reecy, J. M., Weaber, R. L., et al. (2012). Bayesian Genome-wide Association Analysis of Growth and Yearling Ultrasound Measures of Carcass Traits in Brangus Heifers. *J. Anim. Sci.* 90, 3398–3409. doi:10.2527/jas.2011-4507

Porto-Neto, L. R., Reverter, A., Prayaga, K. C., Chan, E. K. F., Johnston, D. J., Hawken, R. J., et al. (2014). The Genetic Architecture of Climatic Adaptation of Tropical Cattle. *PLoS One* 9, e113284. doi:10.1371/journal.pone.0113284

Pryce, J. E., Hayes, B. J., Bolormaa, S., and Goddard, M. E. (2011). Polymorphic Regions Affecting Human Height Also Control Stature in Cattle. *Genetics* 187, 981–984. doi:10.1534/genetics.110.123943

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

Reinsch, N., Thomsen, H., Xu, N., Brink, M., Looft, C., Kalm, E., et al. (1999). A QTL for the Degree of Spotting in Cattle Shows Synteny with the KIT Locus on Chromosome 6. *J. Hered.* 90, 629–634. doi:10.1093/jhered/90.6.629

Rhoad, A. O. (1949). The Santa Gertrudis Breed: The Genesis and the Genetics of a New Breed of Beef Cattle. *J. Hered.* 40, 115–126. doi:10.1093/oxfordjournals.jhered.a106008

Richardson, I. W., Berry, D. P., Wiencko, H. L., Higgins, I. M., More, S. J., McClure, J., et al. (2016). A Genome-wide Association Study for Genetic Susceptibility to Mycobacterium Bovis Infection in Dairy Cattle Identifies a Susceptibility QTL on Chromosome 23. *Genet. Sel. Evol.* 48, 19. doi:10.1186/s12711-016-0197-x

Ring, S. C., Purfield, D. C., Good, M., Breslin, P., Ryan, E., Blom, A., et al. (2019). Variance Components for Bovine Tuberculosis Infection and Multi-Breed Genome-wide Association Analysis Using Imputed Whole Genome Sequence Data. *PLoS One* 14, e0212067. doi:10.1371/journal.pone.0212067

Rolf, M. M., Taylor, J. F., Schnabel, R. D., McKay, S. D., McClure, M. C., Northcutt, S. L., et al. (2012). Genome-wide Association Analysis for Feed Efficiency in Angus Cattle. *Anim. Genet.* 43, 367–374. doi:10.1111/j.1365-2052.2011.02273.x

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., et al. (2020). De Novo assembly of the Cattle Reference Genome with Single-Molecule Sequencing. *Gigascience* 9, giaa021. doi:10.1093/gigascience/giaa021

Rowan, T. N., Hoff, J. L., Crum, T. E., Taylor, J. F., Schnabel, R. D., and Decker, J. E. (2019). A Multi-Breed Reference Panel and Additional Rare Variants Maximize Imputation Accuracy in Cattle. *Genet. Sel. Evol.* 51, 77. doi:10.1186/s12711-019-0519-x

Saatchi, M., Beever, J. E., Decker, J. E., Faulkner, D. B., Freetly, H. C., Hansen, S. L., et al. (2014a). QTLs Associated with Dry Matter Intake, Metabolic Mid-test Weight, Growth and Feed Efficiency Have Little Overlap across 4 Beef Cattle Studies. *BMC Genomics* 15, 1004. doi:10.1186/1471-2164-15-1004

Saatchi, M., Schnabel, R. D., Taylor, J. F., and Garrick, D. J. (2014b). Large-effect Pleiotropic or Closely Linked QTL Segregate within and across Ten US Cattle Breeds. *BMC Genomics* 15, 442. doi:10.1186/1471-2164-15-442

Sahana, G., Guldbrandtsen, B., Bendixen, C., and Lund, M. S. (2010). Genome-wide Association Mapping for Female Fertility Traits in Danish and Swedish Holstein Cattle. *Anim. Genet.* 41, 579–588. doi:10.1111/j.1365-2052.2010.02064.x

Sahana, G., Guldbrandtsen, B., and Lund, M. S. (2011). Genome-wide Association Study for Calving Traits in Danish and Swedish Holstein Cattle. *J. Dairy Sci.* 94, 479–486. doi:10.3168/jds.2010-3381

Sanchez, M.-P., Guatteo, R., Davergne, A., Saout, J., Grohs, C., Deloche, M.-C., et al. (2020). Identification of the ABCC4, IER3, and CBFA2T2 Candidate Genes for Resistance to Paratuberculosis from Sequence-Based GWAS in Holstein and Normande Dairy Cattle. *Genet. Sel. Evol.* 52, 14. doi:10.1186/s12711-020-00535-9

Sanders, J. O. (1980). History and Development of Zebu Cattle in the United States. *J. Anim. Sci.* 50, 1188–1200. doi:10.2527/jas1980.5061188x

Schnabel, R. D., Sonstegard, T. S., Taylor, J. F., and Ashwell, M. S. (2005). Whole-genome Scan to Detect QTL for Milk Production, Conformation, Fertility and Functional Traits in Two US Holstein Families. *Anim. Genet.* 36, 408–416. doi:10.1111/j.1365-2052.2005.01337.x

Schrooten, C., Bink, M. C. A. M., and Bovenhuis, H. (2004). Whole Genome Scan to Detect Chromosomal Regions Affecting Multiple Traits in Dairy Cattle. *J. Dairy Sci.* 87, 3550–3560. doi:10.3168/jds.s0022-0302(04)73492-x

Schrooten, C., Bovenhuis, H., Coppieters, W., and Van Arendonk, J. A. M. (2000). Whole Genome Scan to Detect Quantitative Trait Loci for Conformation and Functional Traits in Dairy Cattle. *J. Dairy Sci.* 83, 795–806. doi:10.3168/jds.s0022-0302(00)74942-3

Schulman, N. F., Sahana, G., Lund, M. S., Viitala, S. M., and Vilkki, J. H. (2008). Quantitative Trait Loci for Fertility Traits in Finnish Ayrshire Cattle. *Genet. Sel. Evol.* 40, 195–214. doi:10.1051/gse:2007044

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide Association Study for Feed Efficiency and Growth Traits in U.S. Beef Cattle. *BMC Genomics* 18, 386. doi:10.1186/s12864-017-3754-y

Seidenspinner, T., Bennewitz, J., Reinhardt, F., and Thaller, G. (2009). Need for Sharp Phenotypes in QTL Detection for Calving Traits in Dairy Cattle. *J. Anim. Breed. Genet.* 126, 455–462. doi:10.1111/j.1439-0388.2009.00804.x

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

Sharma, A., Dang, C. G., Kim, K. S., Kim, J. J., Lee, H. K., Kim, H. C., et al. (2014). Validation of Genetic Polymorphisms on BTA14 Associated with Carcass Trait in a Commercial Hanwoo Population. *Anim. Genet.* 45, 863–867. doi:10.1111/age.12204

Sherman, E. L., Nkrumah, J. D., Li, C., Bartusiak, R., Murdoch, B., and Moore, S. S. (2009). Fine Mapping Quantitative Trait Loci for Feed Intake and Feed Efficiency in Beef Cattle1. *J. Anim. Sci.* 87, 37–45. doi:10.2527/jas.2008-0876

Sikora, K. M., Magee, D. A., Berkowicz, E. W., Berry, D. P., Howard, D. J., Mullen, M. P., et al. (2011). DNA Sequence Polymorphisms within the Bovine Guanine Nucleotide-Binding Protein Gs Subunit Alpha (Gsα)-Encoding (GNAS)

Genomic Imprinting Domain Are Associated with Performance Traits. *BMC Genet.* 12, 4. doi:10.1186/1471-2156-12-4

Smith, J. L., Wilson, M. L., Nilson, S. M., Rowan, T. N., Oldeschulte, D. L., Schnabel, R. D., et al. (2020). Genome-wide Association and Genotype by Environment Interactions for Growth Traits in U.S. Gelbvieh Cattle. *BMC Genomics* 20, 926. doi:10.1186/s12864-019-6231-y

Snelling, W. M., Allan, M. F., Keele, J. W., Kuehn, L. A., McDaneld, T., Smith, T. P. L., et al. (2010). Genome-wide Association Study of Growth in Crossbred Beef Cattle12. *J. Anim. Sci.* 88, 837–848. doi:10.2527/jas.2009-2257

Sollero, B. P., Junqueira, V. S., Gomes, C. C. G., Caetano, A. R., and Cardoso, F. F. (2017). Tag SNP Selection for Prediction of Tick Resistance in Brazilian Braford and Hereford Cattle Breeds Using Bayesian Methods. *Genet. Sel. Evol.* 49, 49. doi:10.1186/s12711-017-0325-2

Spelman, R. J., Huisman, A. E., Singireddy, S. R., Coppieters, W., Arranz, J., Georges, M., et al. (1999). Short Communication: Quantitative Trait Loci Analysis on 17 Nonproduction Traits in the New Zealand Dairy Population. *J. Dairy Sci.* 82, 2514–2516. doi:10.3168/jds.s0022-0302(99)75503-7

Srikanth, K., Lee, S.-H., Chung, K.-Y., Park, J.-E., Jang, G.-W., Park, M.-R., et al. (2020). A Gene-Set Enrichment and Protein-Protein Interaction Network-Based GWAS with Regulatory SNPs Identifies Candidate Genes and Pathways Associated with Carcass Traits in Hanwoo Cattle. *Genes* 11, 316. doi:10.3390/genes11030316

Stafuzza, N. B., Costa e Silva, E. V. d., Silva, R. M. d. O., Costa Filho, L. C. C. d., Barbosa, F. B., Macedo, G. G., et al. (2019). Genome-wide Association Study for Age at Puberty in Young Nelore Bulls. *J. Anim. Breed. Genet.* 137, 234–244. doi:10.1111/jbg.12438

Stone, R. T., Keele, J. W., Shackelford, S. D., Kappes, S. M., and Koohmaraie, M. (1999). A Primary Screen of the Bovine Genome for Quantitative Trait Loci Affecting Carcass and Growth Traits. *J. Anim. Sci.* 77, 1379–1384. doi:10.2527/1999.7761379x

Sugimoto, M., Watanabe, T., and Sugimoto, Y. (2012). The Molecular Effects of a Polymorphism in the 5′UTR of Solute Carrier Family 44, Member 5 that Is Associated with Birth Weight in Holsteins. *PLoS One* 7, e41267. doi:10.1371/journal.pone.0041267

Sun, J., Jin, Q., Zhang, C., Fang, X., Gu, C., Lei, C., et al. (2011). Polymorphisms in the Bovine Ghrelin Precursor (GHRL) and Syndecan-1 (SDC1) Genes that Are Associated with Growth Traits in Cattle. *Mol. Biol. Rep.* 38, 3153–3160. doi:10.1007/s11033-010-9986-4

Tang, H., Peng, J., Wang, P., and Risch, N. J. (2005). Estimation of Individual Admixture: Analytical and Study Design Considerations. *Genet. Epidemiol.* 28, 289–301. doi:10.1002/gepi.20064

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

Thomasen, J. R., Guldbrandtsen, B., Sørensen, P., Thomsen, B., and Lund, M. S. (2008). Quantitative Trait Loci Affecting Calving Traits in Danish Holstein Cattle. *J. Dairy Sci.* 91, 2098–2105. doi:10.3168/jds.2007-0602

Thornton, T. A., and Bermejo, J. L. (2014). Local and Global Ancestry Inference and Applications to Genetic Association Analysis for Admixed Populations. *Genet. Epidemiol.* 38 (Suppl. 1), S5–S12. doi:10.1002/gepi.21819

Wang, Y., Wang, S., Liu, T., Tu, W., Li, W., Dong, G., et al. (2015). CARD15 Gene Polymorphisms Are Associated with Tuberculosis Susceptibility in Chinese Holstein Cows. *PLoS One* 10, e0135085. doi:10.1371/journal.pone.0135085

Warwick, E. J. (1958). Fifty Years of Progress in Breeding Beef Cattle. *J. Anim. Sci.* 17, 922–943. doi:10.2527/jas1958.174922x

Whitacre, L. (2014). *Structural Variation at the KIT Locus Is Responsible for the Piebald Phenotype in Hereford and Simmental Cattle*. Columbia, Missouri: University of Missouri-Columbia. MS thesis. Available at: https://mospace.umsystem.edu/xmlui/handle/10355/44434 (Accessed Aug 7, 2021).

Wu, X., Fang, M., Liu, L., Wang, S., Liu, J., Ding, X., et al. (2013). Genome Wide Association Studies for Body Conformation Traits in the Chinese Holstein Cattle Population. *BMC Genomics* 14, 897. doi:10.1186/1471-2164-14-897

Zhang, C., De Koning, D. J., Herna´ndez-Sa´nchez, J., Haley, C. S., Williams, J. L., and Wiener, P. (2004). Mapping of Multiple Quantitative Trait Loci Affecting Bovine Spongiform Encephalopathy. *Genetics* 167, 1863–1872. doi:10.1534/genetics.104.026401

Zhang, X., Chu, Q., Guo, G., Dong, G., Li, X., Zhang, Q., et al. (2015). Genome-wide Association Studies Identified Multiple Genetic Loci for Body Size at Four Growth Stages in Chinese Holstein Cattle. *PLoS One* 12, e0175971. doi:10.1371/journal.pone.0175971

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 February 2022
doi: 10.3389/fgene.2022.779152

Genome-Wide Assessment of a Korean Composite Pig Breed, Woori-Heukdon

Yong-Min Kim[1,2†], Ha-Seung Seong[1,2†], Young-Sin Kim[1], Joon-Ki Hong[1], Soo-Jin Sa[1], Jungjae Lee[3], Jun-Hee Lee[4], Kyu-Ho Cho[1], Won-Hyong Chung[5], Jung-Woo Choi[2] and Eun-Seok Cho[1]**

[1]Swine Science Division, National Institute of Animal Science, Rural Development Administration, Cheonan, South Korea, [2]Department of Animal Science, College of Animal Life Sciences, Kangwon National University, Chuncheon, South Korea, [3]Department of Animal Science and Technology, College of Biotechnology and Natural Resources, Chung-Ang University, Anseong, South Korea, [4]Institute of Agriculture and Life Science, College of Agriculture and Life Sciences, Gyeongsang National University, Jinju, South Korea, [5]Research Group of Healthcare, Korea Food Research Institute, Wanju, South Korea

OPEN ACCESS

Edited by:
Tiago do Prado Paim,
Goiano Federal Institute (IFGOIANO), Brazil

Reviewed by:
Peipei Ma,
Shanghai Jiao Tong University, China
Guillermo Giovambattista,
CONICET Institute of Veterinary Genetics (IGEVET), Argentina

***Correspondence:**
Jung-Woo Choi
jungwoo.kor@gmail.com
Eun-Seok Cho
segi0486@korea.kr

[†]These authors have contributed equally to this work and share first authorship

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

Received: 18 September 2021
Accepted: 07 January 2022
Published: 02 February 2022

Citation:
Kim Y-M, Seong H-S, Kim Y-S, Hong J-K, Sa S-J, Lee J, Lee J-H, Cho K-H, Chung W-H, Choi J-W and Cho E-S (2022) Genome-Wide Assessment of a Korean Composite Pig Breed, Woori-Heukdon. Front. Genet. 13:779152. doi: 10.3389/fgene.2022.779152

A Korean synthetic pig breed, Woori-Heukdon (WRH; F3), was developed by crossing parental breeds (Korean native pig [KNP] and Korean Duroc [DUC]) with their crossbred populations (F1 and F2). This study in genome-wide assessed a total of 2,074 pigs which include the crossbred and the parental populations using the Illumina PorcineSNP60 BeadChip. After quality control of the initial datasets, we performed population structure, genetic diversity, and runs of homozygosity (ROH) analyses. Population structure analyses showed that crossbred populations were genetically influenced by the parental breeds according to their generation stage in the crossbreeding scheme. Moreover, principal component analysis showed the dispersed cluster of WRH, which might reflect introducing a new breeding group into the previous one. Expected heterozygosity values, which were used to assess genetic diversity, were .365, .349, .336, .330, and .211 for WRH, F2, F1, DUC, and KNP, respectively. The inbreeding coefficient based on ROH was the highest in KNP (.409), followed by WRH (.186), DUC (.178), F2 (.107), and F1 (.035). Moreover, the frequency of short ROH decreased according to the crossing stage (from F1 to WRH). Alternatively, the frequency of medium and long ROH increased, which indicated recent inbreeding in F2 and WRH. Furthermore, gene annotation of the ROH islands in WRH that might be inherited from their parental breeds revealed several interesting candidate genes that may be associated with adaptation, meat quality, production, and reproduction traits in pigs.

Keywords: Woori-Heukdon, Korean native pig, genetic diversity, runs of homozygosity, selection signature, synthetic breed

INTRODUCTION

In the swine industry, crossbreeding has been widely used to exploit the phenomenon of heterosis, or hybrid vigor. The main benefit of the heterosis is increased performance of the resulting crossbred offspring over the average performance of its purebred parent pigs in traits of interest (Johnson, 1981; Falconer and Mackay, 1996). Although not all pig traits that were targeted for hybrid vigor show the same degree of heterosis, there has been significant success in harnessing heterosis to improve

productivity of several economically important traits (Baas et al., 1992; Cassady et al., 2002a; Cassady et al., 2002b). Since 1970, most commercial pig producers have been using a classical terminal crossbreeding system with three breeds (Landrace × Yorkshire dam × Duroc sire, or LYD) to produce market pork in South Korea. The LYD system resulted in higher productivities, including increased growth rate, by taking advantage of crossbreeding (Jin et al., 2006; Choi et al., 2016).

In recent years, the South Korea swine industry has been partly changing from emphasis on efficiently producing more pork, mostly by focusing on growth and reproduction, to addressing taste-oriented consumption. There has been consistent consumer demand in Korea for diversified and great-tasting pork, despite its higher price than the typical market pork that is mostly derived from the LYD system (Kim S. G. et al., 2020). In response to such demand, the National Institute of Animal Science in Korea developed Woori-Heukdon (WRH), which is a synthetic breed derived from crossbreeding Duroc (DUC) and the Korean native pig breed (KNP) (Kim Y.-M. et al., 2020). KNP is known to have great meat qualities, such as high glucose content and a high unsaturated/saturated fatty acid ratio. However, because it also has economically unfavorable characteristics, such as slow growth rate, late maturity, and light carcass weight, it has low productivity compared with the commercial breeds; therefore, the population has decreased (Park et al., 2007; Hur et al., 2013). To take advantage of genetic merits of both the Duroc and KNP populations, WRH were generated by the crossing scheme shown in **Figure 1**. F2 was shown to have a slightly better growth rate compared with WRH; however, meat qualities, such as meat color and shear force, were clearly better in WRH than F2. Therefore, WRH preserved the characteristics of KNP, including superior meat quality, and had an improved growth rate compared with KNP (Kim et al., 2016).

Runs of homozygosity (ROH) are contiguous homozygous stretches that are inherited from each parent. Although ROHs are known to arise from several genetic factors, including genetic drift and population bottlenecks, it has been used as an indicator of selection signatures throughout the genome. In fact, ROHs have been widely used to quantify autozygosity in pigs (Schachler et al., 2020; Wu et al., 2020); because ROH are widely but not randomly distributed across the genome, some ROH overlap with genomic regions associated with economically important traits in pigs. Furthermore, ROH can be used to distinguish between recent and ancient inbreeding (Keller et al., 2011). In particular, shared ROH within a population can be used to identify genomic regions potentially under selection, which could be associated with the environment or production systems. Several studies have assessed autozygosity in livestock species using ROH approaches (Peripolli et al., 2018; Xu Z. et al., 2019; Dzomba et al., 2021). Offspring inherit chromosomal segments from the same ancestor by descent (Broman and Weber, 1999). Consequently, the extent of ROH can be used to estimate the inbreeding coefficient (Bosse et al., 2012; Marras et al., 2015).

Currently, there is still a lack of studies to elucidate genomic characteristics of the composite breeds that were generated by using the indigenous pigs. The main objective of this study were: 1) to estimate various parameters to clarify genomic population structure of the crossbreds with DUC and KNP populations; and 2) to detect and investigate ROH that could be an indicative of selection signature in WRH population.

MATERIALS AND METHODS

Animals and Genotyping

A total of 2,074 pigs (male and female) were used in this study, including DUC, KNP, and their crossbred F1 (DUC × KNP), F2 (F1 × DUC), and F3 (WRH; F1 × F2) populations. The dataset includes previously published genotype data (Kim Y.-M. et al., 2020) that contained 61,565 SNPs for DUC (N = 50), KNP (N = 50), and the composite breed WRH (F3; N = 100). We also generated whole blood samples from 1,874 individuals of DUC (N = 1,029), KNP (N = 158), F1 (N = 11), F2 (N = 144), and WRH (F3; N = 532) at the National Institute of Animal Science, Rural Development Administration, Korea. The genomic DNA was extracted from the blood samples using the phenol–chloroform method and genotyped using the Illumina PorcineSNP60 BeadChip v2 (Illumina, Inc., San Diego, CA, United States), which contains 61,565 SNPs. Genotype data were called using the genotype module in GenomeStudio v2.0 (Illumina, Inc.). The SNP coordinates were updated from the genome assembly of *Sus scrofa* (Sscrofa10.2 to Sscrofa11.1) according to manifest file of Illumina PorcineSNP60 v2 using PLINK v1.9 (Chang et al., 2015).

Quality Control of Genotype Data

The quality control (QC) process for the genotype dataset was conducted separately for SNPs and animals using PLINK v1.9. We applied three different QC procedures for the initial raw SNP dataset to further analyze genetic diversity, population structure, and ROH. All three genotype subsets used the following process: 1) SNPs in sex chromosomes or unmapped in Sscrofa11.1; 2) SNPs with a call rate less than 90%. Individuals with a call rate less than 90% were removed. The ROH analysis was conducted using a subset of data without additional QC process. To assess genetic diversity, we also removed SNPs with a minor allele frequency (MAF) less than .05. Furthermore, SNP filtering based on pairwise linkage disequilibrium (LD) was conducted to minimize the reduction of informativeness of the dataset (Lopes et al., 2013) using the indep-pairwise 50 5 0.5 command for population structure analyses.

Genetic Diversity

To assess genetic diversity within populations, we calculated observed heterozygosity (H_O), expected heterozygosity (H_E), and individual inbreeding coefficient (F_{HOM}) using PLINK v1.9. The inbreeding coefficient was calculated based on homozygous genotypes as follows:

$$F_{HOM} = \frac{\text{Observed homozygous loci} - \text{Expected homozygous loci}}{\text{Total number of nonmissing loci} - \text{Expected homozygous loci}}$$

Population Structure

To explore the pattern of genetic differentiation of samples, we conducted principal component analysis (PCA) using PLINK

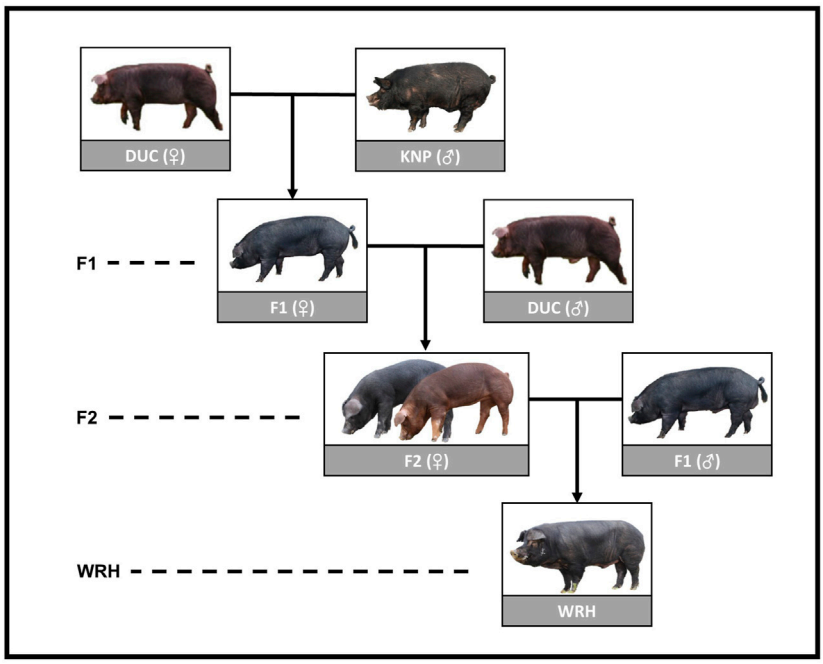

FIGURE 1 | Crossbreeding scheme of WRH using parental breeds (DUC and KNP) and their crossbred populations (F1 and F2). DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).

v1.9, and the first two principal components (PCs) were visualized using R v4.1.0 (www.r-project.org, accessed July 4, 2021). The PCA plot was also considered a QC process because it reveals potential misclassification.

To better understand the relationship among parental breeds and their crossbred populations, an admixture analysis for K = 2 that is based on the number of ancestral populations was performed using ADMIXTURE v1.3 (Alexander and Lange, 2011).

ROH Detection

To detect ROH, we used a dataset of 56,498 SNPs for 2,040 individuals that resulted from QC without filtering based on MAF and LD because of problematic factors in ROH discovery (Marras et al., 2015; Meyermans et al., 2020). To avoid short and common ROH caused by LD, we set the minimum length of ROH to 1 Mb for ROH discovery, as described by several studies (Purfield et al., 2012; Ferencakovic et al., 2013; Marras et al., 2015; Meyermans et al., 2020). ROHs were identified using a sliding window method with the homozyg command in PLINK v1.9. The parameters used in ROH detection were applied as follows: 1) the minimum length of ROH was 1 Mb (--homozyg-kb); 2) an ROH had at least one SNP per 50 kb on average (--homozyg-density); 3) the distance of consecutive SNPs in the same ROH was less than 1,000 kb (--homozyg-gap); 4) no heterozygous SNP (--homozyg-window-het and--homozyg-het) and one SNP with a missing genotype were allowed (--homozyg-window-missing). In addition, the thresholds for the minimal SNP numbers per window (--homozyg-window-snp) and in the ROH (--homozyg-snp) were calculated by the L-parameter for the populations (Lencz et al., 2007; Purfield et al., 2012; Meyermans et al., 2020). In this study, we classified ROH into three categories based on their physical length: short (1 to <3 Mb), medium (3 to <5 Mb) and long (>10 Mb).

ROH islands, which were defined as regions where SNPs in ROHs had *p*-values higher than a specific threshold for each population, were marked as a potential selection signature, and ROH islands were determined using R-script at the Open Science Framework (https://doi.org/10.17605/OSF.IO/XJTKV) provided by Gorssen et al. (2021). The population-specific threshold was determined based on z-scores obtained from the distribution of ROH incidences. Additionally, the top 0.1% of SNPs (*p*-value >.999) were used to form ROH islands (Purfield et al., 2017; Gorssen et al., 2020). Furthermore, a threshold that a ROH should be included in at least 30% of individuals within each population was set for ROH islands. For highly inbred populations such as KNP (F_{ROH} = 41%), the threshold was set to 80% because SNPs with *p*-values higher than .999 were not found in this population, as described by Gorssen et al. (2021).

Using discovered ROHs, we calculated the inbreeding coefficient based on ROH using the following calculation proposed by McQuillan et al. (2008):

$$\mathbf{F_{ROH}} = \frac{\mathbf{L.ROH}}{\mathbf{L.Autosomes}}$$

Where L. ROH is the sum of all ROH of an individual and L. Autosomes is the total length of autosomal genome covered by SNPs (in this study, 2,262.6 Mb). Furthermore, we investigated

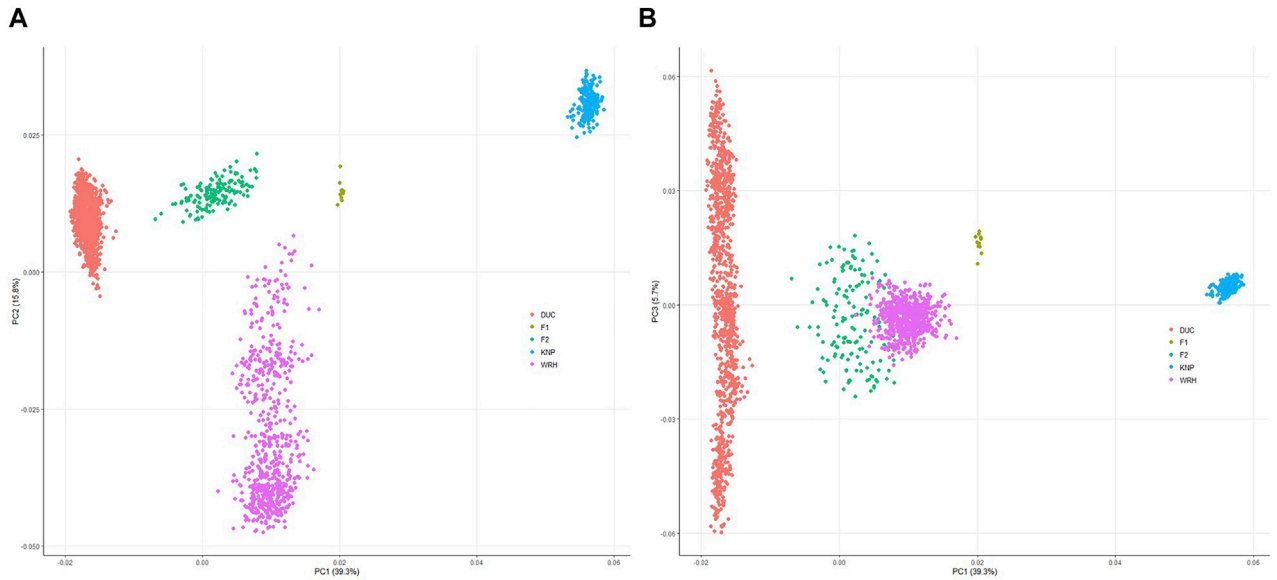

FIGURE 2 | Principal component analysis illustrating stratification of the five pig populations. **(A)** Visualization of PC1 vs PC2, which explains 55.1% of total variation together; **(B)** visualization of PC1 vs PC3, which explains 45%. DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).

patterns of inbreeding estimates based on homozygosity and ROH by calculating Pearson's correlations.

Annotation of Genes and Quantitative Trait Loci

Gene annotation for ROH islands was conducted according to NCBI Sus scrofa Release 106 (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/025/GCF_000003025.6_Sscrofa11.1/GCF_000003025.6_Sscrofa11.1_genomic.gff.gz, accessed on 15 June 2021). Furthermore, we also used pig quantitative trait locus information for ROH islands to annotate potential traits associated with selected regions (https://www.animalgenome.org/, accessed on 20 July 2021).

RESULTS

SNP Characteristics

The SNP coordinates for each chromosome (SSC) were updated from Sscrofa10.2 to Sscrofa11.1 according to manifest file provided from Illumina (**Supplementary Table 1**). After initial QC steps, three additional individuals (DUC = 3) were removed because of breed misclassification according to PCA (**Supplementary Figure 1**). As a result of the QC procedure, we retrieved three SNP subsets for population structure (12,801 SNPs for 2,040 individuals), genetic diversity (43,809 SNPs for 2,040 individuals), and ROH analyses (56,498 SNPs for 2,040 individuals).

Population Structure

The top three PCs (PC1, PC2, and PC3) explained approximately 39.3, 15.8, and 5.7% of total variation, respectively. As shown in **Figure 2A**, DUC and KNP were clearly separated, and the cluster of F1 was located in the middle of DUC and KNP populations based on PC1. F2 and WRH were also located between DUC and KNP; however, both populations were closer to DUC. Furthermore, WRH showed a dispersed cluster based on PC2. Clustering patterns for (F1, F2, and WRH) were clear based on PC1 and PC3, which explained approximately 39.3 and 5.7% of the total variation, respectively (**Figure 2B**). A similar clustering pattern was observed in ADMIXTURE analysis, showing distinct ancestries of 99.3 and 100.0% in DUC and KNP, respectively (**Figure 3**; **Supplementary Table 2**). For their crossbreds, F1, F2, and WRH had 51.1, 25.9, and 34.1% KNP ancestry, respectively, at K = 2.

Genetic Diversity and Inbreeding

Genetic diversity estimates using a subset of 43,809 SNPs for 2,040 individuals are summarized in **Table 1**. All crossbred populations (F1, F2, and WRH) had higher heterozygosity rates than the initial parental breeds (DUC and KNP). F1 had the highest average H_O value (.477 ± .297), followed by WRH (.407 ± .164), F2 (.380 ± .163), DUC (.331 ± .160), and KNP (.223 ± .216). The average H_E value was highest in WRH (.365 ± .133) and lowest in KNP (.211 ± .202). The inbreeding coefficient values (F_{HOM}) were shown to have negative values in all five populations, exhibiting the highest value in DUC (−.002 ± .053), followed by WRH (−.024 ± .083), KNP (−.055 ± .104), F2 (−.115 ± .048), and F1 (−.417 ± .019).

ROH Patterns

The mean number and size of discovered ROH per pig within each population are described in **Table 2**. Among the five pig populations, the largest mean size of ROH was observed in KNP

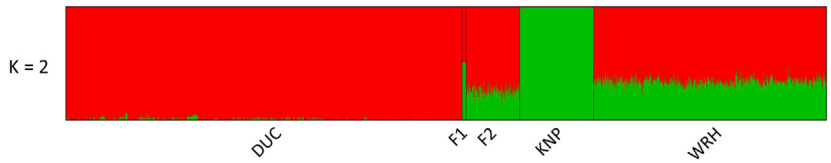

FIGURE 3 | Population structure analysis based on ADMIXTURE at K = 2. Each column represents an individual. DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).

TABLE 1 | Estimates of genetic diversity and inbreeding for five pig populations.

Pop	No. before QC[a]	H_O ± SD	H_E ± SD	F_{HOM} ± SD
DUC	1,079	0.331±0.160	0.330±0.159	−0.002±0.053
KNP	208	0.223±0.216	0.211±0.202	−0.055±0.104
F1	11	0.477±0.297	0.336±0.163	−0.417±0.019
F2	144	0.380±0.163	0.349±0.138	−0.115±0.048
WRH	632	0.407±0.164	0.365±0.133	−0.024±0.083

[a]*Number of samples before quality control (missing genotype rate <90%). Pop, population; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{HOM}, inbreeding coefficient based on excess of homozygosity; SD, standard deviation; DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).*

(14,205.2 ± 17,912.2 kb; N = 65.1), and followed by WRH (7,169.8 ± 8,695.3 kb; N = 58.7), DUC (6,117.5 ± 6,418.3 kb; 65.9), F2 (3,973.9 ± 3,836.9 kb; N = 61.2) and F1 (2,003.5 ± 1,014.0 kb; N = 39.8). Both parental breeds (DUC and KNP) showed to have larger mean size and number of ROH than F1 and F2 crossbred populations, whereas mean length and ROH number were higher in WRH than that of DUC. The ROH length for parental breeds ranged from 1,266 kb (SSC12) to 186,522 kb (SSC1) in DUC and from 2,060 kb (SSC2) to 262,078 kb (SSC1) in KNP. For crossbred populations, the ROH length ranged from 1,014 kb (SSC1) to 9,527 kb (SSC6) in F1, 1,041 kb (SSC12) to 58,419 kb (SSC1) in F2, and 1,247 kb (SSC2) to 152,760 kb (SSC1) in WRH. As shown in **Figure 4** and **Supplementary Table 3**, the proportion of short ROH (1–3 Mb) within each population was the highest in F1 (90.2%; N = 395), followed by F2 (54.9%; N = 4,839), WRH (26.5%; N = 9,691), DUC (24.4%; N = 17,075), and KNP (1.6%; N = 203). The proportion of medium ROH (3–10 Mb) was highest in DUC (62.9%; N = 44,038) and lowest in F1 (9.8%; N = 43), long ROH (>10 Mb) were most frequently observed in KNP (41.9%; N = 5,436), and no long ROH were observed in F1. We also observed parental ROH regions in crossbred populations using the coordinates of concatenated ROH regions that were generated by joining overlapped (at least 1bp) ROH segments within each of the populations (DUC, KNP, F1, F2, and WRH) (**Supplementary Table 4**). Most of ROH regions (>98.9%) in crossbreds were derived from the shared ROH region (approximately 2,218 Mb) between DUC and KNP.

The inbreeding coefficient among all populations was further calculated based on ROH (F_{ROH}), which was defined by McQuillan et al. (2008) (**Table 2**). This result showed that the highest F_{ROH} value was observed in KNP (0.409) followed by WRH (0.186), DUC (0.178), F2 (0.107), and F1 (0.035), which is the same order as mean ROH length. The correlation between F_{HOM} and F_{ROH} for each population was high in most populations, including DUC (0.84), KNP (0.83), F2 (0.86), and WRH (0.88); however, F1 had a correlation of 0.10 (**Supplementary Figure 2**). Because F_{HOM} is influenced by allele frequency and sampling (Zhang et al., 2015) but the level of homozygosity in F_{ROH} is independent of allele frequencies, the low correlation between F_{HOM} and F_{ROH} in F1 might be caused by sampling bias derived from the low sample size of F1 (N = 11) (Dixit et al., 2020). Therefore, we focused on F_{ROH} to assess the degree of inbreeding, which is also better for detecting both common and rare variants than F_{HOM} (Wang et al., 2019).

ROH Islands and Gene Annotation

As shown in **Figure 5** and **Supplementary Table 5**, the ROH islands, which were defined as regions where SNPs in ROH had *p*-values higher than a threshold for each population, were observed for parental (DUC and KNP) and crossbred (F1, F2, and WRH) populations. We identified a total of 365 ROH islands of 1-Mb bins throughout the autosomal regions (**Supplementary Table 6**). DUC had ROH islands on SSC1–3, SSC7, and SSC14, whereas KNP had ROH islands on 11 autosomes, excluding SSC2, SSC3, SSC4, SSC6, SSC14, SSC15, and SSC18. Crossbred populations, especially F1 and F2, had similar occurrence patterns of ROH islands, with ROH islands on SSC1, SSC9, and SSC14–17 in both populations. Additional islands only in F1 were on SSC4, SSC5, SSC13, and SSC18, and those only in F2 were on SSC2 and SSC7. For WRH, SSC1, SSC3, SSC7, SSC9, and SSC14 had ROH islands. For all 1-Mb bin of ROH islands, we annotated 2,165 genes and 11 QTL information and are listed in **Supplementary Table 6**.

DISCUSSION

To investigate the population structure of DUC, KNP, and their crossbred populations (F1, F2, and WRH), we conducted PCA and ADMIXTURE analyses (**Figure 2** and **Figure 3**). The most distant genetic relationship was observed between DUC and KNP based on PCA (**Figure 2**). In addition, KNP separated from DUC at K = 2 and had 100% distinct ancestry (**Figure 3**; **Supplementary Table 2**). The results are consistent with those of previous reports that showed clear genetic difference based on population structure and F_{ST} analyses (Edea et al., 2014;

TABLE 2 | Summary of discovered ROH in five pig populations.

Pop	ROH length (kb)			No. of ROH per individual			F_{ROH} ± SD
	Mean	Min[a]	Max[b]	Mean	Min[c]	Max[d]	
DUC	6,117.5 ± 6,418.3	1,266.3 (SSC12)	186,521.5 (SSC1)	65.9±8.0	4	90	0.178±0.035
KNP	14,205.2 ± 17,912.2	2,060.1 (SSC2)	262,078.0 (SSC1)	65.1±5.5	47	80	0.409±0.046
F1	2,003.5 ± 969.0	1,014.0 (SSC1)	9,527.0 (SSC6)	39.8±3.3	35	44	0.035±0.003
F2	3,973.9 ± 3,836.9	1,040.8 (SSC12)	58,419.8 (SSC1)	61.2±9.9	31	87	0.107±0.025
WRH	7,169.8 ± 8,695.3	1,247.0 (SSC2)	152,760.6 (SSC1)	58.7±9.3	7	89	0.186±0.055

[a]*Minimum length (kb) of ROH, within each population.*
[b]*Maximum length (kb) of ROH, within each population.*
[c]*Minimum number of ROHs, per individual.*
[d]*Maximum number of ROHs, per individual. Pop, population; No., number; F_{ROH}, inbreeding coefficient based on runs of homozygosity; SD, standard deviation; DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).*

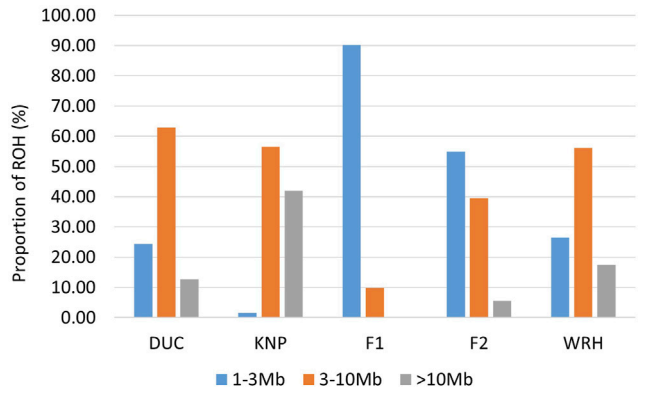

FIGURE 4 | ROH distribution according to size classification (1–3 Mb, 3–10 Mb, or >10 Mb). DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).

Kim Y.-M. et al., 2020; Lee et al., 2020). For the crossbred populations (F1, F2, and WRH) used in this study, all three populations were located between KNP and DUC in the PCA (**Figure 2**). In particular, F1 was located in the middle of KNP and DUC, which was also shown in other F1 populations generated from purebred parental pig breeds (Grossi et al., 2017) and cattle (Gobena et al., 2018). The ADMIXTURE analysis of K = 2 also revealed that F1 had 48.9% DUC and 51.1% KNP ancestry (**Supplementary Table 2**). The genetic distance of F2 and WRH crossbred populations was closer to DUC than that to KNP, which can be explained by the higher genetic composition of DUC than that of KNP in the crossing scheme of F2 and WRH (**Figure 1**). At K = 2 in the ADMIXTURE analysis, F2 had approximately 74.1% DUC and 25.9% of KNP ancestry. In addition, WRH had approximately 65.9% DUC and 34.1% KNP ancestry. This better fits the theoretical genomic composition of DUC (62.5%) and KNP (37.5%) than a previous study that reported the genomic composition of DUC (74.8%) and KNP (25.2%) in WRH (Kim Y.-M. et al., 2020). This study only used parental breeds (DUC and KNP) and crossbred populations (F1 and F2) that were used to develop WRH, even though the previous study used additional Chinese and commercial breeds. Therefore, we inferred that a better estimation of ancestry for WRH was obtained in this study.

However, WRH was shown to have a somewhat dispersed cluster in PCA (**Figure 2A**). This could be explained by newly generated breeding group of WRH using parental breeds (DUC and KNP) and crossbred populations (F1 and F2). The first national project to develop a Korean composite pig breed (WRH) started in 2008 by generating F1 and F2 populations. Subsequently, the first founder population of WRH was developed in 2010. Since then, the initial founder stock of WRH was used for breeding projects as a closed population until 2018. Because of difficulties maintaining a sufficiently large effective population size with a limited population size, they were subject to inbreeding (Dickerson, 1973). To decrease the level of inbreeding of this population, a recent project was initiated to construct new breeding group of WRH population since 2018. Therefore, recent introduction of new breeding group to the previous one might cause the somewhat widely distributed cluster shown in population structure analyses.

The effect of heterosis is difficult to quantify; however, heterozygosity can be used as an indicator of heterosis (Iversen et al., 2019). Genetic diversity levels were assessed by mean expected heterozygosity rate of five pig populations. The expected heterozygosity gradually increased according to crossing stages used in this study (F1, F2, and WRH) (**Table 1**), and those values were higher than in the parental breeds. Similar to this result, a previous study reported that crossbred pigs generated between Dutch Landrace and Dutch Large White had higher heterozygosity levels compared with their parental breeds (Iversen et al., 2019). Among the five pig populations, the lowest degree of genetic diversity was observed in KNP, and this result is concordant with those of previous studies that reported the genetic diversity of KNP relative to most of studied other indigenous and commercial pig breeds using genomic datasets (Kim Y.-M. et al., 2020; Lee et al., 2020). Other studies that assessed genetic diversity of indigenous pigs from other countries also revealed loss of genetic diversity in indigenous pig breeds due to conservation status (Diao et al., 2019; Munoz et al., 2019). Although KNP has not undergone systematic artificial selection during conservation breeding, such loss of genetic diversity could be explained by a small effective

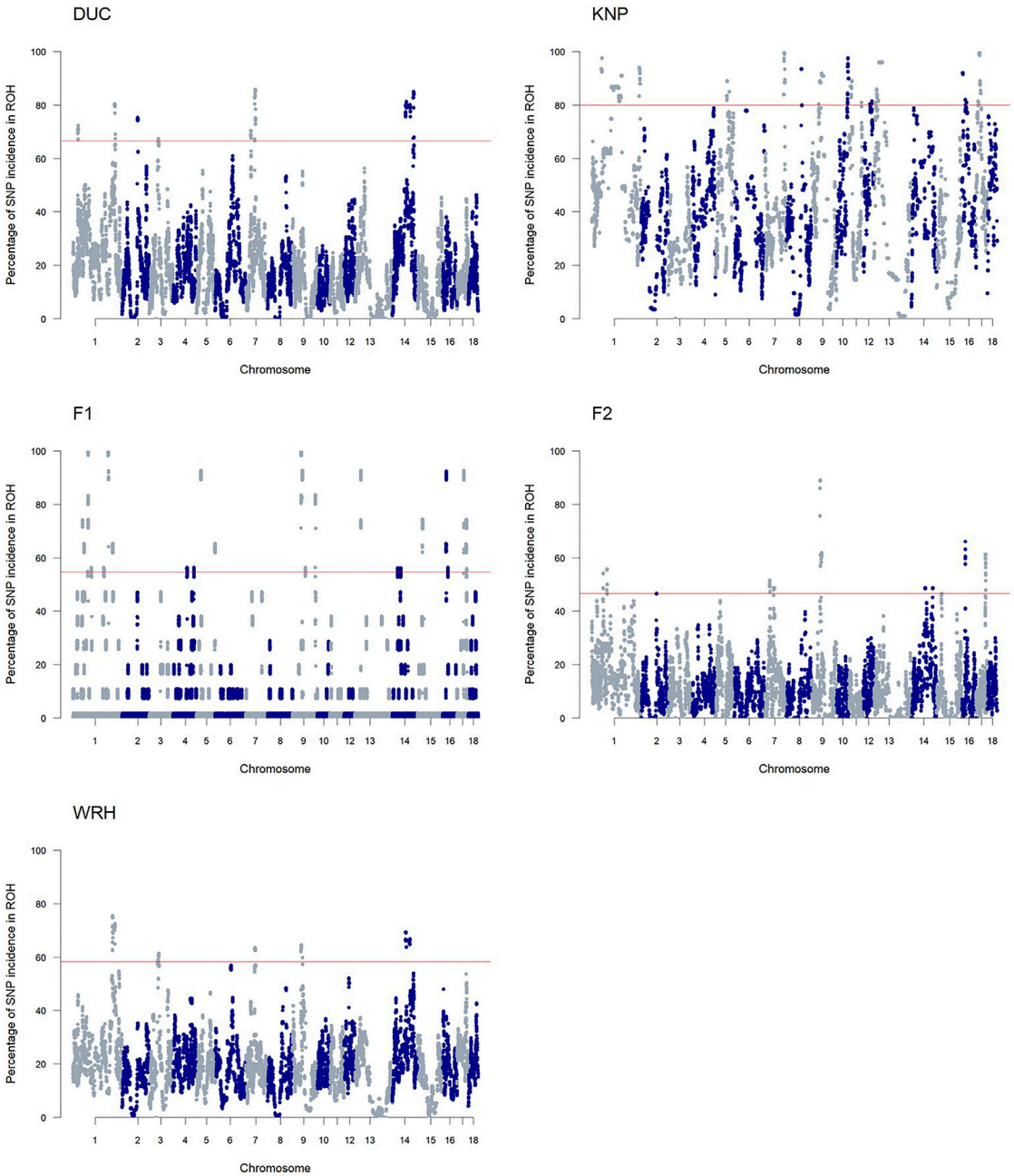

FIGURE 5 | Manhattan plot of SNP frequency in ROH islands. The horizontal line (red) indicates the threshold for ROH islands for each population. DUC, Korean Duroc; KNP, Korean native pig; F1, DUC × KNP; F2, F1 × DUC; WRH, Woori-Heukdon (F1 × F2).

population size and the founder effect or a population bottleneck (Munoz et al., 2019). In terms of breeding history, the KNP restoration project started in 1988 using only nine individuals (four males and five females) as a founder population (Kim et al., 2016); simultaneously, the preservation program was conducted as a closed population (Kim Y.-M. et al., 2020). We suggest that those characteristics of the breeding system might have caused a high level of inbreeding, as KNP had the greatest F_{ROH} value (.409) (**Table 2**).

In ROH analyses, we found a clear difference in ROH patterns between parental breeds and their crossbred populations (F1 and F2). Among parental breeds, DUC had 24.4% short and 12.7% long ROH segments (**Figure 4**). Additionally, KNP had abundant medium (56.5%) and long (41.9%) ROH segments. The short ROH segments may indicate evolutionary events from old inbreeding or selection, whereas long ROH may reflect recent inbreeding (Keller et al., 2011; Mastrangelo et al., 2017; Xu L. et al., 2019). DUC was shown to have more old inbreeding than

recent inbreeding or selection pressure due to intensive selection programs (Bovo et al., 2020), whereas KNP underwent recent inbreeding of small populations (Valluzzi et al., 2021); this was also confirmed by the genetic diversity and inbreeding results in this study. Of the crossbred populations, F1 has the lowest number of total ROH segments (N = 438) among populations used in this study (**Supplementary Table 3**), and most of them are short ROHs (90.2%) and no long ROHs. These ROH patterns in F1 also caused the short length of concatenated ROH regions compared to F2 and WRH (**Supplementary Table 4**), which might be caused by the increase of heterozygous SNPs in the genome due to the hybridization of parental breeds. This is also supported by the highest value of observed heterozygosity in F1 (**Table 1**). However, a caution is required to interpret the ROH results derived from F1, because there was highly likely to be an underestimation in ROH numbers and size due to a low sample size (N = 11). Furthermore, we identified a total of 8,810 and 36,635 ROHs for F2 and WRH, respectively (**Supplementary Table 3**). In both populations, short ROH segments gradually decreased, whereas medium and long ROH segments increased; these changes also increased the inbreeding coefficient based on ROH (**Table 2**) and the length of concatenated ROH region (**Supplementary Table 4**) in those populations. A previous study also revealed that, in a 3-way crossbred population [(Pietrain × Large White) × Duroc)], G0 had smaller homozygous segments than their parental populations, and ROH size was increased in the G1 population, which were the offspring of G0 (Ganteil et al., 2020). Consequently, WRH showed a similar proportion of size-classified ROH to DUC, but WRH had a higher proportion of long ROH than DUC; this indicated recent inbreeding events, as discussed earlier in the population structure analyses. We also observed parental ROHs in crossbred populations using coordinates of concatenated ROH regions for each of the populations (**Supplementary Table 4**). Most ROH regions in crossbreds (>98.8%) were considered to be from the ROH regions shared between parental breeds (DUC and KNP). In particular, the proportions of the ROH regions over the initial total length of ROH region overlapped between parental breeds were 91.55 and 99.55% for F2 and WRH, respectively (data not shown). We suggest that such increased ROH regions in F2 and WRH might be in part due to the inbreeding.

The discovery of ROH islands revealed numerous homozygous regions over five pig populations used in this study (**Figure 5**; **Supplementary Table 6**). Furthermore, we found ROH islands in the parental breeds that were shared with their crossbred populations on SSC1–3, SSC7, SSC9, SSC13, SSC14, and SSC16, which indicated inheritance of homozygous regions (**Supplementary Table 5**). Most of those regions in crossbred populations were shorter than those in parental populations, which might be explained by ROH degeneration due to an increase of heterozygous SNPs or recombination (Bosse et al., 2012). The breakage of the ROHs is supported by the proportion of parental ROH regions in crossbred populations (**Supplementary Table 4**), which indicated that most of ROH regions from the crossbreds (F1, F2 and WRH) belongs to shared ROH regions between parental breeds. Previous studies also revealed ROH persistence in several crossbred pigs: Landrace × Large White (Landrace × Large White) × Duroc, and (Pietrain × Large White) × Duroc (Howard et al., 2016; Gomez-Raya et al., 2019; Ganteil et al., 2020).

In this study, we annotated genes to ROH islands to identify inheritance of homozygous regions that might be potential selection signatures from parental breeds to WRH, which is the last stage of the crossbreeding scheme. First, we found some candidate genes in ROH islands that overlapped between KNP and WRH. We found an ROH island on SSC9 (49–50 Mb) harboring the Cytotoxic And Regulatory T Cell Molecule (*CRTAM*) gene. This gene was reported to have an association with adaptive immune response in cattle (Ben-Jemaa et al., 2021). As KNP is an indigenous pig breed that has been adapted to the local environment of South Korea for a long period, we suggest that *CRTAM* might be a candidate gene that is associated with local adaptation of KNP and WRH populations. At the same location (SSC9; 49–50 Mb) of a ROH island, we found the Heat Shock Protein Family A Member 8 (*HSPA8*) gene, which is also known as Hsp70. *HSPA8* is known to be associated with pork tenderness because this gene was down-regulated in tender samples (Hamill et al., 2012).

We also retrieved ROH islands that were shared between DUC and WRH. We found the ADAMTS Like 3 (*ADAMTSL3*) gene at 52–53 Mb on SSC7. *ADAMTSL3* is a member of the ADAMTS superfamily of proteins, and this gene was previously reported as a candidate gene for body length in Large White pigs (Li et al., 2017) and height in humans (Weedon et al., 2008). In addition, we located the Cytoplasmic Polyadenylation Element Binding Protein 1 (*CPEB1*) gene at 52–53 Mb on SSC7. The *CPEB1* is an RNA-binding protein that regulates mRNA translation by controlling the poly(A) tail length (Nagaoka et al., 2016). *CPEB1* was reported to increase the rate of meiotic resumption and expression of cyclin B when mRNA was injected into immature oocytes (Nishimura et al., 2010). The purpose of using DUC in the crossing scheme includes complementing the body size and reproductive traits of KNP; thus, we suggest that *ADAMTSL3* and *CPEB1*, which are located in ROH islands of DUC and WRH, might be associated with production and reproductive traits in both populations.

CONCLUSION

This study has shown genomic characteristics of crossbred pig populations derived from Korean Duroc and Korean native pigs as founder breeds. Population structure analysis showed genetic influence of founder breeds to crossbred populations. We also observed that WRH had two distinct subgroups due to newly introduced breeding group. For crossbred populations, the genetic diversity was gradually increased according to their crossbreeding stage (F1, F2 and WRH). In ROH analyses, short ROHs were decreased, while medium and long ROHs were increased from F1 to WRH, suggesting that recent

inbreeding is ongoing in WRH. In this study, there is a partial limitation to conclude on F1 due to its small samples size (N = 11). Furthermore, we identified shared ROH islands which contain candidate genes (*CRTAM*, *HSPA8*, *ADAMTSL3* and *CPEB1* genes) between WRH and founder breeds, suggesting inheritance of homozygous region that might be potential signatures of selection.

DATA AVAILABILITY STATEMENT

The SNP dataset for DUC used in this study is deposited in FigShare, https://doi.org/10.6084/m9.figshare.18318584.v1. The datasets for other populations are available from the corresponding author upon request.

ETHICS STATEMENT

All experimental protocols for this study were approved by the Institutional Animal Care and Use Committee at National Institute of Animal Science (approval number: NIAS 2020-479).

AUTHOR CONTRIBUTIONS

Y-MK, J-WC, and E-SC designed the study and revised the manuscript. Y-MK and H-SS performed the data analyses and drafted the manuscript. Y-MK, H-SS, Y-SK, J-KH, and S-JS collected the samples. K-HC, JL, J-HL, and W-HC participated in the experimental design and paper revision. All authors read and approved the final manuscript.

FUNDING

This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01263602)" Rural Development Administration (RDA), Republic of Korea and supported by 2022 the RDA Fellowship Program of National Institute of Animal Science, RDA, Republic of Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.779152/full#supplementary-material

REFERENCES

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinformatics* 12, 246. doi:10.1186/1471-2105-12-246

Baas, T. J., Christian, L. L., and Rothschild, M. F. (1992). Heterosis and Recombination Effects in Hampshire and Landrace Swine: I. Maternal Traits. *J. Anim. Sci.* 70 (1), 89–98. doi:10.2527/1992.70189x

Ben-Jemaa, S., Senczuk, G., Ciani, E., Ciampolini, R., Catillo, G., Boussaha, M., et al. (2021). Genome-Wide Analysis Reveals Selection Signatures Involved in Meat Traits and Local Adaptation in Semi-feral Maremmana Cattle. *Front. Genet.* 12, 675569. doi:10.3389/fgene.2021.675569

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 (11), e1003100. doi:10.1371/journal.pgen.1003100

Bovo, S., Ribani, A., Muñoz, M., Alves, E., Araujo, J. P., Bozzi, R., et al. (2020). Whole-genome Sequencing of European Autochthonous and Commercial Pig Breeds Allows the Detection of Signatures of Selection for Adaptation of Genetic Resources to Different Breeding and Production Systems. *Genet. Sel Evol.* 52 (1), 33. doi:10.1186/s12711-020-00553-7

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 (6), 1493–1500. doi:10.1086/302661

Cassady, J. P., Young, L. D., and Leymaster, K. A. (2002a). Heterosis and Recombination Effects on Pig Growth and Carcass Traits. *J. Anim. Sci.* 80 (9), 2286–2302. doi:10.2527/2002.8092286x

Cassady, J. P., Young, L. D., and Leymaster, K. A. (2002b). Heterosis and Recombination Effects on Pig Reproductive Traits. *J. Anim. Sci.* 80 (9), 2303–2315. doi:10.2527/2002.8092303x

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, 7. doi:10.1186/s13742-015-0047-8

Choi, Y.-S., Lee, J.-K., Jung, J.-T., Jung, Y.-C., Jung, J.-H., Jung, M.-O., et al. (2016). Comparison of Meat Quality and Fatty Acid Composition of Longissimus Muscles from Purebred Pigs and Three-Way Crossbred LYD Pigs. *Korean J. Food Sci. Anim. Resour.* 36 (5), 689–696. doi:10.5851/kosfa.2016.36.5.689

Diao, S., Huang, S., Xu, Z., Ye, S., Yuan, X., Chen, Z., et al. (2019). Genetic Diversity of Indigenous Pigs from South China Area Revealed by SNP Array. *Animals* 9 (6), 361. doi:10.3390/ani9060361

Dickerson, G. E. (1973). Inbreeding and Heterosis in Animals. *J. Anim. Sci.* 1973 (Symposium), 54–77. doi:10.1093/ansci/1973.Symposium.54

Dixit, S. P., Singh, S., Ganguly, I., Bhatia, A. K., Sharma, A., Kumar, N. A., et al. (2020). Genome-Wide Runs of Homozygosity Revealed Selection Signatures in Bos indicus. *Front. Genet.* 11, 92. doi:10.3389/fgene.2020.00092

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 (1), 7. doi:10.1186/s12864-020-07314-2

Edea, Z., Kim, S.-W., Lee, K.-T., Kim, T. H., and Kim, K.-S. (2014). Genetic Structure of and Evidence for Admixture between Western and Korean Native Pig Breeds Revealed by Single Nucleotide Polymorphisms. *Asian Australas. J. Anim. Sci.* 27 (9), 1263–1269. doi:10.5713/ajas.2014.14096

Falconer, D. S., and Mackay, T. F. C. (1996). *An Introduction to Quantitative Genetics*. Essex, UK: Longman Group.

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 (4), 286–293. doi:10.1111/jbg.12012

Ganteil, A., Rodriguez-Ramilo, S. T., Ligonesche, B., and Larzul, C. (2020). Characterization of Autozygosity in Pigs in Three-Way Crossbreeding. *Front. Genet.* 11, 584556. doi:10.3389/fgene.2020.584556

Gobena, M., Elzo, M. A., and Mateescu, R. G. (2018). Population Structure and Genomic Breed Composition in an Angus-Brahman Crossbred Cattle Population. *Front. Genet.* 9, 90. doi:10.3389/fgene.2018.00090

Gomez-Raya, L., Rauw, W. M., Dunkelberger, J. R., and Dekkers, J. C. M. (2019). Autozygosity and Genetic Differentiation of Landrace and Large White Pigs as Revealed by the Genetic Analyses of Crossbreds. *Front. Genet.* 10, 739. doi:10.3389/fgene.2019.00739

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

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. *Genetics Selection Evolution* 53 (1), 1–10.

Grossi, D. A., Jafarikia, M., Brito, L. F., Buzanskas, M. E., Sargolzaei, M., and Schenkel, F. S. (2017). Genetic Diversity, Extent of Linkage Disequilibrium and Persistence of Gametic Phase in Canadian Pigs. *BMC Genet.* 18 (1), 6. doi:10.1186/s12863-017-0473-y

Hamill, R. M., McBryan, J., McGee, C., Mullen, A. M., Sweeney, T., Talbot, A., et al. (2012). Functional Analysis of Muscle Gene Expression Profiles Associated with Tenderness and Intramuscular Fat Content in Pork. *Meat Sci.* 92 (4), 440–450. doi:10.1016/j.meatsci.2012.05.007

Howard, J. T., Tiezzi, F., Huang, Y., Gray, K. A., and Maltecca, C. (2016). Characterization and Management of Long Runs of Homozygosity in Parental Nucleus Lines and Their Associated Crossbred Progeny. *Genet. Sel Evol.* 48 (1), 91. doi:10.1186/s12711-016-0269-y

Hur, S. J., Jeong, T. C., Kim, G. D., Jeong, J. Y., Cho, I. C., Lim, H. T., et al. (2013). Comparison of Live Performance and Meat Quality Parameter of Cross Bred (Korean Native Black Pig and Landrace) Pigs with Different Coat Colors. *Asian Australas. J. Anim. Sci.* 26 (7), 1047–1053. doi:10.5713/ajas.2013.13005

Iversen, M. W., Nordbø, Ø., Gjerlaug-Enger, E., Grindflek, E., Lopes, M. S., and Meuwissen, T. (2019). Effects of Heterozygosity on Performance of Purebred and Crossbred Pigs. *Genet. Sel Evol.* 51 (1), 8. doi:10.1186/s12711-019-0450-1

Jin, S., Kim, I., Hur, S., Kim, S., and Jeong, K. (2006). The Influence of Pig Breeds on Qualities of Loin. *J. Anim. Sci. Techn* 48 (5), 747–758. doi:10.5187/JAST.2006.48.5.747

Johnson, R. K. (1981). Crossbreeding in Swine: Experimental Results. *J. Anim. Sci.* 52 (4), 906–923. doi:10.2527/jas1981.524906x

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 (1), 237–249. doi:10.1534/genetics.111.130922

Kim, S. G., Bae, H. H., Son, J. Y., Shin, J. S., and Ha, G. H. (2020a). *The Trends of Consumption of Pork Meat in Korean 2019 NIAS, RDA, Republic of Korea*. Wanju, South Korea: NIAS.

Kim, Y.-M., Seong, H.-S., Lee, J.-J., Son, D.-H., Kim, J.-S., Sa, S.-J., et al. (2020b). Genome-wide Investigation of a Korean Synthetic Breed, Woori-Heukdon Using the Illumina PorcineSNP60K BeadChip. *Genes Genom* 42 (12), 1443–1453. doi:10.1007/s13258-020-01008-5

Kim, Y. M., Cho, G. H., Park, J. C., Kim, D. W., Hong, J. K., Sa, S. J., et al. (2016). *Preparing Method of Synthetic Pig by a Cross between Duroc and Korean Native Pig, and Synthetic Pig Using the Same*. Republic of Korea patent application.

Lee, S. H., Seo, D. W., Cho, E. S., Choi, B. H., Kim, Y. M., Hong, J. K., et al. (2020). Genetic Diversity and Ancestral Study for Korean Native Pigs Using 60K SNP Chip. *Animals* 10 (5), 760. doi:10.3390/ani10050760

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.* 104 (50), 19942–19947. doi:10.1073/pnas.0710021104

Li, X., Yang, S., Dong, K., Tang, Z., Li, K., Fan, B., et al. (2017). Identification of Positive Selection Signatures in Pigs by Comparing Linkage Disequilibrium Variances. *Anim. Genet.* 48 (5), 600–605. doi:10.1111/age.12574

Lopes, M. S., Silva, F. F., Harlizius, B., Duijvesteijn, N., Lopes, P. S., Guimarães, S. E., et al. (2013). Improved Estimation of Inbreeding and Kinship in Pigs Using Optimized SNP Panels. *BMC Genet.* 14, 92. doi:10.1186/1471-2156-14-92

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., et al. (2015). Analysis of Runs of Homozygosity and Their Relationship with Inbreeding in Five Cattle Breeds Farmed in Italy. *Anim. Genet.* 46 (2), 110–121. doi:10.1111/age.12259

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

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2021). Runs of Homozygosity in European Populations. *The American Journal of Human Genetics* 83 (3), 359–372.

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

Muñoz, M., Bozzi, R., García-Casco, J., Núñez, Y., Ribani, A., Franci, O., et al. (2019). Genomic Diversity, Linkage Disequilibrium and Selection Signatures in European Local Pig Breeds Assessed with a High Density SNP Chip. *Sci. Rep.* 9 (1), 13546. doi:10.1038/s41598-019-49830-6

Nagaoka, K., Fujii, K., Zhang, H., Usuda, K., Watanabe, G., Ivshina, M., et al. (2016). CPEB1 Mediates Epithelial-To-Mesenchyme Transition and Breast Cancer Metastasis. *Oncogene* 35 (22), 2893–2901. doi:10.1038/onc.2015.350

Nishimura, Y., Kano, K., and Naito, K. (2010). Porcine CPEB1 Is Involved in Cyclin B Translation and Meiotic Resumption in Porcine Oocytes. *Anim. Sci. J.* 81 (4), 444–452. doi:10.1111/j.1740-0929.2010.00755.x

Park, B. Y., Kim, N. K., Lee, C. S., and Hwang, I. H. (2007). Effect of Fiber Type on Postmortem Proteolysis in Longissimus Muscle of Landrace and Korean Native Black Pigs. *Meat Sci.* 77 (4), 482–491. doi:10.1016/j.meatsci.2007.04.022

Peripolli, E., Metzger, J., de Lemos, M. V. A., Stafuzza, N. B., Kluska, S., Olivieri, B. F., et al. (2018). Autozygosity Islands and ROH Patterns in Nellore Lineages: Evidence of Selection for Functionally Important Traits. *BMC Genomics* 19 (1), 680. doi:10.1186/s12864-018-5060-8

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

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

Schiavo, G., Bovo, S., Bertolini, F., Dall'Olio, S., Costa, L. N., Tinarelli, S., et al. (2020). Runs of Homozygosity Islands in Italian Cosmopolitan and Autochthonous Pig Breeds Identify Selection Signatures in the Porcine Genome. *Livestock Sci.* 240, 104219. doi:10.1016/j.livsci.2020.104219

Schiavo, G., Bovo, S., Muñoz, M., Ribani, A., Alves, E., Araújo, J. P., et al. (2021). Runs of Homozygosity Provide a Genome Landscape Picture of Inbreeding and Genetic History of European Autochthonous and Commercial Pig Breeds. *Anim. Genet.* 52 (2), 155–170. doi:10.1111/age.13045

Valluzzi, C., Rando, A., Macciotta, N. P. P., Gaspa, G., and Di Gregorio, P. (2021). The Nero Lucano Pig Breed: Recovery and Variability. *Animals* 11 (5), 1331. doi:10.3390/ani11051331

Wang, L., Mu, Y., Xu, L., Li, K., Han, J., Wu, T., et al. (2019). Genomic Analysis Reveals Specific Patterns of Homozygosity and Heterozygosity in Inbred Pigs. *Animals* 9 (6), 314. doi:10.3390/ani9060314

Weedon, M. N., Lango, H., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., et al. (2008). Genome-wide Association Analysis Identifies 20 Loci that Influence Adult Height. *Nat. Genet.* 40 (5), 575–583. doi:10.1038/ng.121

Wu, F., Sun, H., Lu, S., Gou, X., Yan, D., Xu, Z., et al. (2020). Genetic Diversity and Selection Signatures within Diannan Small-Ear Pigs Revealed by Next-Generation Sequencing. *Front. Genet.* 11, 733. doi:10.3389/fgene.2020.00733

Xu, L., Zhao, G., Yang, L., Zhu, B., Chen, Y., Zhang, L., et al. (2019a). Genomic Patterns of Homozygosity in Chinese Local Cattle. *Sci. Rep.* 9 (1), 16977. doi:10.1038/s41598-019-53274-3

Xu, Z., Sun, H., Zhang, Z., Zhao, Q., Olasege, B. S., Li, Q., et al. (2019b). Assessment of Autozygosity Derived from Runs of Homozygosity in Jinhua Pigs Disclosed by Sequencing Data. *Front. Genet.* 10, 274. doi:10.3389/fgene.2019.00274

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fgene.2022.812753

Sequencing Reveals Population Structure and Selection Signatures for Reproductive Traits in Yunnan Semi-Fine Wool Sheep (*Ovis aries*)

Yi Guo[1], Jiachong Liang[2], Chunrong Lv[2], Yi Wang[1], Guoquan Wu[2], Xiangdong Ding[1] and Guobo Quan[2]**

[1]*Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture and Rural Affairs, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China,* [2]*Yunnan Provincial Genebank of Livestock and Poultry Genetic Resources, Yunnan Provincial Engineering Laboratory of Animal Genetic Resource Conservation and Germplasm Enhancement, Yunnan Animal Science and Veterinary Institute, Kunming, China*

OPEN ACCESS

Edited by:
El Hamidi Hay,
Agricultural Research Service (USDA), United States

Reviewed by:
Ran Li,
Northwest A&F University, China
Ottmar Distl,
University of Veterinary Medicine Hannover, Germany
Natalia A. Zinovieva,
L. K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia
Bang Liu,
Huazhong Agricultural University, China

***Correspondence:**
Xiangdong Ding
xding@cau.edu.cn
Guobo Quan
waltq20020109@163.com

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

Received: *10 November 2021*
Accepted: *18 February 2022*
Published: *07 March 2022*

Citation:
Guo Y, Liang J, Lv C, Wang Y, Wu G, Ding X and Quan G (2022) Sequencing Reveals Population Structure and Selection Signatures for Reproductive Traits in Yunnan Semi-Fine Wool Sheep (Ovis aries). Front. Genet. 13:812753. doi: 10.3389/fgene.2022.812753

Yunnan semi-fine wool sheep are among the most important cultivated sheep breeds in China. However, their population structure, genetic characteristics and traits of interest are poorly studied. In this study, we systematically studied the population characteristics and selection signatures of 40 Yunnan semi-fine wool sheep using SNPs obtained from whole-genome resequencing data. A total of 1393 Gb of clean data were acquired. The mapping rate against the reference genome was 91.23% on average (86.01%–92.26%), and the average sequence depth was 9.51X. After filtering, 28,593,198 SNPs and 4,725,259 indels with high quality were obtained. The heterozygosity rate, inbreeding coefficient and effective population size of the sheep were calculated to preliminarily explore their genetic characteristics. The average heterozygosity rate was 0.264, the average inbreeding coefficient was 0.0099, and the effective population size estimated from the heterozygote excess (HE) was 242.9. Based on the Tajima's D and integrated haplotype score (iHS) approaches, 562 windows and 11,356 core SNPs showed selection signatures in the Yunnan semi-fine wool sheep population. After genome annotation and gene enrichment analysis, we found traces of early domestication in sensory organs, behavioural activity and the nervous system as well as adaptive changes in reproductive and wool traits under selection in this population. Some selected genes related to litter size, including *FSHR*, *BMPR1B* and *OXT*, were identified as being under selection. Specific missense mutations of the *FSHR* gene that differed from the reference genome were also identified in the population, and we found some SNP variations that may affect litter size. Our findings provide a theoretical basis for the conservation and utilization of Yunnan semi-fine wool sheep. Furthermore, our results reveal some changes common to sheep after domestication and provide a new opportunity to investigate the genetic variation influencing fecundity within a population evolving under artificial selection.

Keywords: Yunnan semi-fine wool sheep, sequencing, selection signatures, population structure, reproduction

INTRODUCTION

Yunnan semi-fine wool sheep are among the most important cultivated breeds in Yunnan Province, China. This breed has excellent wool quality, high adaptability and strong robustness. Currently, it is widely used to produce wool and meat products (Wang, 2009). From 1954 to 1971, to improve the wool fineness of ZhaoTong sheep, they were successively crossbred with Rambouillet hybrid sheep, Caucasian sheep, XinJiang fine-wool sheep, and New Zealand Romney sheep, among others. In 1977, Lincoln sheep were introduced to improve wool length, and in 1979, the ideal cross combination was obtained (National Commission On Livestock and Poultry Genetic Resources, 2011). In 2000, this breed was approved by the Chinese Livestock and Poultry Breed Approval Committee, becoming the first approved breed of coarse semi-fine wool sheep in China. Due to the long-term cultivation and growth of these sheep in alpine regions, they have developed high adaptability to cold climates, high altitudes and hypoxic conditions. In general, Yunnan semi-fine wool sheep are medium in body size, and their wool quality is comparable to that of Romney and Lincoln sheep. Due to its competitive advantages and rapid adaptation to the local environment, the population of Yunnan semi-fine wool sheep has been continuously expanded, with the population size exceeding 100,000 individuals in 2005 (Yuan and Sun, 2014). However, its conception rate and lamb survival rate are not high, and more than 90% of ewes give birth to one lamb per parity. Therefore, the improvement of reproductive traits is important for Yunnan semi-fine wool sheep.

Selection signatures are the imprints left on the genome of a species under long-term natural and artificial selection during the process of evolution. Genome selection signatures are an effective tool for studying the adaptability of domestic animals in different natural environments and exploring the genetic mechanisms underlying phenotypic differences. Various methods are currently used to assess genetic diversity and detect selection signatures. These methods can be divided into three categories depending on the different types of examined genomic information: methods based on the allele frequency spectrum, e.g., Tajima's D (Tajima, 1989) and the composite likelihood ratio (CLR) (Nielsen et al., 2005); methods based on linkage disequilibrium (LD), e.g., extended haplotype homozygosity (EHH) (Sabeti et al., 2002) and the integrated haplotype score (iHS) (Voight et al., 2006); and methods based on population differentiation, e.g., the fixation index (FST) (Weir and Cockerham, 1984). In this study, we employed the iHS and Tajima's D tests to detect selection signatures within the population.

With the development of sequencing technologies and the improvement of biological analysis methods, selection signature detection has been used to mine the imprints left by domestication and artificial selection and the genes associated with certain important traits (Porto-Neto et al., 2014; Liu et al., 2016; E et al., 2019; Abied et al., 2020; Zhao et al., 2020; Zhong et al., 2020; Zhu et al., 2020). However, due to certain limitations, it is difficult to obtain a complete picture of the germplasm and selection characteristics of indigenous sheep breeds in China. Therefore, the objectives of this study were to explore the population characteristics and genetic structure of Yunnan semi-fine wool sheep, to screen selection regions related to important traits and to identify variants by using genome-wide selection signature detection. Furthermore, the present study will provide a theoretical basis for the breeding and conservation of Yunnan semi-fine wool sheep.

MATERIALS AND METHODS

Ethics Statement

The ethical committee of the Yunnan Animal Science and Veterinary Institute (Kunming city, Yunnan Province, China) approved all experiments in this study (201909006). In addition, during the study, all authors strictly complied with the Regulations on the Administration of Laboratory Animals (Order No. 2 of the State Science and Technology Commission of the People's Republic of China, 1988) and the Regulations on the Administration of Experimental Animals of Yunnan Province (the Standing Committee of Yunnan Provincial People's Congress 2007.10). There was no use of human participants, data or tissues.

Animals and Phenotypes

A total of 40 female Yunnan semi-fine wool sheep with similar body weights (approximately 50 kg) and ages (3 years old) were sampled from the Laishishan sheep farm in Qiaojia City, Yunnan Province (N26°54′55.29″, E102°55′40.02″), which is one representative Yunnan semi-fine wool sheep farm in Yunnan Province, and genetic exchange (rams or sperm) was carried out with other Yunnan semi-fine wool sheep farms. The samples were unrelated according to the pedigree. Based on their litter sizes in two successive parities, these ewes were classified into two groups: ewes with a litter size of 1 (20 individuals) and those with a litter size of 2 (20 individuals). In addition, the sequencing data of 4 Zhaotong sheep and 3 Romney sheep were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database to further explore the population structure of Yunnan semi-fine wool sheep.

Sample Collection, DNA Extraction, and Whole-Genome Sequencing

Ear tissues of the selected sheep were collected, transferred to sterile centrifuge tubes, and stored in a freezer at −80°C. DNA was extracted from the sample of each individual and analysed following standard experimental procedures to guarantee quality. Sequencing libraries of each individual were constructed by Beijing BerryGenomics Biotechnology Co., Ltd., using the Illumina NovaSeq6000 PE150™ platform.

Genomic Data Processing and SNP Calling

Trimmomatic software (version 0.38) (Bolger et al., 2014) was applied to remove adaptors and low-quality reads. After quality control, the reads of each sample were mapped to the GCF_002742125.1_Oar_rambouillet_v1.0 genome with the

'mem' algorithm of BWA software (version 0.7.17) (Li and Durbin, 2009). The bam and sorted bam files were generated using SAMtools (version 1.9) (Li et al., 2009). Then, the Genome Analysis Toolkit (GATK, version 3.7.0) (McKenna et al., 2010) pipeline was used to call SNPs and indels.

To ensure high quality of the variations for the following analysis, BCFtools (version 1.10.2) (Danecek and McCarthy, 2017) was used to filter the minimum or maximum number of alleles (bcftools view -m 2 -M 2) from autosomes and chromosome X. In addition, SNPs with minor allele frequencies lower than 0.01 or missing rates greater than 0.2 were excluded by VCFtools-0.1.16 (Danecek et al., 2011). To identify the functions of the variants, SnpEff 4.3 software (Cingolani et al., 2012) was utilized to annotate the filtered SNPs.

Population Structure Analysis

Based on the filtered SNPs of Yunnan semi-fine wool sheep and their parental sheep breeds, principal component analysis (PCA), admixture analysis and LD decay analysis were performed to explore the population structure of Yunnan semi-fine wool sheep. For PCA, PLINK (v1.90b4) software (Purcell et al., 2007) was used to generate the input file (.bed.bim.fam). gcta64 software (v1.26.0) (Yang et al., 2011) was applied to construct a genetic relationship matrix (–make-grm) and calculate principal components. For admixture analysis, ADMIXTURE (version 1.3.0) (Alexander and Lange, 2011) was employed to carry out unsupervised hierarchical clustering. PopLDdecay software (version 3.41) was used to calculate LD decay, LD was evaluated on the basis of the squared coefficient of correlation (r^2) between loci (Zhang et al., 2019), and the parameters were set as follows: -MaxDist 300 -OutType 1. All of the results were presented in graphs created using R software.

Heterozygosity Rate, Inbreeding Coefficients and Effective Population Size

As important population parameters, the heterozygosity rate, inbreeding coefficient and effective population size were calculated in this study. PLINK (v1.90b4) software (Purcell et al., 2007) was used to estimate the heterozygosity rate, which was calculated with the formula $\frac{N(NM)-O(HOM)}{N(NM)}$, where $O(HOM)$, $E(HOM)$ and $N(NM)$ represent the observed number of homozygotes, expected number of homozygotes and number of nonmissing genotypes, respectively, which were obtained by using PLINK. The inbreeding coefficients (F_{ROH}) of individuals were estimated based on runs of homozygosity (ROHs). PLINK software was used to detect ROHs with the following parameters: 1) minimum density of 1 SNP per 100 kb, 2) minimum ROH length of 1000 kb, 3) 1 heterozygote allowed per window, 4) 5 missing calls allowed per window, 5) minimum number of SNPs in a window of 50, and 6) 0.05 missing calls allowed per window. The following formula was used to calculate F_{ROH}:

$$F_{ROH} = \frac{\sum_i ROH_i}{L_{auto}} \quad \text{(McQuillan et al., 2008)}$$

where $\sum ROH_i$ is the total ROH length on autosomes and L_{auto} is the total length of autosomes. The effective population size (Ne) can indicate dynamic changes in population size, and Ne was estimated according to the random mating model based on the heterozygote excess (HE) method using the default parameters in NeEstimator (version 2.1) software (Do et al., 2014). NeEstimator calculates Ne as follows:

$$Ne = \frac{1}{2D} + \frac{1}{2(D+1)} \quad \text{(Pudovkin et al., 1996)}$$

where Ne is the effective population size and D is Selander's index, calculated as follows: $D_j(i) = \frac{H_j^{obs}(i) - H_j^{\exp}(i)}{H_j^{\exp}(i)}$. $H_j^{\exp}(i)$ is the expected heterozygote frequency of allele i at locus j. The actual D value is the weighted mean value, calculated as follows: $D = \frac{\sum w_j D_j}{\sum_{j=1}^{k}(n_j - 1)\sqrt{N_j}}$, where w_j is the weight of locus j and is calculated as $w_j = \sqrt{N_j}\,\frac{n_j - 1}{n_j}$, where N_j is the sample size, and n_j is the number of alleles at locus j.

Detection of Selection Signatures

Selection signature detection within populations can reveal selection dynamics and the history of evolution. In this study, two metrics, Tajima's D (Tajima, 1989) and the iHS (Voight et al., 2006), were utilized to identify selection signatures in the whole genome of Yunnan semi-fine wool sheep. Tajima's D is based on neutral mutation theory. When Tajima's D is equal to 0, it indicates that the population is in a neutral evolutionary state, whereas Tajima's D values lower or greater than 0 indicate that the population has experienced purifying selection events/population expansion or balancing selection, respectively. Tajima's D values were calculated in nonoverlapping 50-kb sliding windows. Only the windows with the 1% highest and the 1% lowest Tajima's D values were identified as subject to selection. Different from the single locus-based Tajima's D metric, the iHS is mainly based on haplotypes. A negative iHS value indicates that the mutated allele may be affected by positive selection. In this study, haplotypes were first constructed by using SHAPEIT (Delaneau and Marchini, 2014), and iHS statistics were then calculated using Selscan software (version 1.2.0) (Szpiech et al., 2014). Thereafter, the iHS values were standardized to follow a standard normal distribution, and the statistics exceeding the 0.1% quantile ($|Z(iHS)\ value| = 3$) indicated that the core SNP experienced selection.

Functional Annotation of Selected Regions

Based on the core SNPs or windows under selection detected based on the iHS and Tajima's D analyses, selected regions were defined for the bioinformatics analysis. From each core SNP detected based on iHS analysis, the region was extended 20 kb upstream and downstream to define the selected region. Under the Tajima's D approach, each selected 50-kb sliding window was extended 150 kb upstream and downstream to define the selected region. Gene information in the selected regions was obtained using the Biomart database (http://asia.ensembl.org/biomart/martview/). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses were performed for further gene function analysis

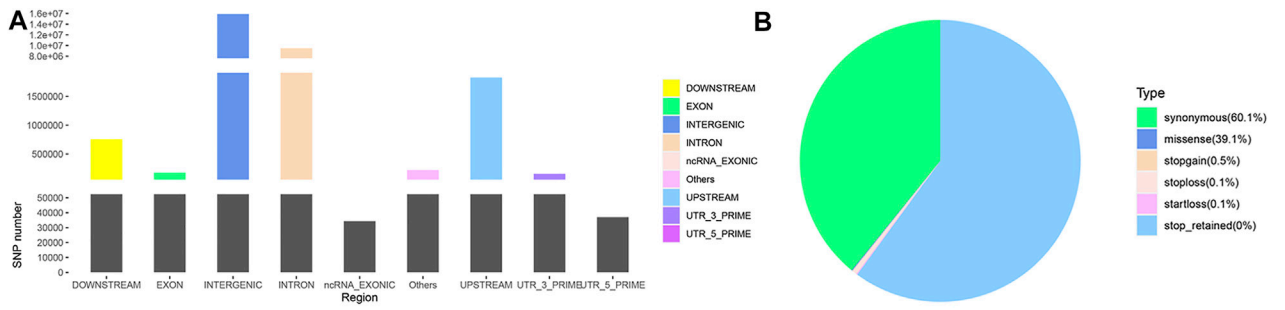

FIGURE 1 | SNP annotation of Yunnan semi-fine wool sheep. **(A)** Genomic annotation of SNPs according to SnpEff. "Others" contained a few other "splice site acceptor," "splice site donor" and "splice site region" sites. **(B)** Pie chart of SNPs annotated in exonic regions.

using DAVID Bioinformatics Resources 6.8 (Huang et al., 2009) (https://david.ncifcrf.gov/). The GO terms and KEGG pathways with P values less than 0.05 were considered significant.

Mutation Analysis of Selected Genes Relevant to Reproductive Traits

In this study, 40 Yunnan semi-fine wool sheep were classified into two groups on the basis of litter size: 1 lamb (20 individuals) or 2 lambs (20 individuals) in two successive parities. To investigate whether SNP mutations or frequency differences existed between the two groups, three significantly selected genes (*FSHR*, *BMPR1B*, and *OXT*) related to reproductive traits were studied. PLINK (v1.90b4) software was employed to perform a standard case (individuals with two lambs)/control (individuals with one lamb) association analysis using Fisher's exact test with the option plink --file gene --fisher.

RESULTS

Genetic Variants in Yunnan Semi-Fine Wool Sheep

In the present study, a total of 40 female Yunnan semi-fine wool sheep were sequenced. After filtering out contaminated reads, including adaptors, low-quality reads and reads with an N ratio greater than 10%, 1393 Gb of clean data were acquired, and 35,178,304 raw SNPs and 5,256,453 raw indels were obtained. The mapping rate against the reference genome was 91.23% on average (86.01%–92.26%), and the average sequence depth was 9.51X (8.09X~12.10X) (see **Supplementary Material S1** for details). After filtering, 28,593,198 SNPs and 4,725,259 indels with high quality were retained.

Following annotation, 27,623,037 SNPs on autosomes and 970,161 SNPs on chromosome X (**Supplementary Material S2**) were obtained. These SNPs were partitioned according to their locations: intergenic (15,897,566; 56.03%), intronic (9,491,317; 33.45%), exonic (177516; 0.63%) and other gene regulatory regions (**Figure 1A**). Among the SNPs in exonic regions, as shown in **Figure 1B**, synonymous mutations constituted the overwhelming majority (60.07%), followed by missense mutations (39.14%), and the other types of observed mutations included stop-gain, stop-loss, and start-loss mutations (less than 1%). The transition/transversion (Ts/Tv) ratio was 2.4721.

Genetic Diversity of the Yunnan Semi-Fine Wool Sheep Population

To describe the genetic diversity of the Yunnan semi-fine wool sheep population, two metrics were used. For the whole population, the average heterozygosity rate and inbreeding coefficient were 0.264 (0.231–0.374) and 0.0099 (0.0004–0.0239), respectively (see **Supplementary Material S4** for details), indicating that the heterozygosity rates and inbreeding coefficients varied greatly within the population. The effective population size estimated from the HE of the Yunnan semi-fine wool sheep population using the lowest allele frequency of 0.05 as the threshold was 242.9 (95% CI: 232.6–254.2).

Population Structure Analysis

The population structure is illustrated in **Figure 2**. ADMIXTURE analysis indicated that the ancestors of Yunnan semi-fine wool sheep included Romney and Zhaotong sheep, with Romney sheep contributing more (K = 2); **Figure 2A** also demonstrates other ancestors of Yunnan semi-fine wool sheep (K = 3). Moreover, PCA (**Figure 2B**) further showed that the population of Yunnan semi-fine wool sheep was aggregated and separated from Romney and Zhaotong sheep, indicating that it is a new breed. LD decay analysis of two groups with litter sizes of 1 or 2 lambs (**Figure 2C**) confirmed the absence of obvious population stratification within Yunnan semi-fine wool sheep, and the difference between litter sizes might be due to ongoing selection.

Detection of Genome-Wide Signatures of Selection

As shown in **Figure 3**, regardless of whether the iHS or Tajima's D approach was used, the observed selection signatures were distributed across all chromosomes. Under the iHS approach, the 0.1% quantile ($|Z(iHS)\ value| = 3$) was used as the threshold to extract selected core SNPs. A total of 11,356 core SNPs showing the strongest selection were identified, accounting for 1% of all

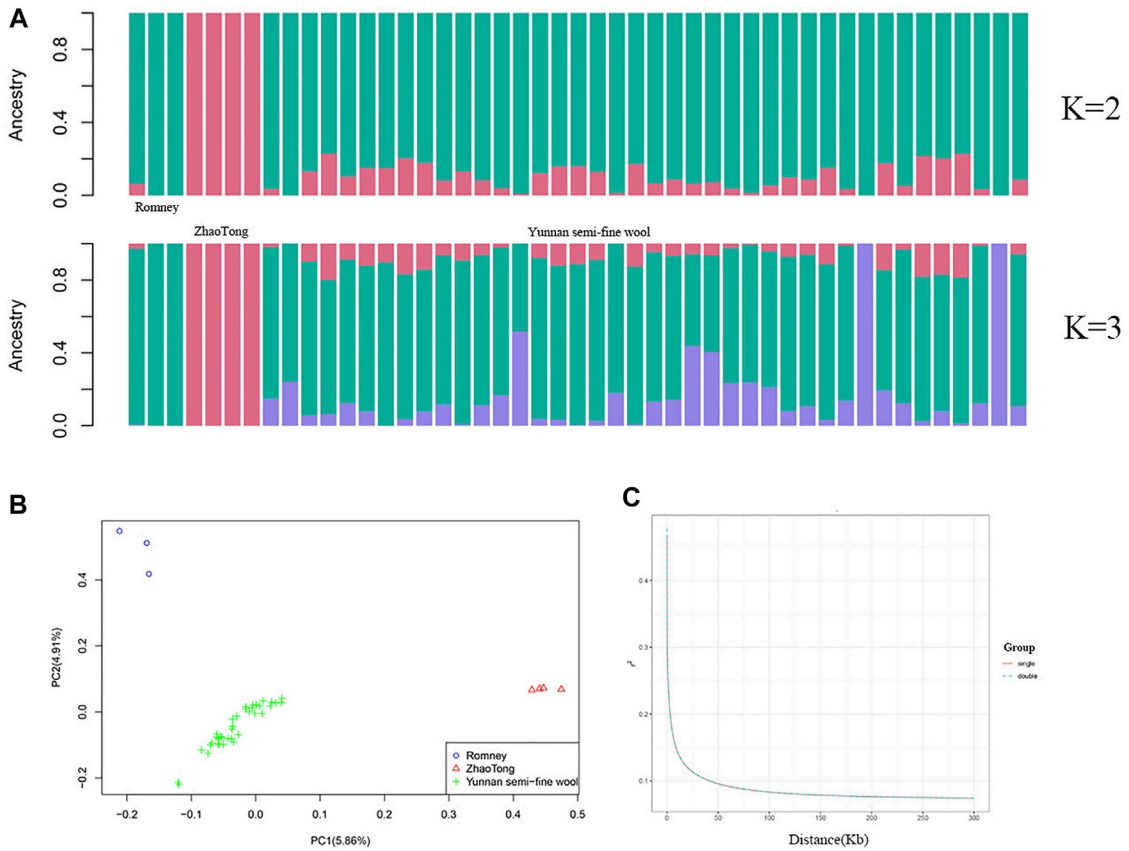

FIGURE 2 | Genetic differentiation of 40 Yunnan semi-fine wool sheep. **(A)** ADMIXTURE analysis of Yunnan semi-fine wool sheep and their parental sheep breeds (Romney and Zhaotong sheep). **(B)** PCA of 40 Yunnan semi-fine wool sheep and their parental sheep breeds using 710,412 common SNPs located on autosomes. **(C)** LD decay curves (based on r^2) of two groups (single or double lambs).

selected SNPs. The iHS values ranged from −7.69 to 8.26 (**Figures 3A,C**). Considering the LD among SNPs, the selected SNPs within each 50-kb interval were combined into one selected region. In total, 3512 selected regions were identified. Chromosomes 1, 2, 3 and 4 harboured the most selected regions, i.e., 422, 323, 297 and 206 selected regions, respectively (**Figure 3F**).

A total of 56,115 nonoverlapping 50-kb windows were assessed, and only the top and bottom 1% of the windows with high Tajima's D values were determined to be regions of selective sweeps (**Figures 3B,D**). A total of 562 windows showing the strongest selection and 449 equivalent selected regions were identified by combining the selected windows. Selected regions were identified on all chromosomes (**Figure 3F**). Among all chromosomes, chromosomes 1, 2, 3 and X harboured the most selected windows, i.e., 42, 41, 41 and 41 windows, respectively. The most significant regions were located on chromosome 20 from 16.70–16.85 Mb, with an average Tajima's D value of −2.77432, and this region harboured 495 SNPs. Five known genes were identified within this putative selective sweep region, namely, *CUL7*, *KLC4*, *MEA1*, *MRPL2* and *PTK7*, which may regulate cell activity and microtubule motor activity.

Functional Enrichment of Candidate Genes

According to the two methods of selection signature detection, 1856 genes were found in the 20-kb regions upstream and downstream of the core SNPs according to the iHS test, and 1753 selected genes were obtained in the 350-kb intervals upstream and downstream of the selected windows based on the Tajima's D test. We combined the genes obtained via the two methods described above, and a total of 108 genes were identified (**Figure 3E**). To evaluate the functions of the selected genes obtained by the two methods, gene annotation and enrichment analyses (GO and KEGG) were conducted.

For the iHS, the enrichment analysis identified a total of 53 significant GO terms (**Supplementary Material S3**; **Figure 1**), namely, 13 cellular component terms, 17 molecular function terms, and 23 biological process terms. Notably, most of the selected genes were enriched in the categories of immune system processes, intracellular signal transduction and neural synapses. In the KEGG analysis results, 19 pathways were annotated, as shown in **Figure 4A**. Some known reproduction-related, domestication-related and wool-related pathways were found to be significantly enriched, including the vascular smooth muscle contraction, calcium signalling, salivary secretion, gastric acid secretion, long-term depression and long-term potentiation pathways.

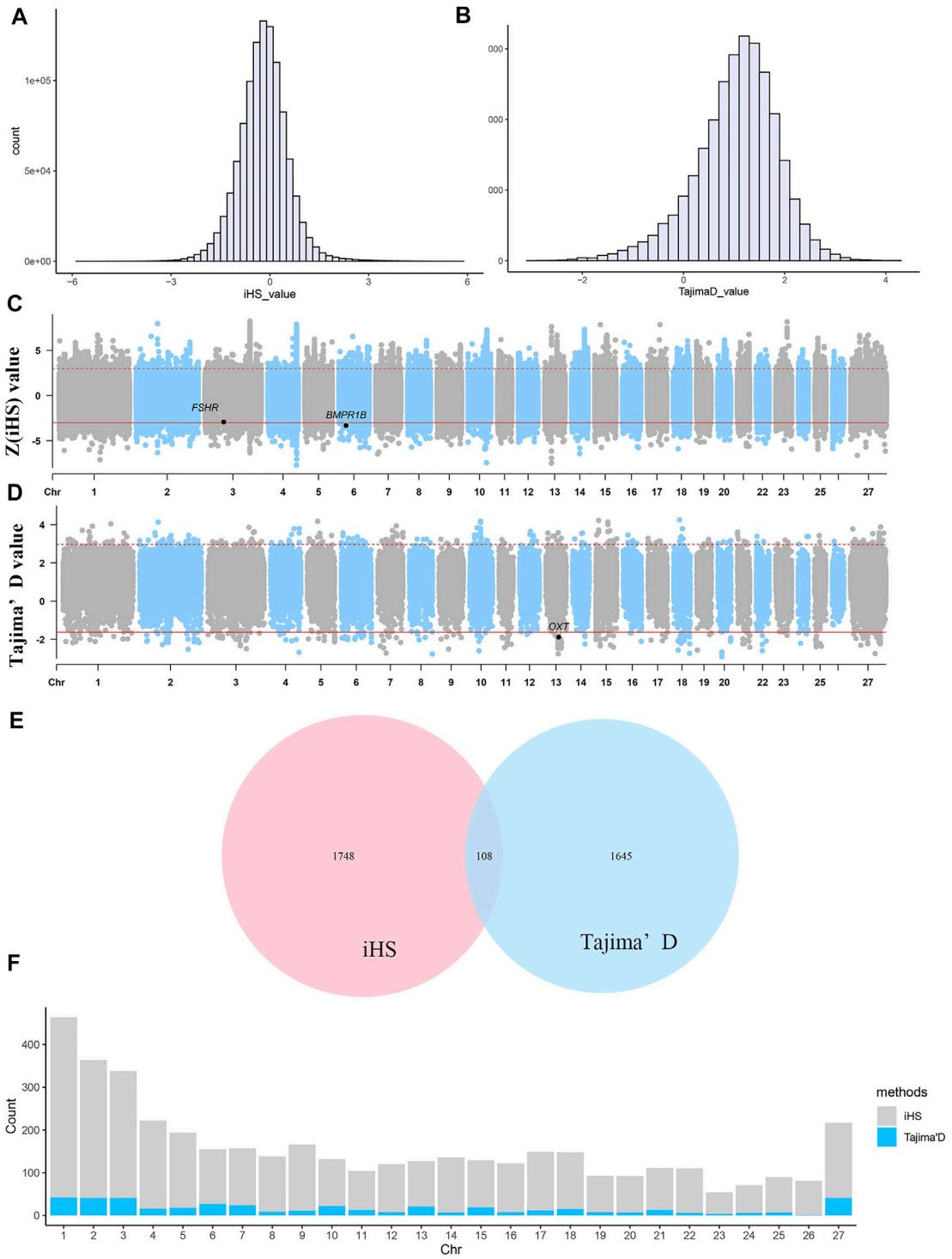

FIGURE 3 | Selective sweep analysis of 40 Yunnan semi-fine wool sheep. **(A)** Distribution of iHS values for selected regions. **(B)** Distribution of Tajima's D values for selected regions. **(C)** Manhattan plot of *Z* (*iHS*) values of sheep. The dashed line denotes a threshold of |*Z* (*iHS*) *value*| = 3 (*p* = 0.001). **(D)** Manhattan plot of Tajima's D values. The black line is the "suggestive line": the top 1% of values are >2.97345, and the bottom 1% of values are < −1.63182. **(E)** The 108 common selected genes detected by both the iHS and Tajima's D approaches. **(F)** The distribution of the selected SNPs on each chromosome obtained by the iHS and Tajima's D approaches. In the figure, "27" on the X-axis represents chromosome X.

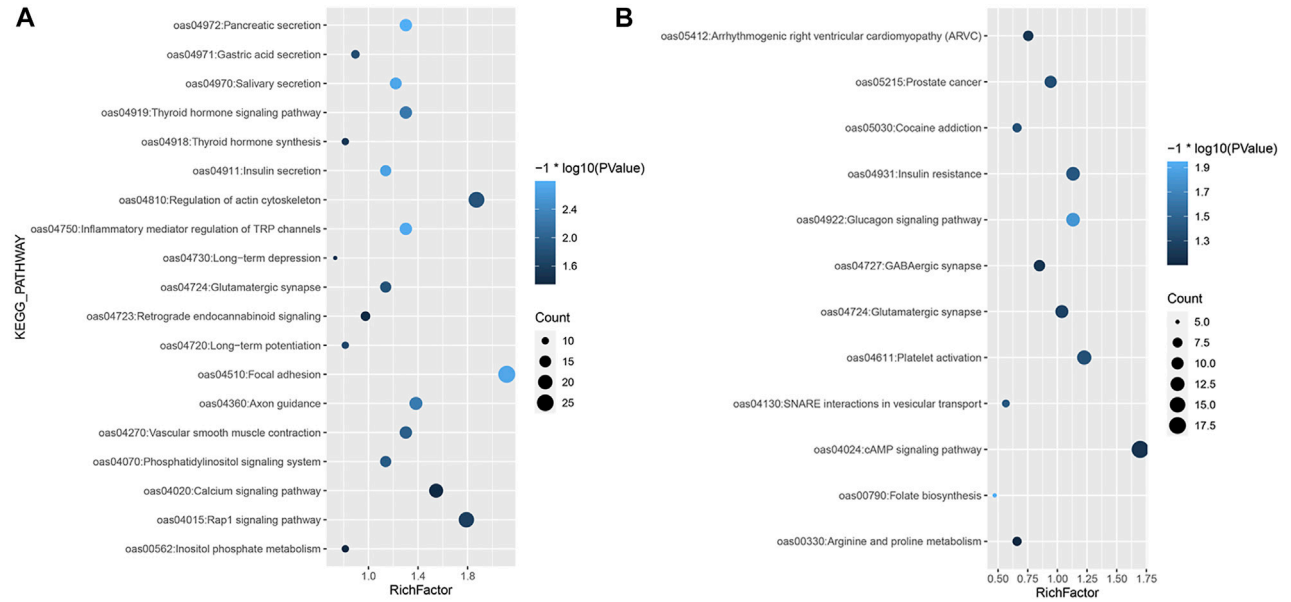

FIGURE 4 | KEGG enrichment analysis of selected genes. **(A)** KEGG analysis of the selected genes based on the **(A)** iHS and **(B)** Tajima's D approaches.

TABLE 1 | GO terms or KEGG pathways associated with domestication, reproduction and immune system processes.

		Category	Term	Count	P Value
Domestication	iHS	GOTERM_BP_DIRECT	Synapse organization	5	0.0499
		KEGG_PATHWAY	Salivary secretion	15	0.0020
		KEGG_PATHWAY	Glutamatergic synapse	14	0.0148
		KEGG_PATHWAY	Gastric acid secretion	11	0.0170
		KEGG_PATHWAY	Long-term potentiation	10	0.0197
		KEGG_PATHWAY	Long-term depression	9	0.0329
		KEGG_PATHWAY	Retrograde endocannabinoid signalling	12	0.0418
	Tajima's D	GOTERM_MF_DIRECT	Olfactory receptor activity	36	0.0130
		GOTERM_BP_DIRECT	Positive regulation of dendrite morphogenesis	3	0.0446
		KEGG_PATHWAY	Glutamatergic synapse	11	0.0526
		KEGG_PATHWAY	GABAergic synapse	9	0.0650
Reproduction	iHS	KEGG_PATHWAY	Vascular smooth muscle contraction	16	0.0107
	Tajima's D	KEGG_PATHWAY	Folate biosynthesis	5	0.0112
Immune system process	iHS	KEGG_PATHWAY	Inflammatory mediator regulation of TRP channels	16	0.0018
		KEGG_PATHWAY	Thyroid hormone signalling pathway	16	0.0058
		KEGG_PATHWAY	Thyroid hormone synthesis	10	0.0306
Wool	iHS	GOTERM_BP_DIRECT	Negative regulation of BMP signalling pathway	7	0.0393
		GOTERM_BP_DIRECT	Positive regulation of peptidyl-tyrosine phosphorylation	9	0.0355
	Tajima's D	GOTERM_BP_DIRECT	Hyaluronan catabolic process	3	0.0446
		GOTERM_BP_DIRECT	Regulation of protein phosphorylation	6	0.0041
		GOTERM_MF_DIRECT	Protein tyrosine phosphatase activity	10	0.0378

Among the results based on Tajima's D method, 18 significant GO terms were identified (**Supplementary Material S3**; **Figure 2**), including olfactory receptor activity, immune process and dendrite morphogenesis terms. A total of 12 KEGG pathways were obtained (**Figure 4B**), including folate biosynthesis related to reproductive traits and the GABAergic synapse pathway related to inhibition of the nervous system.

Overall, our results revealed traces of early domestication in Yunnan semi-fine wool sheep, such as modifications in sensory organs, behavioural activity and the nervous system, as well as adaptive changes in economic traits, including reproductive and wool traits, under artificial selection (**Table 1**).

Mutation and Association Analysis of Genes Related to Reproduction

Genes related to reproductive traits, such as follicle-stimulating hormone receptor (*FSHR*), bone morphogenetic protein receptor

TABLE 2 | Mutations in the *BMPR1B* and *FSHR* genes in exonic regions of the Yunnan semi-fine wool sheep genome.

Gene	Chromosome	Position	cDNA variation	Position of mRNA	Amino acid variation	Mutation type	db SNP	MAF
BMPR1B	6	33998156	1059A > C	1059	Arg353Arg	Synonymous variant	rs429416173	0.43
		33992536	1416C > A	1416	Thr472Thr	Synonymous variant	rs413854373	0.28
		34009717	825C > T	825	Ser275Ser	Synonymous variant	rs598261578	0.04
		34009732	810T > C	810	Tyr270Tyr	Synonymous variant	rs159952533	0.41
		34011008	543A > G	543	Thr181Thr	Synonymous variant	rs427897187	0.33
		34011011	540G > A	540	Arg180Arg	Synonymous variant	rs408447622	0.08
		34051987	52G > A	52	Ala18Thr	Missense variant	rs605658565	0.16
FSHR	3	80789093	696C > T	696	Ser232Ser	Synonymous variant	rs412817989	0.41
		80794664	1113C > T	1113	Phe371Phe	Synonymous variant	rs416291965	0.11
		80605490	28G > A	28	Ala10Thr	Missense variant	rs399253678	0.25
		80789089	692G > A	692	Arg231His	Missense variant	rs399612350	0.40
		80789180	783T > G	783	Phe261Leu	Missense variant	rs422112895	0.04
		80789251	854C > T	854	Thr285Ile	Missense variant	rs398233545	0.06
OXT	13	—	—	—	—	—		

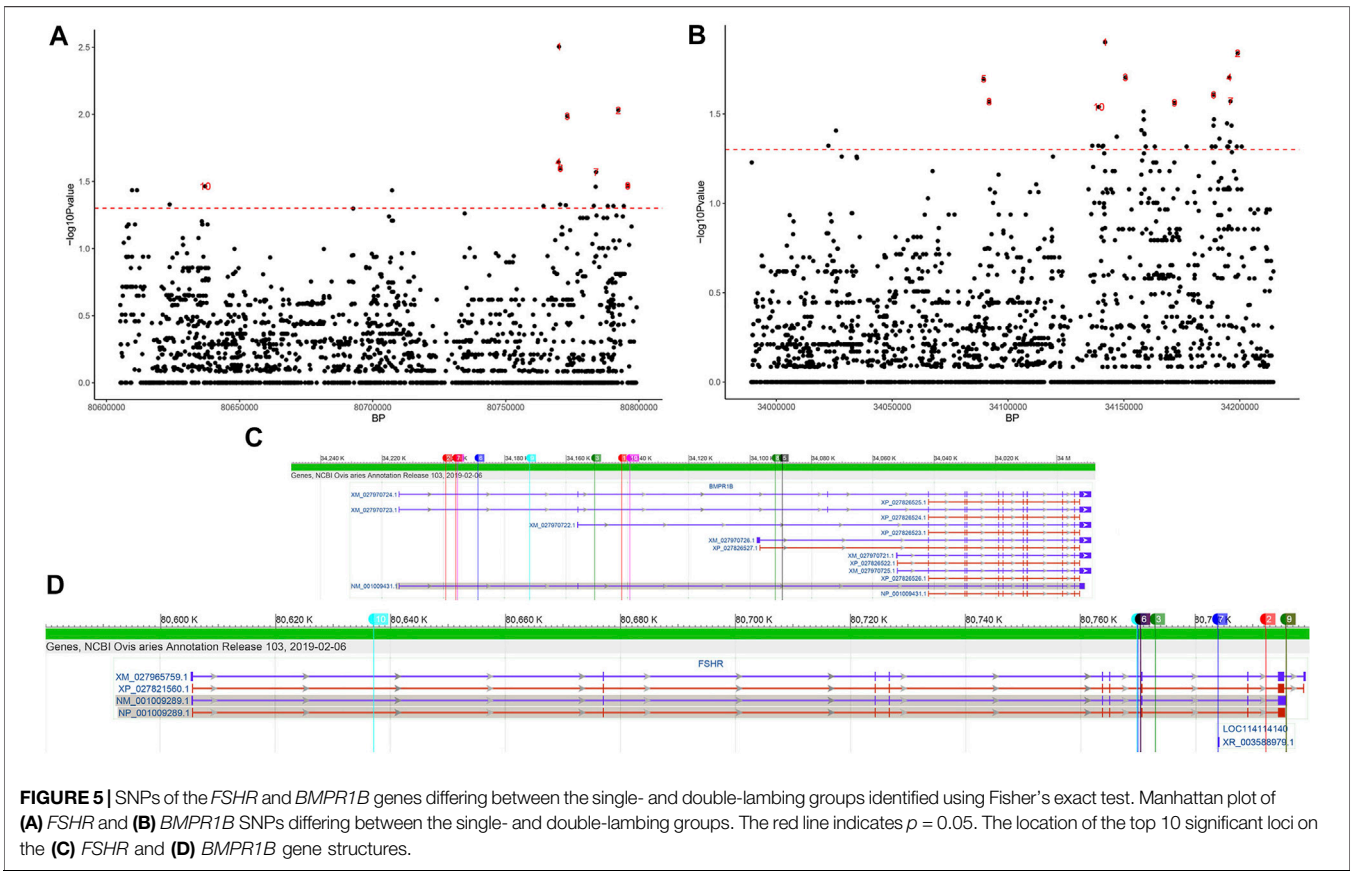

FIGURE 5 | SNPs of the *FSHR* and *BMPR1B* genes differing between the single- and double-lambing groups identified using Fisher's exact test. Manhattan plot of **(A)** *FSHR* and **(B)** *BMPR1B* SNPs differing between the single- and double-lambing groups. The red line indicates $p = 0.05$. The location of the top 10 significant loci on the **(C)** *FSHR* and **(D)** *BMPR1B* gene structures.

type 1B (*BMPR1B*) and oxytocin/neurophysin I prepropeptide (*OXT*), were identified as being under selection in this study. *OXT* was identified by the Tajima's D test, and *BMPR1B* and *FSHR* were identified by the iHS test. Based on the whole-genome sequencing results obtained for Yunnan semi-fine wool sheep, the polymorphic loci in the exons of the *BMPR1B*, *FSHR* and *OXT* genes were analysed by comparison with the reference genome (**Table 2**).

In the *FSHR* gene, a total of 2437 SNPs were identified. Six mutations were detected in its exonic region (**Table 2**), four of which were missense mutations (c.28G < A, c.692G > A, c.783T > G, and c.854C > T, known mutation loci). At the same time, we calculated the distribution of minor alleles in the two groups based on litter size. Twenty-three SNPs with significant differences as detected by Fisher's exact test were finally obtained (**Figure 5A**). Although these SNPs were all located in

intronic regions, they are probably related to reproductive performance in the two groups.

In the *BMPR1B* gene, among the 2645 identified SNPs, 7 SNP mutations were located in the exonic region, only one of which was a missense variant (c.52G > A, known mutation locus, rs605658565), and 44 SNPs showed significant differences between the single- and double-lambing groups (**Figure 5B**). In the *OXT* gene, there were only 3 identified SNPs (Chr13:54216488, 54216953, and 54216953). Although none of the SNPs showed significant differences, we found that in the two-lamb group, the frequency of minor alleles increased *de novo*.

DISCUSSION

Population Situation of Yunnan Semi-Fine Wool Sheep

Yunnan semi-fine wool sheep were the first coarse-grade semi-fine wool sheep breed cultivated in China. In the last 2 decades, the breed has become an indispensable component of sheep germplasm resources in China. This study is the first to explore the population characteristics, genetic structure and selection signatures based on sequencing data of this breed. We used the heterozygosity rate and inbreeding coefficient to evaluate genetic diversity within the population. ROHs are continuous homozygous segments, and they can reflect the inbreeding level of a population at the genome level. Using ROHs to calculate the inbreeding coefficient (F_{ROH}) is more accurate than estimating the inbreeding coefficient from pedigree data (Kim et al., 2015). In this study, the average inbreeding coefficient (0.0099: 0.0004–0.0239) was lower than that of other sheep breeds, e.g., Chaka sheep (F_{ROH} = 0.032) (Cheng et al., 2020), Italian Alpagota sheep (F_{ROH} = 0.053) (Mastrangelo et al., 2018)) and other species, e.g., white leghorn (F_{ROH} = 0.28) (Zhang et al., 2020) and Laiwu pigs (F_{ROH} = 0.133) (Fang et al., 2021), indicating that the population structure of Yunnan semi-fine wool sheep is ideal. Combining the average inbreeding coefficient and heterozygosity rate, we could preliminarily infer that although these two parameters varied greatly within the Yunnan semi-fine wool sheep population, sufficient genetic diversity was maintained overall during cultivation and artificial selection. This conclusion is well supported by the Ne results. In the long-term selection process, the greater the selection intensity is, the smaller the effective population size and the lower the genetic diversity of the population. According to previous studies, when the effective population size is below 50 individuals, the population is threatened (Meuwissen and Woolliams, 1994). According to our results, the Ne of Yunnan semi-fine wool sheep is large enough to indicate a low risk of excessive inbreeding at present.

The breeding of Yunnan semi-fine wool sheep is carried out under a strict system: new genetic materials cannot be introduced into the core group, and sheep in propagation groups must be obtained from foundation seed farms, which maintains the good breeding status of these sheep. However, the herd size in breeding farms is small, e.g., only 1010 and 270 sheep in the Lashishan breeding farm and Xiaohai breeding farm in Qiaojia County, respectively, both of which are representative breeding farms of Yunnan semi-fine wool sheep in Yunnan Province. Such a small population size will lead to inbreeding, resulting in a decrease in the effective population size.

Genomic Signatures of Selection in Yunnan Semi-Fine Wool Sheep

Genetic variation profoundly affects phenotypic variation. When either natural or artificial selection occurs, it will leave traces in the genome. Therefore, we adopted intrapopulation detection methods (Tajima's D and iHS) to discover genome-wide footprints caused by natural and artificial selection in Yunnan semi-fine wool sheep. In this study, overlap of the selected genes simultaneously identified by the two approaches was observed in a few cases (108 genes, **Figure 3E**). The iHS approach shows higher efficiency in detecting partial sweeps; thus, it can detect an event in which a favourable allele increases rapidly from a low frequency but has not yet reached a fixed state (Pritchard et al., 2010). The Tajima's D test is especially powerful for the detection of fixation signatures. According to the characteristics of the above two signature detection methods, the two approaches can be employed together to identify more traces of selection in the genome. Considering the short breeding history of Yunnan semi-fine wool sheep, positive selection may still be affecting some genes related to wool, meat quality and reproductive traits. Furthermore, some alleles for genes associated with domestication and adaptation may be fixed.

Domestication refers to the process by which wild animals come to be maintained under domestic conditions, in which they reproduce over generations and are used by humans. Domestic animals undergo fundamental changes in their phenotypes, morphology, and behaviour relative to those of wild animals (Naval-Sanchez et al., 2018). Some key traits were selected and fixed in most domestic sheep breeds in the early stages of domestication. The changes in these traits mainly manifest as reductions in brain volume and weight, which lead to dulled sensory organ function, more docile personalities and slower activities, such as vision, smell and motor abilities (Kruska, 1996). In this study, we identified some biological pathways associated with nervous system regulation and olfactory receptor activity, consistent with the results of previous studies (Axelsson et al., 2013; Li et al., 2020).

Due to the spread of domestic sheep with human migration activities worldwide, some traits associated with specific human needs have been fixed, leading to greater diversity of some phenotypes. In the Yunnan semi-fine wool sheep population investigated in this study, in addition to the findings related to domestication, we also found many

selected regions and novel functional genes that may be responsible for traits such as reproduction and wool production. Some biological pathways associated with reproductive traits, including vascular smooth muscle contraction and folate biosynthesis, were identified, which may be related to the development of the endometrial vascular system and pregnancy (Cullinan-Bove et al., 1993).

In previous studies on a segregated flock based on QTL analysis and GWAS mapping, some mutations such as *FecB* (Mulsant et al., 2001), *FecX* (Galloway et al., 2000), and *FecG* (Nicol et al., 2009) that may affect ovulation in sheep were identified. Interestingly, we also found that the *BMPR1B* gene was located in a selected region in this study. The *BMPR1B* gene encodes a member of the bone morphogenetic protein (*BMP*) receptor family of transmembrane serine/threonine kinases and regulates animal cell growth and differentiation, embryonic development and reproductive performance (Wilson et al., 2001). Since the mutant *FecB* [nonsynonymous substitution (Q249R)] form of the sheep *BMPR1B* gene was proven to be the main gene regulating high-fertility performance in Booroola sheep, researchers have attempted to identify the *FecB* gene in some indigenous breeds such small-tail Han and Hu sheep (Chu et al., 2011). However, some researchers have reported no significant link between this gene and high fertility traits in breeds such as Tan sheep (Chong et al., 2019). In this study, the *FecB* mutation was not detected in Yunnan semi-fine wool sheep, suggesting that this *BMPR1B* gene mutation may not affect the reproductive performance of the breed. Of course, it is also possible that the examined sample size was too small for detection of this mutation in this study. Nevertheless, we detected 44 SNPs with significantly different frequencies between the single- and double-lambing groups. However, these SNPs were mainly located in intronic regions. Many mutations in intronic regions have previously been found to (**Figure 5D**) lead to functional changes through aberrant splicing (Busslinger et al., 1981; Vaz-Drago et al., 2017), so it is necessary to expand the studied population or conduct experimental verification.

Two other genes that participate in reproduction-related hormone secretion are *FSHR* and *OXT*. The *OXT* gene encodes a precursor protein that is processed to produce oxytocin and neurophysin I. This precursor seems to be activated as it is transported along the axon to the posterior pituitary. This hormone causes the contraction of smooth muscle during parturition and lactation and functions as a neurotransmitter in the central nervous system, playing a role in cognition, tolerance, adaptation, and complex sexual and maternal behaviours (Gimpl and Fahrenholz, 2001). In the present study, there was little evidence that this gene was associated with litter size, but we did find differences in SNPs in this gene region between the two litter size groups, suggesting that *OXT* may be related to litter size. Most relevant studies have shown that the *FSHR* gene is closely related to reproductive traits, such as ovarian function (Laan et al., 2012) and testicular development (Lend et al., 2010), in mammals. SNPs in the 10th exon of *FSHR* are significantly correlated with fertility traits of the Chinese native pig breed Xiaomeishan (Wu and Wang, 2012). Mutations in the 5′ flanking region of the ovine *FSHR* gene may also be associated with litter size (Pan et al., 2014). In this study, we identified specific missense mutations in the Yunnan semi-fine wool sheep population that differed from the reference genome. However, their functions still need to be further explored. On this basis, 23 SNPs with significantly different frequencies were also identified between the single- and double-lambing groups. All of these SNPs were located in intronic regions (**Figure 5C**), but their effect on lambing size needs to be further investigated.

CONCLUSION

In conclusion, using second-generation sequencing technology, the population structure of Yunnan semi-fine wool sheep was detected, and strategies for the conservation of this breed were proposed. In addition, some genes related to adaptation and reproductive traits were shown to have experienced strong selection. The results of this study increase our knowledge of the genetic basis of litter size in Yunnan semi-fine wool sheep, shed light on the changes in heritable phenotypes during the processes of adaptation and artificial breeding, and provide a theoretical basis for the breeding and conservation of Yunnan semi-fine wool 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 below: NCBI, PRJNA783661.

ETHICS STATEMENT

The animal study was reviewed and approved by the ethical committee of the Yunnan Animal Science and Veterinary Institute (201909006).

AUTHOR CONTRIBUTIONS

YG, GQ, and XD conceived the study. JL and CL collected the samples and recorded the phenotypes. YG and YW completed the DNA data analysis. JL, CL, and GW contributed to the visualization of the data. GQ and XD supervised the study and proposed revisions to the manuscript. YG, GQ, and XD wrote and revised the manuscript. All authors read and approved the manuscript.

FUNDING

This investigation was funded by the National Natural Science Foundation of China (Grant No. 31960663), the National Wool Caprine Industrial Technology System (Grant No. CARS-39), Yunnan Applied Basic Research Projects (Grant No. 202001AS070001), the Yunnan Young Academic Leaders Program (Grant No. 202005AC160004), and the Major Science and Technology Project of Yunnan Province (Grant No. 202102AE090039). This work was also supported by grants from the National Key Research and Development Project (2019YFE0106800).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviewers for their constructive comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.812753/full#supplementary-material

REFERENCES

Abied, A., Bagadi, A., Bordbar, F., Pu, Y., Augustino, S. M. A., and Xue, X. (2020). Genomic Diversity, Population Structure, and Signature of Selection in Five Chinese Native Sheep Breeds Adapted to Extreme Environments. *Genes* 11 (5), 494. doi:10.3390/genes11050494

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinformatics* 12, 246. doi:10.1186/1471-2105-12-246

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The Genomic Signature of Dog Domestication Reveals Adaptation to a Starch-Rich Diet. *Nature* 495 (7441), 360–364. doi:10.1038/nature11837

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Busslinger, M., Moschonas, N., and Flavell, R. A. (1981). Beta + Thalassemia: Aberrant Splicing Results from a Single point Mutation in an Intron. *Cell* 27 (2 Pt 1), 289–298. doi:10.1016/0092-8674(81)90412-8

Cheng, J., Zhao, H., Chen, N., Cao, X., Hanif, Q., Pi, L., et al. (2020). Population Structure, Genetic Diversity, and Selective Signature of Chaka Sheep Revealed by Whole Genome Sequencing. *BMC Genomics* 21 (1), 520. doi:10.1186/s12864-020-06925-z

Chong, Y., Liu, G., and Jiang, X. (2019). Effect of BMPRIB Gene on Litter Size of Sheep in China: A Meta-Analysis. *Anim. Reprod. Sci.* 210, 106175. doi:10.1016/j.anireprosci.2019.106175

Chu, M., Jia, L., Zhang, Y., Jin, M., Chen, H., Fang, L., et al. (2011). Polymorphisms of Coding Region of BMPR-IB Gene and Their Relationship with Litter Size in Sheep. *Mol. Biol. Rep.* 38 (6), 4071–4076. doi:10.1007/s11033-010-0526-z

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., and Wang, L. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6 (2), 80–92. doi:10.4161/fly.19695

Cullinan-Bove, K., and Koos, R. D. (1993). Vascular Endothelial Growth Factor/vascular Permeability Factor Expression in the Rat Uterus: Rapid Stimulation by Estrogen Correlates with Estrogen-Induced Increases in Uterine Capillary Permeability and Growth. *Endocrinology* 133 (2), 829–837. doi:10.1210/endo.133.2.8344219

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., and DePristo, M. A. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330

Danecek, P., and McCarthy, S. A. (2017). BCFtools/Csq: Haplotype-Aware Variant Consequences. *Bioinformatics* 33 (13), 2037–2039. doi:10.1093/bioinformatics/btx100

Delaneau, O., Marchini, J., and Marchini, J. (2014). Integrating Sequence and Array Data to Create an Improved 1000 Genomes Project Haplotype Reference Panel. *Nat. Commun.* 5, 3934. doi:10.1038/ncomms4934

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimatorv2: Re-implementation of Software for the Estimation of Contemporary Effective Population Size (Ne) from Genetic Data. *Mol. Ecol. Resour.* 14 (1), 209–214. doi:10.1111/1755-0998.12157

E, G.-X., Duan, X.-H., Zhang, J.-H., Huang, Y.-F., Zhao, Y.-J., Na, R.-S., et al. (2019). Genome-wide Selection Signatures Analysis of Litter Size in Dazu Black Goats Using Single-Nucleotide Polymorphism. *Biotech.* 9 (9), 336. doi:10.1007/s13205-019-1869-3

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

Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P., Juengel, J. L., Jokiranta, T. S., et al. (2000). Mutations in an Oocyte-Derived Growth Factor Gene (BMP15) Cause Increased Ovulation Rate and Infertility in a Dosage-Sensitive Manner. *Nat. Genet.* 25 (3), 279–283. doi:10.1038/77033

Gimpl, G., and Fahrenholz, F. (2001). The Oxytocin Receptor System: Structure, Function, and Regulation. *Physiol. Rev.* 81 (2), 629–683. doi:10.1152/physrev.2001.81.2.629

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Kim, E.-S., Sonstegard, T. S., Van Tassell, C. P., Wiggans, G., and Rothschild, M. F. (2015). The Relationship between Runs of Homozygosity and Inbreeding in Jersey Cattle under Selection. *PLoS One* 10 (7), e0129967. doi:10.1371/journal.pone.0129967

Kruska, D. (1996). The Effect of Domestication on Brain Size and Composition in the Mink (Mustela Vison). *J. Zoolog.* 239 (4), 645–661. doi:10.1111/j.1469-7998.1996.tb05468.x

Laan, M., Grigorova, M., and Huhtaniemi, I. T. (2012). Pharmacogenetics of Follicle-Stimulating Hormone Action. *Curr. Opin. Endocrinol. Diabetes Obes.* 19 (3), 220–227. doi:10.1097/MED.0b013e3283534b11

Lend, A. K., Belousova, A., Haller-Kikkatalo, K., Punab, M., Poolamets, O., Peters, M., et al. (2010). Follicle-Stimulating Hormone Receptor Gene Haplotypes and Male Infertility in Estonian Population and Meta-Analysis. *Syst. Biol. Reprod. Med.* 56 (1), 84–90. doi:10.3109/19396360903456676

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., and Homer, N. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, X., Yang, J., Shen, M., Xie, X. L., Liu, G. J., Xu, Y. X., et al. (2020). Whole-genome Resequencing of Wild and Domestic Sheep Identifies Genes Associated with Morphological and Agronomic Traits. *Nat. Commun.* 11 (1), 2815. doi:10.1038/s41467-020-16485-1

Liu, Z., Sun, C., Qu, L., Wang, K., and Yang, N. (2016). Genome-Wide Detection of Selective Signatures in Chicken through High Density SNPs. *PLoS One* 11 (11), e166146. doi:10.1371/journal.pone.0166146

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., and Portolano, B. (2018). Runs of Homozygosity Reveal Genome-wide Autozygosity in Italian Sheep Breeds. *Anim. Genet.* 49 (1), 71–81. doi:10.1111/age.12634

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

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

Meuwissen, T. H., and Woolliams, J. A. (1994). Effective Sizes of Livestock Populations to Prevent a Decline in Fitness. *Theor. Appl. Genet.* 89 (7-8), 1019–1026. doi:10.1007/BF00224533

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., et al. (2001). Mutation in Bone Morphogenetic Protein Receptor-IB Is Associated with Increased Ovulation Rate in Booroola Merino Ewes. *Proc. Natl. Acad. Sci. U S A.* 98 (9), 5104–5109. doi:10.1073/pnas.091577598

National Commission On Livestock and Poultry Genetic Resources (2011). *Animal Genetic Resources in China: Sheep and Goats*. Beijing: China Agriculture Press.

Naval-Sanchez, M., Nguyen, Q., McWilliam, S., Porto-Neto, L. R., Tellam, R., Vuocolo, T., et al. (2018). Sheep Genome Functional Annotation Reveals Proximal Regulatory Elements Contributed to the Evolution of Modern Breeds. *Nat. Commun.* 9 (1), 859. doi:10.1038/s41467-017-02809-1

Nicol, L., Bishop, S. C., Pong-Wong, R., Bendixen, C., Holm, L. E., Rhind, S. M., et al. (2009). Homozygosity for a Single Base-Pair Mutation in the Oocyte-specific GDF9 Gene Results in Sterility in Thoka Sheep. *Reproduction* 138 (6), 921–933. doi:10.1530/REP-09-0193

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic Scans for Selective Sweeps Using SNP Data. *Genome Res.* 15 (11), 1566–1575. doi:10.1101/gr.4252305

Pan, X., Liu, S., Li, F., Wang, W., Li, C., Ma, Y., et al. (2014). Molecular Characterization, Expression Profiles of the Ovine FSHR Gene and its Association with Litter Size. *Mol. Biol. Rep.* 41 (12), 7749–7754. doi:10.1007/s11033-014-3666-8

Porto-Neto, L. R., Lee, S. H., Sonstegard, T. S., Van Tassell, C. P., Lee, H. K., Gibson, J. P., et al. (2014). Genome-wide Detection of Signatures of Selection in Korean Hanwoo Cattle. *Anim. Genet.* 45 (2), 180–190. doi:10.1111/age.12119

Pritchard, J. K., Pickrell, J. K., and Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr. Biol.* 20 (4), R208–R215. doi:10.1016/j.cub.2009.11.055

Pudovkin, A. I., Zaykin, D. V., and Hedgecock, D. (1996). On the Potential for Estimating the Effective Number of Breeders from Heterozygote-Excess in Progeny. *Genetics* 144 (1), 383–387. doi:10.1093/genetics/144.1.383

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

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., 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:10.1038/nature01140

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31 (10), 2824–2827. doi:10.1093/molbev/msu211

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

Vaz-Drago, R., Custodio, N., and Carmo-Fonseca, M. (2017). Deep Intronic Mutations and Human Disease. *Hum. Genet.* 136 (9), 1093–1111. doi:10.1007/s00439-017-1809-4

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 (3), e72. doi:10.1371/journal.pbio.0040072

Wang, W. (2009). Conservation and Utilization of Yunnan Semi-fine Wool Sheep. *Yunnan J. Anim. Sci. Vet. Med.* (S1), 42–43.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38 (6), 1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x

Wilson, T., Wu, X. Y., Juengel, J. L., Ross, I. K., Lumsden, J. M., Lord, E. A., et al. (2001). Highly Prolific Booroola Sheep Have a Mutation in the Intracellular Kinase Domain of Bone Morphogenetic Protein IB Receptor (ALK-6) that Is Expressed in Both Oocytes and Granulosa Cells. *Biol. Reprod.* 64 (4), 1225–1235. doi:10.1095/biolreprod64.4.1225

Wu, J. S., and Wang, J. Y. (2012). Polymorphism of Exon10 of FSHR Gene and its Relationship with Litter Size in Xiaomeishan Pigs. *Scientia Agricultura Sinica* 45 (13), 2728–2736.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88 (1), 76–82. doi:10.1016/j.ajhg.2010.11.011

Yuan, Y. Y., and Sun, L. M. (2014). *Animal Genetic Resources in Yunnan*. Kunming: Yunnan Scientific Publishing Press.

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, J., Nie, C., Li, X., Ning, Z., Chen, Y., Jia, Y., et al. (2020). Genome-Wide Population Genetic Analysis of Commercial, Indigenous, Game, and Wild Chickens Using 600K SNP Microarray Data. *Front. Genet.* 11, 543294. doi:10.3389/fgene.2020.543294

Zhao, F., Deng, T., Shi, L., Wang, W., Zhang, Q., Du, L., et al. (2020). Genomic Scan for Selection Signature Reveals Fat Deposition in Chinese Indigenous Sheep with Extreme Tail Types. *Animals (Basel)* 10 (5). doi:10.3390/ani10050773

Zhong, X. U., Hao, S., Zhe, Z., Qing-bo, Z., Olasege, B. S., Qiu-meng, L. I., et al. (2020). Genome-wide Detection of Selective Signatures in a Jinhua Pig Population. *J. Integr. Agr.* 19 (5), 1314–1322. doi:10.1016/S2095-3119(19)62833-9

Zhu, C., Li, M., Qin, S., Zhao, F., and Fang, S. (2020). Detection of Copy Number Variation and Selection Signatures on the X Chromosome in Chinese Indigenous Sheep with Different Types of Tail. *Asian-australas J. Anim. Sci.* 33 (9), 1378–1386. doi:10.5713/ajas.18.0661

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fgene.2022.833503

Whole-Genome Analyses Reveal Genomic Characteristics and Selection Signatures of Lincang Humped Cattle at the China–Myanmar Border

Luyang Sun[1,2], Kaixing Qu[3], Xiaohui Ma[2], Quratulain Hanif[4], Jicai Zhang[1], Jianyong Liu[1], Ningbo Chen[2], Quji Suolang[5], Chuzhao Lei[2] and Bizhi Huang[1]**

[1]Yunnan Academy of Grassland and Animal Science, Kunming, China, [2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, [3]Academy of Science and Technology, Chuxiong Normal University, Chuxiong, China, [4]National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, [5]Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Science, Lhasa, China

OPEN ACCESS

Edited by:
El Hamidi Hay, Agricultural Research Service (USDA), United States

Reviewed by:
Ran Li, Northwest A&F University, China
Yue Yaojing, International Livestock Research Institute, Ethiopia
Xia Zhou, Soochow University, China

***Correspondence:**
Chuzhao Lei
leichuzhao1118@126.com
Bizhi Huang
hbz@ynbp.cn

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

Received: *11 December 2021*
Accepted: *01 March 2022*
Published: *22 March 2022*

Citation:
Sun L, Qu K, Ma X, Hanif Q, Zhang J, Liu J, Chen N, Suolang Q, Lei C and Huang B (2022) Whole-Genome Analyses Reveal Genomic Characteristics and Selection Signatures of Lincang Humped Cattle at the China–Myanmar Border. Front. Genet. 13:833503. doi: 10.3389/fgene.2022.833503

The location on the Yunnan border with Myanmar and its unique cultural landscape has shaped Lincang humped cattle over time. In the current study, we investigated the genetic characteristics of 22 Lincang humped cattle using whole-genome resequencing data. We found that Lincang humped cattle derived from both Indian indicine and Chinese indicine cattle depicted higher levels of genomic diversity. Based on genome-wide scans, candidate genomic regions were identified that were potentially involved in local thermal and humid environmental adaptions, including genes associated with the body size (*TCF12*, *SENP2*, *KIF1C*, and *PFN1*), immunity (*LIPH*, *IRAK3*, *GZMM*, and *ELANE*), and heat tolerance (*MED16*, *DNAJC8*, *HSPA4*, *FILIP1L*, *HELB*, *BCL2L1*, and *TPX2*). Missense mutations were detected in candidate genes *IRAK3*, *HSPA4*, and *HELB*. Interestingly, eight missense mutations observed in the *HELB* gene were specific to the indicine cattle pedigree. These mutations may reveal differences between indicine and taurine cattle adapted to variable climatic conditions. Our research provides new insights into the genetic characteristics of Lincang humped cattle representing Lincang and Pu'er areas as an important channel for the migration of Indian indicine from domestication centers toward southwestern China.

Keywords: whole-genome resequencing, Lincang humped cattle, genetic characteristics, selection signatures, *HELB*

INTRODUCTION

Domestic cattle comprise two subspecies, humpless taurine (*Bos taurus*) and humped indicine or zebu (*Bos indicus*), both of which are derived from extinct wild aurochs (*Bos primigenous*) (Decker et al., 2014). Existing research has recognized that worldwide cattle can be divided into five continental groups, European taurine, Eurasian taurine, East Asian taurine, Chinese indicine, and Indian indicine, through whole-genome sequencing analysis (Chen et al., 2018). As for current distribution patterns, modern cattle live in different geographical and climatic zones

worldwide. Taurine cattle mainly inhabit temperate environments. In contrast, indicine cattle adapt to continuous high and variable temperate climates (Barendse, 2017).

Yunnan province in China is traversed by the Tropic of Cancer, which is mainly a tropical and subtropical climate zone. Recent studies have identified the complex genetic diversity and admixture patterns of cattle breeds in Yunnan province (Chen et al., 2018; R.; Li et al., 2019; Liu et al., 2020). Lincang humped cattle is the more primitive regional livestock breed mainly distributed in the southern part of Lincang and Pu'er cities in Yunnan province bordering Myanmar (Gan, 2011). The exact history of the formation of Lincang humped cattle has not been verified. However, the reason for the formation of this breed is the adaptive selection and breeding of local farmers according to their social needs. Moreover, the local ethnicity (Wa ethnic) regards cattle as a totem, and the birth of the Wa ethnicity culture is closely associated with cattle. Lincang humped cattle displays superior characteristics of heat tolerance and resistance to disease and are one kind of Yunnan high-humped cattle (Y. Zhang, 2011). Long-term strong natural selection and human-mediated selection might have potentially affected the structure of the Lincang humped cattle genome by forming detectable selection signals in functional genes (Andersson and Georges, 2004; Hoffmann, 2010).

The unique adaptive characterization of indigenous African and Asian cattle breeds has become a hot topic based on whole-genome sequencing. (Ben-Jemaa, Mastrangelo, Lee, Lee, and Boussaha, 2020; J.; Kim et al., 2017; Xia et al., 2021). In the current study, the whole genome of 22 Lincang humped cattle was resequenced and compared with the sequence data of 61 cattle from five continental groups. Our analysis reports the genome characterization of Lincang humped cattle using whole-genome resequencing, providing many insights into their candidate signatures of positive selection.

METHODS

Samples and Resequencing

A total of 22 domestic Lincang humped cattle (NCBI: PRJNA781760) from Cangyuan Wa Ethnicity Autonomous County, Lincang City, Yunnan Province, China, were sequenced. The purity of that breed was ensured throughout the sampling. However, one cattle might fall under hybrid cattle. Genomic DNA was extracted from the ear tissue samples. Twenty-two paired-end DNA libraries were constructed for the 22 pieces (500 bp insert size) and subjected to Illumina NovaSeq sequencing at the Novogene Bioinformatics Institute, Beijing, China. The genome sequence data of 61 cattle from five continental groups including European cattle breeds [Angus and Simmental (n = 17)], Chinese native breeds (Leiqiong, Guangfeng, Jian, Jingjiang, Wannan, and Wenshan n = 24), Indian cattle breeds (Sahiwal, Hariana, Tharparkar, Nelore, Gir, and Brahman, n = 10), and Korean native breed (Hanwoo, n = 10) were used for the combined analysis (**Supplementary Table S1**).

Reads Mapping and SNP Calling

The clean reads were mapped to the latest *Bos taurus* reference genome (ARS-UCD1.2_Btau5.0.1Y.fa) using BWA-MEM (0.7.13-r1126) (H. Li & Durbin, 2009). The average mapping rate of the reads was 99.72%, and the sequencing coverage was approximately 9.75 × (ranging from 8.94 to 11.78) per individual. Duplicate reads were removed using Picard Tools (http://broadinstitute.github.io/picard). The genome analysis toolkit (GATK, version 3.8) (Using the HaplotypeCaller, GenotypeGVCFs, and Select Variants module) was used to detect SNPs. Previous studies have referred to the SNP calling parameters (Chen et al., 2018). Moreover, ANNOVAR (Using the table_annovar.pl module) (K. Wang, Li, & Hakonarson, 2010) was used to annotate the functions of the SNPs.

Population Genetic Analysis

VCFtools (Danecek et al., 2011) was used to estimate the nucleotide diversity of each breed in window sizes of 50 kb with 20 kb increment. Furthermore, the linkage disequilibrium (LD) decay between pairwise SNPs was calculated by PopLDdecay software (C. Zhang, Dong, Xu, He, & Yang, 2019). PLINK was used to detect the runs of homozygosity (ROH) in each cattle population. The number and length of ROH for each population were estimated and classified into three categories: 0.5–1 Mb, 1–2 Mb, and 2–4 Mb (Sun et al., 2021). PLINK (version 1.9) (Purcell et al., 2007) was again used to remove the linkage sites in genomic data (--indep-pair-wise 50 5 0.2) to perform principal component analysis (PCA) and ADMIXTURE analysis. To accurately identify the components of Lincang humped cattle, ADMIXTURE software (Alexander & Lange, 2011) was used to analyze the population structure with a kinship (K) set from 2 to 5. The aforementioned results were visualized *via* RStudio software (Loraine et al., 2015). A phylogenetic tree was constructed using the neighbor-joining (NJ) method by PLINK with the matrix of pairwise genetic distances and visualized in MEGA7 (Kumar, Stecher, & Tamura, 2016) and FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Selective Sweep Identification

To identify selective sweep regions in Lincang humped cattle, three methods were used (I) The fixation index (F_{ST}) values (Weir & Cockerham, 1984; Porto-Neto et al., 2013), (II) Cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007), and (III) The composite likelihood ratio (CLR) (Nielsen et al., 2005). F_{ST} and XP-EHH were calculated with a 50-kb sliding window and 20-kb steps along the autosomes using VCFtools and in-house scripts between Lincang humped cattle and the reference group. The CLR test was calculated for sites in non-overlapping 50-kb windows using "SweepFinder". Tajima's D statistics and nucleotide diversity were calculated for each candidate gene using VCFtools. Furthermore, functional annotation (GO analysis) and KEGG pathway enrichment were performed by DAVID 6.8 (Huang da, Sherman, & Lempicki, 2009), and FDR <0.05 was used as a threshold to detect significantly enriched genes and pathways. The LD heatmap

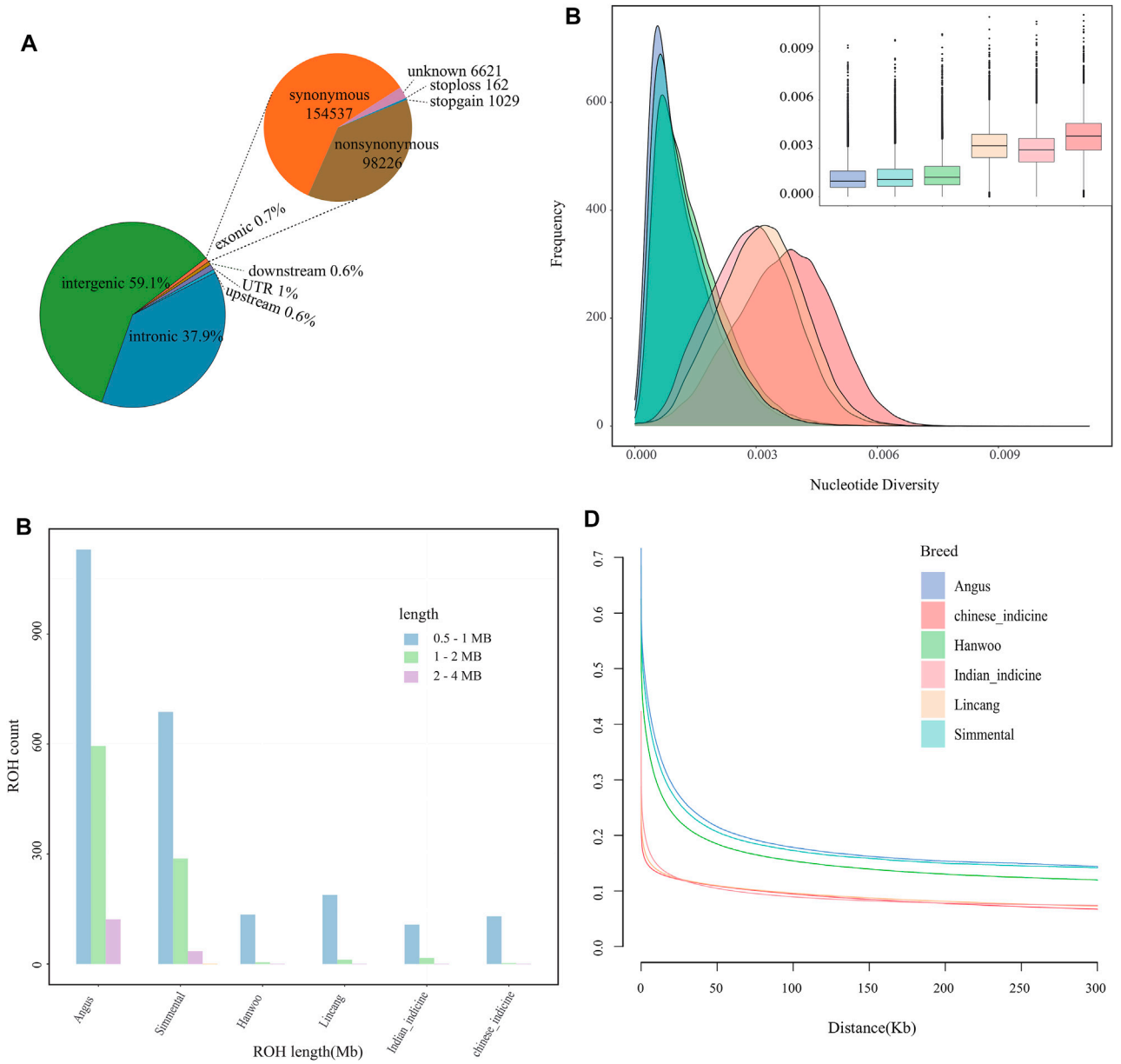

FIGURE 1 | Genomic characteristics of cattle population. **(A)** Functional classification of the detected SNPs in Lincang cattle. **(B)** The estimation of number and length of ROH for each group. **(C)** Estimated nucleotide diversity for each group. **(D)** Linkage disequilibrium (r2) decay on cattle autosomes estimated from each group.

was visualized for candidate genes using VCFtools based on LDBlockShow (Dong et al., 2021).

RESULTS

Resequencing, Identification, and Diversity of Single Nucleotide Polymorphisms

The Lincang humped cattle (n = 22) were selected for genome resequencing (**Supplementary Figure S1A**). The data were combined with the available dataset (n = 61) from 15 breeds, giving a total of 83 individuals (**Supplementary Table S1**). Nelore, Gir, and Brahman were also used to represent Indian indicine cattle in this study. In total, 2.58 billion clean reads were generated and aligned to the reference genome ARS-UCD1.2_Btau5.0.1Y.fa with an average alignment rate of 99.72% and an average depth of 11.87 × in Lincang humped cattle.

Furthermore, 34,636,187 SNPs were detected in mapped reads across 22 Lincang humped cattle. Approximately, 0.7% of SNPs with 98,226 nonsynonymous and 154,537 synonymous SNPs were detected in exonic regions. Then, 37.9% of SNPs were found in intronic regions, 59.1% were found in intergenic regions, 1% of SNPs were observed in untranslated regions

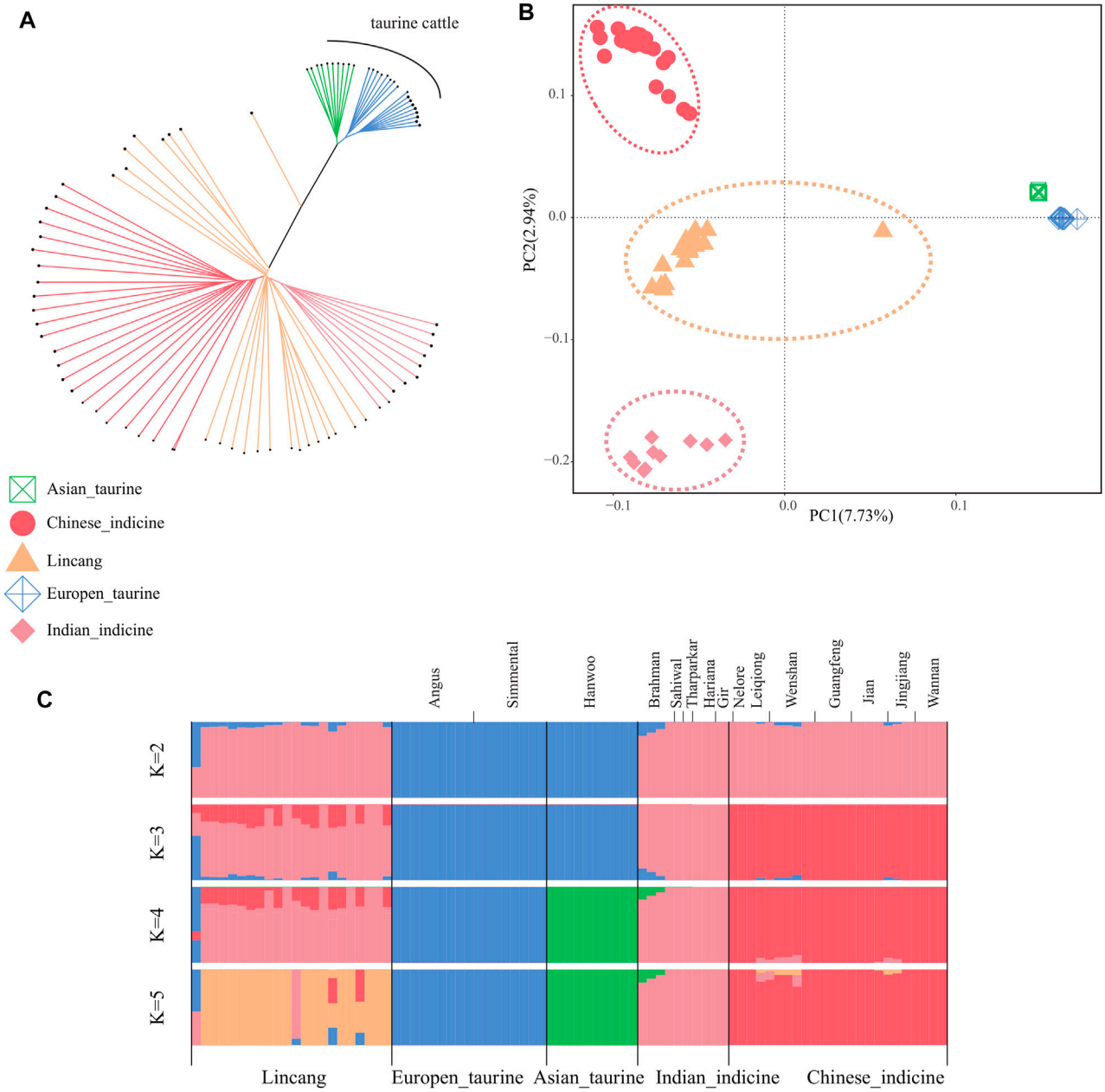

FIGURE 2 | Population structure and relationships of Lincang cattle. **(A)** Neighbor-joining tree of the relationships between the cattle group. **(B)** Principal component analysis of the cattle group. **(C)** Model-based clustering of the cattle group using ADMIXTURE (with K = 2 to K = 5). Breeds are colored by geographic regions and labeled with breed names.

(UTR), and 1.2% of SNPs were found in upstream and downstream of genes (**Figure 1A**). The highest number of SNPs was observed in the Chinese indicine, while the number was lower in taurine cattle than in indicine cattle (**Supplementary Table S2**). Moreover, the highest and second highest numbers of specific SNPs were found in Chinese indicine and Lincang humped cattle, respectively (**Supplementary Figure S1B**). The difference in the number of SNPs and specific SNPs might indicate the differences in the cattle numbers and different populations.

In the absence of pedigree records, ROH may help to infer the level of inbreeding. At the ROH threshold of >2 Mb, Lincang humped cattle showed low levels of genomic inbreeding. Cattle of European origin appeared to be more inbred than other groups (**Figure 1B**). The nucleotide diversity was the highest in Chinese indicine cattle, followed by Lincang and Indian indicine cattle

(**Figure 1C**). The lowest nucleotide diversity was found in taurine cattle. The genome-wide LD was lower for the indicine cattle than for the taurine cattle, which might indicate faster LD decay in indicine cattle than in taurine cattle (**Figure 1D**).

Phylogenetic Relationship, Principal Component Analysis, and Population Structure

The phylogenetic relationship among 83 cattle samples was explored based on the autosomal SNPs. The NJ tree separated taurine and indicine in its clade. The taurine clade clustered Angus, Simmental, and Hanwoo, whereas Chinese indicine, Indian indicine, and Lincang humped cattle were clustered into the indicine clade (**Figure 2A**). An individual of Lincang humped cattle appeared alone between the taurine and the indicine clade, indicating it as a hybrid. Principal component analysis (PCA) demonstrated a clear genetic structure. PC1 explained 7.73% of the total variation and was driven by the difference between taurine and indicine cattle. Within indicine, a separation was found between Chinese indicine and Indian indicine along PC2. The Lincang humped cattle were found at an intermediate position between Chinese indicine and Indian indicine (**Figure 2B**). The admixture estimated from K = 2 to K = 5 showed gradual separation of Lincang humped cattle. When K = 2, the CV error value was the lowest, which means the most reasonable biological explanation was obtained. Lincang humped cattle belonged to *Bos indicus*, composed of crosses with Indian–Chinese indicine genotypes (**Figure 2C**). In particular, hybrid cattle between taurine and indicine cattle appeared in Lincang humped cattle, which might represent the recent introduction of Simmental cattle. To ensure the accuracy of studying Lincang humped cattle, this particular sample was removed in the follow-up analysis.

Candidate Regions and Genes Under Positive Selection

The composite likelihood method (CLR) was applied to detect the selection signals in Lincang humped cattle (**Figure 3A**). The top 1% signal window was selected as candidate regions, while 618 genes were annotated with selection characteristics (**Supplementary Table S3**). KEGG pathway and gene ontology (GO) analyses were used to perform functional enrichment analysis. However, no significant enrichment pathway was found. Surprisingly, the *TCF12* gene was annotated in the top 10 signal windows of CLR. The primary biological process of *TCF12* is to orchestrate the activity of myogenic factors through myogenic differentiation (Parker, Perry, Fauteux, Berkes, & Rudnicki, 2006). Throughout the entire *TCF12* region, Lincang humped cattle showed low nucleotide diversity and constant haplotype diversity patterns (**Figure 3B**).

The fixation index (F_{ST}) test was performed on various groups (I) Lincang humped cattle and Indian indicine; (II) Lincang humped cattle and Chinese indicine; (III) Lincang humped cattle and Hanwoo cattle; and (IV) Lincang humped cattle and European taurine, averaging 0.047, 0.048, 0.32, and 0.34, respectively. Moreover, the F_{ST} and XP-EHH methods were performed on Lincang and Hanwoo cattle to detect the significant positive selection signatures, owing to their vast genetic differences between both breeds pertaining to the ecological adaption (**Figure 3C**). The outlier regions were screened in the top 1% of the empirical distribution of F_{ST} and XP-EHH statistics (**Supplementary Table S4**) and annotated (316 genes). The pathway "negative regulation of protein kinase activity" (FDR = 0.03) was significantly enriched in the KEGG pathway, which might play an important role in the adaptation of Lincang humped cattle to stressful environments.

Moreover, it is noticeable that the regions scanned by F_{ST} on BTA1(81.60–81.65 MB), BTA5 (47.58–47.65 MB), BTA7 (43.16–43.27, 43.40–43.45 MB), BTA16 (50.56–50.61 MB), and BTA19 (26.38–26.45 MB) showed a strong positive selection signal, while using XP-EHH, BTA2 (125.14–125.19 MB), BTA7 (44.54–44.63 MB), BTA11 (55.84–55.89 MB), and BTA12 (24.12–24.19, 76.56–76.65 MB) regions showed strong positive selection signals (**Table 1**). Overall, genes related to the body size (*SENP2*, *KIF1C*, and *PFN1*) (**Figure 3D** and **Supplementary Figure 2A**), immunity (*LIPH*, *IRAK3*, *GZMM*, and *ELANE*) (**Figure 3E** and **Supplementary Figure 2B**), and heat resistance (*MED16*, *DNAJC8*, and *HSPA4*) (**Figure 3F** and **Supplementary Figure 2C,D**) were identified in the 11 candidate genomic regions. Most target genes exhibited lower nucleotide diversity, higher F_{ST}, and differential Tajima's D values than Hanwoo genomic regions, indicating strong selective sweeps. Furthermore, a missense mutation (rs521365524) was found in the *IRAK3* gene, an immune-related gene. This mutation presented a predominant divergence between Lincang humped cattle (allele G frequency = 0.9) and Hanwoo cattle (allele T frequency = 1). Another missense mutation (rs210913195) was detected in the heat-related gene *HSPA4*, which showed a widespread pattern in Lincang humped cattle (frequency 0.98) and the opposite pattern in Hanwoo cattle (frequency 0.2).

It is worth noting that eight overlapped genomic regions and 19 genes were detected among the three mentioned selection methods (**Table 1**), indicating that these were strongly selected in Lincang humped cattle. Among them, *FILIP1L*, *HELB*, *BCL2L1*, and *TPX2* genes were all associated with heat stress. Candidate genes showed a stable haplotype diversity pattern in Lincang humped cattle or discrepant Tajima's D and nucleotide diversity (**Figure 4A** and **Supplementary Figure S2E,F**).

Eight Missense Mutations in *HELB* to Indicine Cattle

The low diversity of the *HELB* gene haplotype in Lincang humped cattle (**Figure 4A**), corresponding to the LD heatmap, showed strong linkage disequilibrium (**Figure 4B**). Surprisingly, we detected eight missense mutations (rs433576296, rs517104855, rs478515513, rs447583631, rs432042680, rs479117197, rs447470311, and rs525001520) that were located within the *HELB* gene, indicating significant genomic differences between Lincang humped cattle and Hanwoo cattle. The allele frequencies

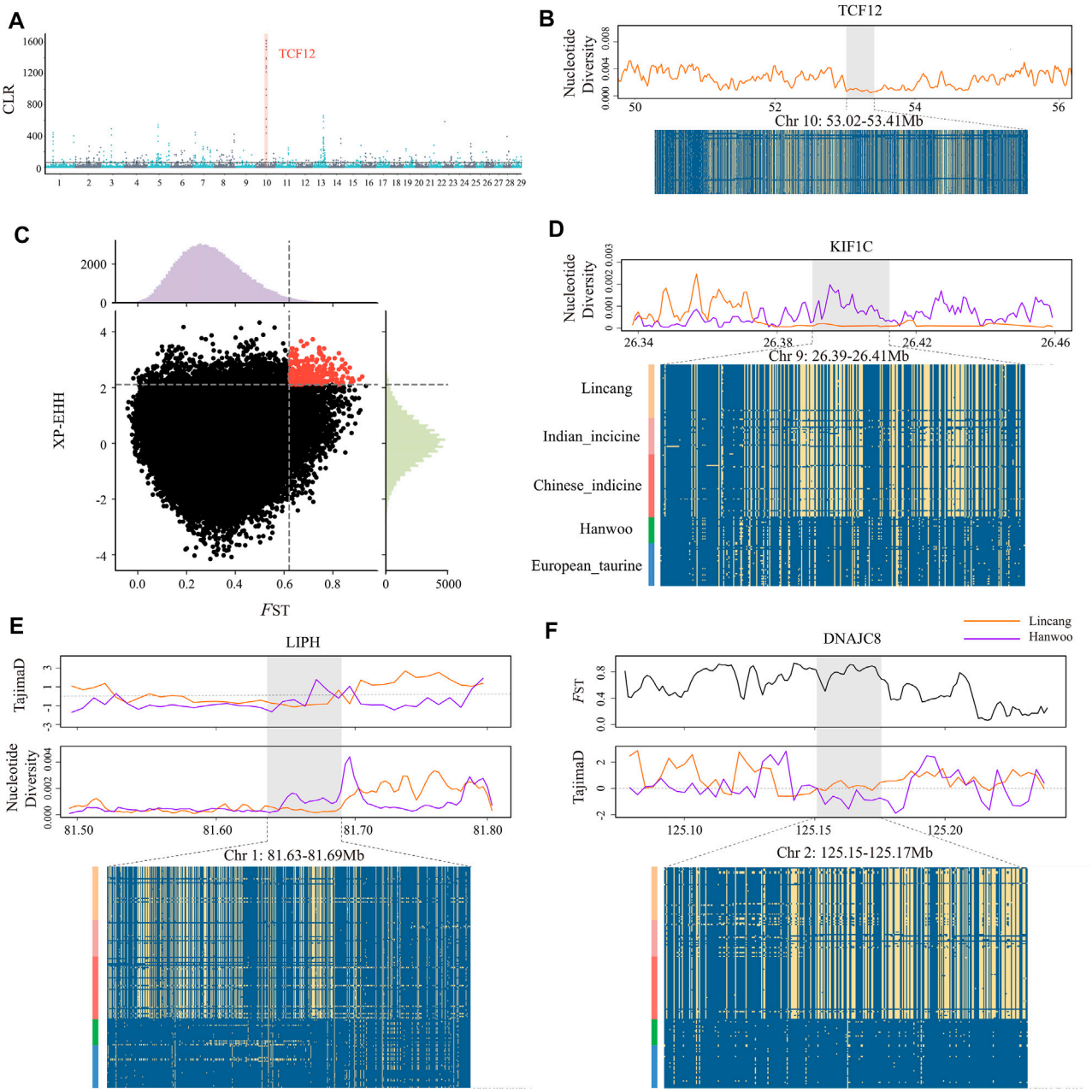

FIGURE 3 | Selective signals in Lincang cattle. **(A)** Manhattan plot of selective sweeps by the CLR method. The horizontal lines indicate the significance threshold (where CLR >61) used for extracting outliers. **(B)** Nucleotide diversity plots and haplotype pattern heatmaps of the *TCF12* gene region. **(C)** Distribution of the F_{ST} (x axis) and XP-EHH (y axis) between Lincang and Hanwoo cattle. The dashed vertical and horizontal lines indicate the significance threshold (where F_{ST} > 0.62 and XP-EHH > 2.1) used for extracting outliers. **(D)** Nucleotide diversity and degree of haplotype sharing across populations at the *KIF1C* gene region. **(E)** TajimaD, nucleotide diversity, and degree of haplotype sharing across populations at the *LIPH* gene region. **(F)** F_{ST}, TajimaD, and degree of haplotype sharing across populations at the *DNAJC8* gene region. The major allele at each SNP position is colored in yellow.

of these missense mutations were estimated in the five major cattle populations. These missense mutations only occur in indicine cattle in our samples (**Figure 4C**), while one of the missense mutations (rs447470311) has been confirmed to be specific to indicine cattle (Naval-Sánchez et al., 2020).

To check whether these mutations are also specific for indicine cattle in a wider cattle breed population, the frequencies of these eight missense mutations were searched among 432 individuals from 54 cattle breeds around the world (**Figure 4D**, **Supplementary Table S5, S3**), adopting BGVD (Bovine Genome Variation Database and Selective Signatures) (Chen et al., 2020) (**Figure 4D**, **Supplementary Table S5, S3**). It was observed that 37 breeds harbored rs479117197 and rs447470311 mutations, where 36 out of 37 were either indicine breeds or

TABLE 1 | Genomic regions and associated genes putatively under selection identified using F_{ST}, XP-EHH, and CLR statistics.

Test	BTA	Start (pb)	End (pb)	Max F_{ST}	Max XP-ehh	Max-CLR	Genes
Top F_{ST}	1	81600001	81650000	0.875749	−−	−−	*SENP2*, *LIPH*
	5	47580001	47650000	0.879839	−−	−−	*TRNAK-CUU*, *IRAK3*, and *TMBIM4*
	7	43160001	43270000	0.920406	−−	−−	*MADCAM1*, *TPGS1*,*CDC34*, *GZMM*, *BSG*, *HCN2*, *TRNAE-UUC*, *FGF22*, *POLRMT*, and *RNF126*
	7	43400001	43450000	0.908626	−−	−−	*PRTN3*, *ELANE*, *R3HDM4*, *CFD*, and *MED16*
	16	50560001	50610000	0.900547	−−	−−	*SKI*, *PRKCZ*, and *FAAP20*
	19	26380001	26450000	0.890845	−−	−−	*INCA1, KIF1C*, *CAMTA2*, *SPAG7*, *PFN1*, and *ENO3*
Top XP-EHH	2	125140001	1.25E+08	−−	3.26	−−	*ATP5IF1* and *DNAJC8*
	7	44540001	44630000	−−	3.74	−−	*ZCCHC10*, *AFF4*, and *HSPA4*
	11	58840001	58890000	−−	3.36	−−	*LRRTM4*
	12	24120001	24190000	−−	3.66	−−	*TRPC4*
	12	76560001	76650000	−−	3.57	−−	*CLYBL*
F_{ST}, XP-EHH, and CLR	1	44080001	44250000	0.796722	2.8	443.703861	*CMSS1* and *FILIP1L*
	2	61500001	61570000	0.710715	2.94	82.469142	*LCT* and *MCM6*
	5	47500001	47570000	0.736184	2.64	92.679749	*HELB* and *IRAK3*
	7	90280001	90430000	0.776616	2.68	220.938087	*ADGRV1* and *MIR2464*
	11	73880001	73930000	0.642026	2.17	73.228089	*DTNB*
	13	61160001	61390000	0.781211	2.54	185.266809	*COX4I2*, *ID1*, *BCL2L1*, *TPX2*, *MYLK2*, and *FOXS1*
	19	39780001	39850000	0.674589	2.53	79.785159	*FBXL20*, *TRNAW-CCA*, and *MED1*
	19	44040001	44090000	0.764124	2.22	71.273024	*SLC4A1*

mixed with indicine. The remaining one was Yanbian cattle from Northeast Asia, which may have individual deviations. At the same time, in rs433576296, rs517104855, rs478515513, rs447583631, rs432042680, and rs525001520, only breeds with indicine ancestry showed the mutations, revealing a higher frequency of these mutations in indicine ancestry alone. It is worth mentioning that the species from Mongolian, Chaidamu, Kazakh from northwest China, and Tibetan depicted low-frequency mutations, probably due to minute indicine introgression (Chen et al., 2018).

DISCUSSION

Lincang humped cattle is one of the groups of Yunnan high-humped cattle in the southern part of Lincang and Pu'er cities bordering Myanmar. Lincang humped cattle has become a valuable genetic resource in the region because of its good environmental adaptability and particularity to the Wa ethnicity (Gan, 2011). As cattle genetic resources are being exhausted (Y. Zhang, 2011), the genetic characterization of Lincang humped cattle is of great significance to this vital genetic resource.

The genetic diversity of genomes may reflect differences in cattle management or breed history. Compared to taurine cattle with long breeding histories, indicine cattle has a lower degree of selection history and higher genetic diversity, consistent with our results and matching those observed in earlier studies (Mei et al., 2018). The Myanmar border is a habitat of Indian indicine, whereas Lincang and Pu'er areas inhabit Lincang humped cattle admixed with Chinese and Indian indicines. Prior studies mark Yunnan as an admixture zone of *Bos taurus* and *Bos indicus* (R. Li et al., 2019), and the entry of Indian indicine in China is also proposed through the Yunnan border (Jia et al., 2007). Our research results indicated that Lincang humped cattle are composed of Indian–Chinese indicine cross genotypes in the genetic structure. This demonstrated that Indian indicine may have entered East Asia through the Lincang and Pu'er areas, which may be an important route for Indian indicine to migrate from the domestication sites (Utsunomiya et al., 2019).

The smaller body size of Lincang humped cattle is associated with its ecological adaptation to hot and humid conditions (Gardner, Peters, Kearney, Joseph, & Heinsohn, 2011). Interestingly, our study detected genes related to skeletal muscle development (*TCF12*, *SENP2*, *KIF1C*, and *PFN1*). Skeletal muscle development is a complex biological process involving multiple key genes. The protein encoded by *TCF12* acts as a complex to positively regulate itself during muscle development (Fu et al., 2020). Related pathways include extracellular signal-regulated kinase signal transduction and CDO in myogenesis (Parker et al., 2006). The *SENP2* gene plays an essential role in the regulation of the muscle growth inhibitor expression and myogenesis, which encodes SUMO-specific protease 2 and is an important regulator of fatty acid metabolism in skeletal muscle (Koo et al., 2015). *KIF1C* plays a role in maintaining membrane circulation during myogenesis and adult muscle (Ginkel & Wordeman, 2000). *PFN1* is a critical factor in skeletal development and regulates sternal bone development and endochondral bone formation (Miyajima et al., 2012). It should be noted that Lincang humped cattle weigh less than 300 kg and have an average height of 1 m (Y. Zhang, 2011). They were small body size cattle. The change in the body size can be explained as an adaptive response to the climate, which means positively selected genes associated with the body size may contribute Lincang humped cattle in humid and hot conditions.

The superior adaptability of Lincang humped cattle is partly attributed to their resistance to disease and parasites (Turner,

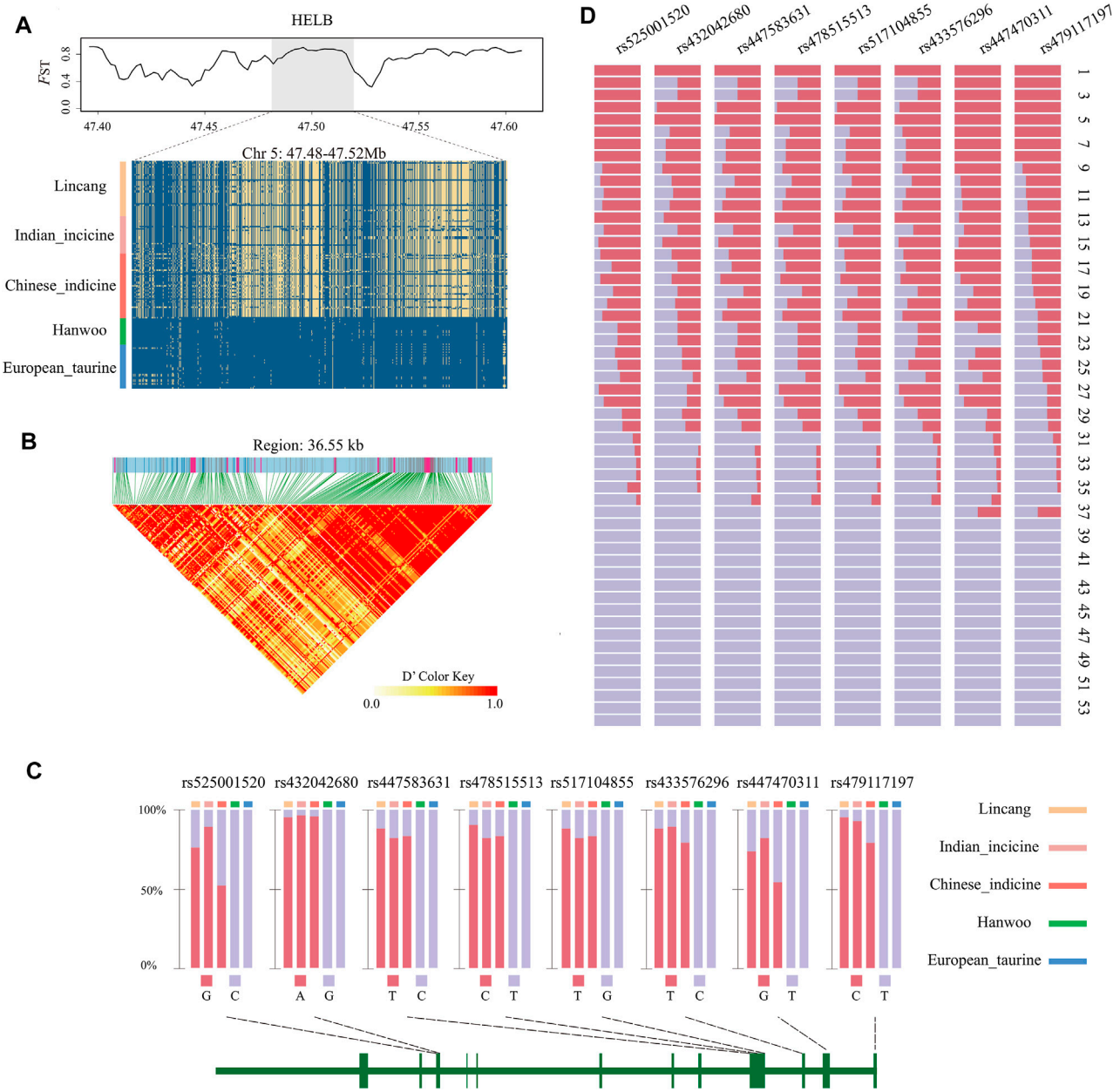

FIGURE 4 | Characteristics of the *HELB* gene. **(A)** F_{ST}, TajimaD, and degree of haplotype sharing across populations. The major allele at each SNP position is colored in yellow. **(B)** Allele frequency of the eight missense SNPs in each group and the schematic structure of genes. The vertical bars in the structure diagram represent the exon region. **(C)** LD plot of SNPs. LD values (D′) between two loci are detailed in boxes (D′ = 0–1). D′ = 1 indicates perfect disequilibrium. **(D)** Allele frequency of the eight missense SNPs across the 54 populations in the BGVD. Population names associated with serial numbers are as follows. 1 Tharparkar; 2 Sahiwal; 3 Hariana; 4 Brahman; 5 Nelore; 6 Gir; 7 Dabieshan; 8 Wandong; 9 Jinjiang; 10 Lingnan; 11 Shorthorn Zebu; 12 Kenana; 13 Srilanka; 14 Bashan; 15 Dianzhong; 16 Wenshan; 17 Guangfeng; 18 Jian; 19 Weining; 20 Ogaden; 21 Leiqiong; 22 Luxi; 23 Nganda; 24 Boran; 25 JiaxianRed; 26 Kazakh; 27 Wannan; 28 Zaobei; 29 Nanyang; 30 Rashoki; 31 Tibetan; 32 Ankole; 33 NDama; 34 Chaidamu; 35 Mongolian; 36 BohaiBlack; 37 Yanbian; 38 Nsongora; 39 Mishima; 40 Kuchinoshima; 41 Hanwoo; 42 Jerse; 43 Simmental; 44 Gelbvieh; 45 Piedmontese; 46 Limousin; 47 Salers; 48 Charolais; 49 MaineAnjou; 50 Devon; 51 Holstein; 52 Hereford; 53 RedAngus; and 54 Angus.

1980). Our study detected genes related to immune response and parasite resistance (*LIPH*, *IRAK3*, *GZMM*, and *ELANE*). A previous study identified the association of *LIPH* with cattle immunity (Zhuang et al., 2021). Similarly, the *LIPH* gene is also found in Dehong cattle (the same Yunnan high-humped cattle as Lincang humped cattle) (R. Li R et al., 2020). *IRAK3* is thought to be a negative regulator of innate immune signaling (Lange, Nelen, Cohen, & Kulathu, 2021). In addition, this gene is also found in selective scans of other indicine cattle (Naval-Sánchez et al., 2020). *GZMM* can affect the killing efficacy against

intracellular pathogens (S. Wang, Xia, Shi, & Fan, 2012). *ELANE* mutations may trigger neutrophil precursors' death and lead to neutropenia (Garg et al., 2020). Furthermore, *ELANE* and *GZMM* have also been demonstrated in African N'Dama cattle, depicting multiple biological functions in parasitic infections (Ben-Jemaa et al., 2020).

For indicine cattle that have lived in tropical and subtropical climatic conditions, several reports have shown that the indicine breed exhibits stronger heat tolerance (Hansen, 2004; J.; Kim et al., 2017). This was also demonstrated by our screening of candidate genes for DNA damage repair and apoptosis associated with heat resistance (*MED16*, *DNAJC8*, *HSPA4*, *FILIP1L, HELB*, *BCL2L1*, and *TPX2*). *MED16* is recruited as the *HSP* gene promoter in response to heat stress (S. Kim & Gross, 2013). Previous studies have established that knockdown of *DNAJC8* decreases antioxidant defenses and increases oxidative damage in honeybees, while *DNAJC8* has been shown to function significantly under heat stress in honeybees (G. Li G et al., 2020). Moreover, *HSPA4* promotes cellular protection against thermal damage and prevents protein denaturation (Niu et al., 2006). Furthermore, several recent genome-wide analyses had detected selective scans for *HSPA4* and highlighted it as a candidate gene for adaptation to hot climates in African indicine cattle (Edea et al., 2018; J.; Kim et al., 2017). *FILIP1L* interacts with *HSF1* to modulate the heat shock response (Hu & Mivechi, 2011). *BCL2L1* acts as an anti-apoptotic gene to control apoptosis inducers (Zinkel, Gross, & Yang, 2006), and it has been suggested that it may be a valuable candidate for heat stress studies in dairy cattle (Khan et al., 2020). *TPX2* functions in the amplification of the DNA damage response (Neumayer, Belzil, Gruss, & Nguyen, 2014). Considering that Lincang humped cattle are well-adapted to hot climates, these genes may play a vital role in the thermal adaptability of Lincang humped cattle.

HELB is involved in DNA damage response as a DNA end-excision inhibitor (Tkáč et al., 2016). Multiple mutations in *HELB* have been identified in mouse cell lines with temperature-sensitive DNA replication (Tada et al., 2001). Furthermore, the mutation rs447470311 in *HELB* revealed in tropical cattle may allow better adaptation to the environment (Naval-Sánchez et al., 2020). An important finding in our study was not only the identification of rs447470311 specific to cattle with indicine cattle pedigree but also the identification of seven missense mutations (rs479117197, rs433576296, rs517104855, rs478515513, rs447583631, rs432042680, and rs525001520) which were only found in cattle with indicine pedigree. Meanwhile, the strong linkage disequilibrium of *HELB* implied that the significant association of a few SNPs in the gene with the trait may be sufficient to indicate association with the majority of SNPs in the gene and implied a substantial enrichment of the biological function (Qanbari, 2019). Therefore, these results may support the hypothesis that missense mutations in *HELB* caused alterations in its DNA damage response function, making indicine cattle more adapted to the hot environment. Additional studies may be required in the future to fully understand the effects of *HELB* on adaptation in indicine cattle.

CONCLUSION

This study explored the genomic variation in the local cattle population at the China–Myanmar border for the first time *via* whole-genome resequencing data. The genomic diversity of Lincang humped cattle was explored and identified as indicine cattle. It is proposed that the Indian indicine might have migrated to southwestern China through the Lincang and Pu'er areas. In addition, we identified candidate genes associated with environmental adaptations such as the body size, immunity, and heat tolerance. Finally, we identified missense mutations in *HELB* that were specific to indicine cattle and were presumed to be associated with adaptation to hot environments. Overall, these results provided a basis for a proper genetic assessment of Lincang humped cattle and further studies on the relationship between *HELB* and heat tolerance in indicine cattle.

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 study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University, following the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China. Specific consent procedures were not required for this study, following the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

AUTHOR CONTRIBUTIONS

NC and CL conceived and designed the experiments. LS and KQ performed the experiments. JZ, JL, and QS contributed analysis tools. LS wrote the manuscript. XM and QH revised the manuscript and provided suggestions. KQ, CL, and BH contributed in the funding for the research. All authors contributed to the manuscript and approved the submitted version.

FUNDING

This work was funded by the China Agriculture Research System of MOF and MARA (CARS-37), the Program of Yunling Scholar, and the Young and Middle-aged Academic Technology Leader Backup Talent Cultivation Program in Yunnan Province, China

(No. 2018HB045), and Yunnan Provincial Major S and T Project (No. 2019ZG007), and the Doctoral Startup Project of Chuxiong Normal University (No. BSQD2101).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.833503/full#supplementary-material

Supplementary Figure S1 | Information for each group. **(A)** Locations for each cattle population. **(B)** The specific and shared SNPs for each group. The numbers in the circle components show specific SNPs for each group or overlapping SNPs among groups.

Supplementary Figure S2 | F_{ST}, TajimaD, nucleotide diversity, or degree of haplotype sharing across populations. The major allele at each SNP position is colored in yellow. **(A)** *MED16* gene region **(B)** *IRAK3* gene region **(C)** *SENP2* gene region **(D)** *FILIP1L* gene region **(E)** *HSPA4* gene region **(F)** *BCL2L1* gene region **(G)** *TPX2* gene region.

Supplementary Figure S3 | Allele frequency of the eight missense SNPs across the 432 individuals from 54 cattle breeds around the world by the BGVD.

REFERENCES

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinformatics* 12, 246. doi:10.1186/1471-2105-12-246

Andersson, L., and Georges, M. (2004). Domestic-animal Genomics: Deciphering the Genetics of Complex Traits. *Nat. Rev. Genet.* 5 (3), 202–212. doi:10.1038/nrg1294

Barendse, W. (2017). Climate Adaptation of Tropical Cattle. *Annu. Rev. Anim. Biosci.* 5, 133–150. doi:10.1146/annurev-animal-022516-022921

Ben-Jemaa, S., Mastrangelo, S., Lee, S.-H., Lee, J. H., and Boussaha, M. (2020). Genome-wide Scan for Selection Signatures Reveals Novel Insights into the Adaptive Capacity in Local North African Cattle. *Sci. Rep.* 10 (1), 19466. doi:10.1038/s41598-020-76576-3

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome Resequencing Reveals World-wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nat. Commun.* 9 (1), 2337. doi:10.1038/s41467-018-04737-0

Chen, N., Fu, W., Zhao, J., Shen, J., Chen, Q., Zheng, Z., et al. (2020). BGVD: An Integrated Database for Bovine Sequencing Variations and Selective Signatures. *Genomics, Proteomics & Bioinformatics* 18 (2), 186–193. doi:10.1016/j.gpb.2019.03.007

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

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *Plos Genet.* 10 (3), e1004254. doi:10.1371/journal.pgen.1004254

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., and Yang, T.-L. (2021). LDBlockShow: a Fast and Convenient Tool for Visualizing Linkage Disequilibrium and Haplotype Blocks Based on Variant Call Format Files. *Brief Bioinform* 22 (4), 227. doi:10.1093/bib/bbaa227

Edea, Z., Dadi, H., Dessie, T., Uzzaman, M. R., Rothschild, M. F., Kim, E.-S., et al. (2018). Genome-wide Scan Reveals Divergent Selection Among Taurine and Zebu Cattle Populations from Different Regions. *Anim. Genet.* 49 (6), 550–563. doi:10.1111/age.12724

Fu, L., Wang, H., Liao, Y., Zhou, P., Xu, Y., Zhao, Y., et al. (2020). miR-208b Modulating Skeletal Muscle Development and Energy Homoeostasis through Targeting Distinct Targets. *RNA Biol.* 17 (5), 743–754. doi:10.1080/15476286.2020.1728102

Gan, S. L. (2011). Exploitation Status and Prospect Analysis of Cangyuan High Hump Yellow Cattle Cured-Beef Industry. *China Cattle Sci.* 37 (2), 66–68.

Gardner, J. L., Peters, A., Kearney, M. R., Joseph, L., and Heinsohn, R. (2011). Declining Body Size: a Third Universal Response to Warming? *Trends Ecol. Evol.* 26 (6), 285–291. doi:10.1016/j.tree.2011.03.005

Garg, B., Mehta, H. M., Wang, B., Kamel, R., Horwitz, M. S., and Corey, S. J. (2020). Inducible Expression of a Disease-Associated ELANE Mutation Impairs Granulocytic Differentiation, without Eliciting an Unfolded Protein Response. *J. Biol. Chem.* 295 (21), 7492–7500. doi:10.1074/jbc.RA120.012366

Ginkel, L. M., and Wordeman, L. (2000). Expression and Partial Characterization of Kinesin-Related Proteins in Differentiating and Adult Skeletal Muscle. *MBoC* 11 (12), 4143–4158. doi:10.1091/mbc.11.12.4143

Hansen, P. J. (2004). Physiological and Cellular Adaptations of Zebu Cattle to thermal Stress. *Anim. Reprod. Sci.* 82-83, 349–360. doi:10.1016/j.anireprosci.2004.04.011

Hoffmann, I. (2010). Climate Change and the Characterization, Breeding and Conservation of Animal Genetic Resources. *Anim. Genet.* 41 (Suppl. 1), 32–46. doi:10.1111/j.1365-2052.2010.02043.x

Hu, Y., and Mivechi, N. F. (2011). Promotion of Heat Shock Factor Hsf1 Degradation via Adaptor Protein Filamin A-Interacting Protein 1-like (FILIP-1L). *J. Biol. Chem.* 286 (36), 31397–31408. doi:10.1074/jbc.M111.255851

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Jia, S., Chen, H., Zhang, G., Wang, Z., Lei, C., Yao, R., et al. (2007). Genetic Variation of Mitochondrial D-Loop Region and Evolution Analysis in Some Chinese Cattle Breeds. *J. Genet. Genomics* 34 (6), 510–518. doi:10.1016/s1673-8527(07)60056-3

Khan, A., Dou, J., Wang, Y., Jiang, X., Khan, M. Z., Luo, H., et al. (2020). Evaluation of Heat Stress Effects on Cellular and Transcriptional Adaptation of Bovine Granulosa Cells. *J. Anim. Sci Biotechnol* 11, 25. doi:10.1186/s40104-019-0408-8

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The Genome Landscape of Indigenous African Cattle. *Genome Biol.* 18 (1), 34. doi:10.1186/s13059-017-1153-y

Kim, S., and Gross, D. S. (2013). Mediator Recruitment to Heat Shock Genes Requires Dual Hsf1 Activation Domains and Mediator Tail Subunits Med15 and Med16. *J. Biol. Chem.* 288 (17), 12197–12213. doi:10.1074/jbc.M112.449553

Koo, Y. D., Choi, J. W., Kim, M., Chae, S., Ahn, B. Y., Kim, M., et al. (2015). SUMO-specific Protease 2 (SENP2) Is an Important Regulator of Fatty Acid Metabolism in Skeletal Muscle. *Diabetes* 64 (7), 2420–2431. doi:10.2337/db15-0115

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Lange, S. M., Nelen, M. I., Cohen, P., and Kulathu, Y. (2021). Dimeric Structure of the Pseudokinase IRAK3 Suggests an Allosteric Mechanism for Negative Regulation. *Structure* 29 (3), 238–251. doi:10.1016/j.str.2020.11.004

Li, G., Zhao, H., Guo, H., Wang, Y., Cui, X., Li, H., et al. (2020). Analyses of the Function of DnaJ Family Proteins Reveal an Underlying Regulatory Mechanism of Heat Tolerance in Honeybee. *Sci. Total Environ.* 716, 137036. doi:10.1016/j.scitotenv.2020.137036

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, R., Li, C., Chen, H., Li, R., Chong, Q., Xiao, H., et al. (2020). Genome-wide Scan of Selection Signatures in Dehong Humped Cattle for Heat Tolerance and Disease Resistance. *Anim. Genet.* 51 (2), 292–299. doi:10.1111/age.12896

Li, R., Li, C., Chen, H., Liu, X., Xiao, H., and Chen, S. (2019). Genomic Diversity and Admixture Patterns Among Six Chinese Indigenous Cattle Breeds in Yunnan. *Asian-australas J. Anim. Sci.* 32 (8), 1069–1076. doi:10.5713/ajas.18.0605

Liu, Y., Xu, L., Yang, L., Zhao, G., Li, J., Liu, D., et al. (2020). Discovery of Genomic Characteristics and Selection Signatures in Southern Chinese Local Cattle. *Front. Genet.* 11, 533052. doi:10.3389/fgene.2020.533052

Loraine, A. E., Blakley, I. C., Jagadeesan, S., Harper, J., Miller, G., and Firon, N. (2015). Analysis and Visualization of RNA-Seq Expression Data Using RStudio, Bioconductor, and Integrated Genome Browser. *Methods Mol. Biol.* 1284, 481–501. doi:10.1007/978-1-4939-2444-8_24

Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., et al. (2018). Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome Resequencing. *Mol. Biol. Evol.* 35 (3), 688–699. doi:10.1093/molbev/msx322

Miyajima, D., Hayata, T., Suzuki, T., Hemmi, H., Nakamoto, T., Notomi, T., et al. (2012). Profilin1 Regulates Sternum Development and Endochondral Bone Formation. *J. Biol. Chem.* 287 (40), 33545–33553. doi:10.1074/jbc.M111.329938

Naval-Sánchez, M., Porto-Neto, L. R., Cardoso, D. F., Hayes, B. J., Daetwyler, H. D., Kijas, J., et al. (2020). Selection Signatures in Tropical Cattle Are Enriched for Promoter and Coding Regions and Reveal Missense Mutations in the Damage Response Gene HELB. *Genet. Sel Evol.* 52 (1), 27. doi:10.1186/s12711-020-00546-6

Neumayer, G., Belzil, C., Gruss, O. J., and Nguyen, M. D. (2014). TPX2: of Spindle Assembly, DNA Damage Response, and Cancer. *Cell. Mol. Life Sci.* 71 (16), 3027–3047. doi:10.1007/s00018-014-1582-7

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic Scans for Selective Sweeps Using SNP Data. *Genome Res.* 15 (11), 1566–1575. doi:10.1101/gr.4252305

Niu, P., Liu, L., Gong, Z., Tan, H., Wang, F., Yuan, J., et al. (2006). Overexpressed Heat Shock Protein 70 Protects Cells against DNA Damage Caused by Ultraviolet C in a Dose-dependent Manner. *Cell Stress Chaper* 11 (2), 162–169. doi:10.1379/csc-175r.1

Parker, M. H., Perry, R. L. S., Fauteux, M. C., Berkes, C. A., and Rudnicki, M. A. (2006). MyoD Synergizes with the E-Protein HEBβ to Induce Myogenic Differentiation. *Mol. Cel Biol* 26 (15), 5771–5783. doi:10.1128/mcb.02404-05

Porto-Neto, L. R., Sonstegard, T. S., Liu, G. E., Bickhart, D. M., Da Silva, M. V., Machado, M. A., et al. (2013). Genomic Divergence of Zebu and Taurine Cattle Identified through High-Density SNP Genotyping. *BMC Genomics* 14 (1), 876. doi:10.1186/1471-2164-14-876

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 (3), 559–575. doi:10.1086/519795

Qanbari, S. (2019). On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. *Front. Genet.* 10, 1304. doi:10.3389/fgene.2019.01304

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 (7164), 913–918. doi:10.1038/nature06250

Sun, L., Qu, K., Liu, Y., Ma, X., Chen, N., Zhang, J., et al. (2021). Assessing Genomic Diversity and Selective Pressures in Bashan Cattle by Whole-Genome Sequencing Data. *Anim. Biotechnol.*, 1–12. doi:10.1080/10495398.2021.1998094

Tada, S., Kobayashi, T., Omori, A., Kusa, Y., Okumura, N., Kodaira, H., et al. (2001). Molecular Cloning of a cDNA Encoding Mouse DNA Helicase B, Which Has Homology to *Escherichia coli* RecD Protein, and Identification of a Mutation in the DNA Helicase B from tsFT848 Temperature-Sensitive DNA Replication Mutant Cells. *Nucleic Acids Res.* 29 (18), 3835–3840. doi:10.1093/nar/29.18.3835

Tkáč, J., Xu, G., Adhikary, H., Young, J. T. F., Gallo, D., Escribano-Díaz, C., et al. (2016). HELB Is a Feedback Inhibitor of DNA End Resection. *Mol. Cel* 61 (3), 405–418. doi:10.1016/j.molcel.2015.12.013

Turner, J. W. (1980). Genetic and Biological Aspects of Zebu Adaptability. *J. Anim. Sci.* 50 (6), 1201–1205. doi:10.2527/jas1980.5061201x

Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., et al. (2019). Genomic Clues of the Evolutionary History of *Bos indicus* Cattle. *Anim. Genet.* 50 (6), 557–568. doi:10.1111/age.12836

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Wang, S., Xia, P., Shi, L., and Fan, Z. (2012). FADD Cleavage by NK Cell Granzyme M Enhances its Self-Association to Facilitate Procaspase-8 Recruitment for Auto-Processing Leading to Caspase cascade. *Cell Death Differ* 19 (4), 605–615. doi:10.1038/cdd.2011.130

Weir, B. S., and Cockerham, C. C. (1984). Estimating F -Statistics for the Analysis of Population Structure. *Evolution* 38 (6), 1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22 (1), 43. doi:10.1186/s12864-020-07340-0

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: a Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, Y. (2011). *Animal Genetic Resources in China Bovines*. Beijing, China: China Agriculture Press.

Zhuang, H., Chen, X., Wang, Y., Huang, S., Chen, B., Zhang, C., et al. (2021). Identification of LIPH as an Unfavorable Biomarkers Correlated with Immune Suppression or Evasion in Pancreatic Cancer Based on RNA-Seq. *Cancer Immunol. Immunother.* 71, 601–612. doi:10.1007/s00262-021-03019-x

Zinkel, S., Gross, A., and Yang, E. (2006). BCL2 Family in DNA Damage and Cell Cycle Control. *Cel Death Differ* 13 (8), 1351–1359. doi:10.1038/sj.cdd.4401987

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.

The reviewer declared an affiliation with several of the authors, LS, XM, NC, and CL to the handling editor at the time of review.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fgene.2022.830393

Population Genomic Sequencing Delineates Global Landscape of Copy Number Variations that Drive Domestication and Breed Formation of in Chicken

Xia Chen[1†], *Xue Bai*[2,3†], *Huagui Liu*[1], *Binbin Zhao*[2,3], *Zhixun Yan*[1], *Yali Hou*[2,3,4*] *and Qin Chu*[1*]

[1]*Institute of Animal Husbandry and Veterinary Medicine, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China,* [2]*CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China,* [3]*China National Center for Bioinformation, Beijing, China,* [4]*College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China*

OPEN ACCESS

Edited by:
El Hamidi Hay, United States Department of Agriculture, United States

Reviewed by:
John William Keele, United States Department of Agriculture (USDA), United States
Yan Li, Yunnan University, China

***Correspondence:**
Yali Hou
houyl@big.ac.cn
Qin Chu
chuqinsd@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: *07 December 2021*
Accepted: *14 February 2022*
Published: *22 March 2022*

Citation:
Chen X, Bai X, Liu H, Zhao B, Yan Z, Hou Y and Chu Q (2022) Population Genomic Sequencing Delineates Global Landscape of Copy Number Variations that Drive Domestication and Breed Formation of in Chicken. Front. Genet. 13:830393. doi: 10.3389/fgene.2022.830393

Copy number variation (CNV) is an important genetic mechanism that drives evolution and generates new phenotypic variations. To explore the impact of CNV on chicken domestication and breed shaping, the whole-genome CNVs were detected via multiple methods. Using the whole-genome sequencing data from 51 individuals, corresponding to six domestic breeds and wild red jungle fowl (RJF), we determined 19,329 duplications and 98,736 deletions, which covered 11,123 copy number variation regions (CNVRs) and 2,636 protein-coding genes. The principal component analysis (PCA) showed that these individuals could be divided into four populations according to their domestication and selection purpose. Seventy-two highly duplicated CNVRs were detected across all individuals, revealing pivotal roles of nervous system (*NRG3*, *NCAM2*), sensory (*OR*), and follicle development (*VTG2*) in chicken genome. When contrasting the CNVs of domestic breeds to those of RJFs, 235 CNVRs harboring 255 protein-coding genes, which were predominantly involved in pathways of nervous, immunity, and reproductive system development, were discovered. In breed-specific CNVRs, some valuable genes were identified, including *HOXB7* for beard trait in Beijing You chicken; *EDN3*, *SLMO2*, *TUBB1*, and *GFPT1* for melanin deposition in Silkie chicken; and *SORCS2* for aggressiveness in Luxi Game fowl. Moreover, *CSMD1* and *NTRK3* with high duplications found exclusively in White Leghorn chicken, and *POLR3H*, *MCM9*, *DOCK3*, and *AKR1B1L* found in Recessive White Rock chicken may contribute to high egg production and fast-growing traits, respectively. The candidate genes of breed characteristics are valuable resources for further studies on phenotypic variation and the artificial breeding of chickens.

Keywords: chicken, copy number variation, evolution, domestication, breed-specific

Abbreviations: CNV, Copy number variation; CNVR, Copy number variation region; DEGs, Differentially expressed genes; GO, Gene ontology; IPA, Ingenuity pathways analysis; KEGG, Kyoto encyclopedia of genes and genomes; LXG, Luxi game; PCA, Principal component analysis; RJF, Red jungle fowl; RW, Recessive white rock; SILK, Silkie; SNP, Single nucleotide polymorphism; SRA, Sequence read archive; WL, White leghorn; XH, Xinghua; YOU, Beijing YOU.

INTRODUCTION

Since the days of Darwin, it has been recognized that a succession of livestock species leads to significant differences in behavior, morphology, and physiology in response to domestication compared with their wild ancestors (Charles Darwin, 1859). Deciphering the genetic basis, molecular mechanisms, and evolutionary driving forces of the complex traits that have been innovated or reshaped by human manipulation during livestock domestication and breed formation has long captured the interest of animal biologists in resource utilization and animal breeding. The development of genomic sequencing technologies provides more powerful tools for a comprehensive understanding of the genetic architectures and evolutionary trajectories of complicated traits in humans and animals (International Chicken Genome Sequencing, 2004; Scally et al., 2012). In addition to the single nucleotide polymorphisms (SNPs), structural variation has been recognized as another crucial factor for driving phenotypic variations, complex diseases, and developmental abnormalities, such as obesity, diabetes, psychiatric diseases, and cancers (Sebat et al., 2004; Feuk et al., 2006; Freeman et al., 2006; Redon et al., 2006). Copy number variation (CNV) is a major type of structural variation in the genome and generally defined by the insertion, duplication, or deletion of a relatively large size of DNA with length >50 bp; this process contributes to much more variability than SNPs (Iafrate et al., 2004; Freeman et al., 2006). Various molecular mechanisms have been proposed for the formation of CNVs, with non-allelic homologous recombination being regarded as a major source of structural variation in regions of extended homology (Hastings et al., 2009; Arlt et al., 2012). CNVs typically affect gene expression and phenotypic specialization directly through dosage compensation (Zhou et al., 2011), or indirectly through altering gene expression by reshaping the three-dimensional genome architecture and local DNA accessibility; thus, they influence the regulatory relationships or intensities between the regulatory elements and targeted genes (Beroukhim et al., 2016; Franke et al., 2016; Rice and Mclysaght, 2017).

CNVs play an important role in the evolutionary adaptation of an organism under both natural and artificial selection, affecting fitness and reproductive ability, which indicates a significant source of adaptive potential. The copy number of *AMY1* is strongly correlated with the evolution of diets; specifically, individuals with high-starch diets have more *AMY1* copies than those with low-starch diets (Perry et al., 2007). Minias et al. (2019) reported that the evolution of the *MHC* copy number in birds was driven by different selective pressures, such as by intra- and extracellular pathogens and parasites. Besides, the CNVs have been implicated in phenotypic variability of traits important for domestication and breed formation in many livestock species. The duplication of *KIT* is significantly associated with white coat color in pigs (Giuffra et al., 2002) and cattle (Durkin et al., 2012). The chicken pea-comb is caused by the duplication of *SOX5* in intron 1 (Wright et al., 2009). The CNV of *ZNF280AY* was negatively correlated with male reproduction trait in Holstein and Simmental bulls (Pei et al., 2019).

Domestic chickens (*Gallus gallus domesticus*), which were initially domesticated from the red jungle fowl (RJF) subspecies *Gallus gallus spadiceus* in East Asia (Peters et al., 2016; Wang M.-S. et al., 2020), have undergone climate and environmental changes, artificial domestication, and commercial breeding for a long period. After thousands of years of domestication and selection, domestic chickens have been separated into several hundreds of distinct breeds spreading globally, and their appearance, behavior, growth, and reproduction traits vary from breed to breed. Hence, chickens are an excellent model to explore the evolution and domestication of animals and to identify the evolutionary spectrum of CNVs controlling domestication, breed formation, and important economic traits. However, most researches mainly focused on yielding CNV maps for chickens (Crooijmans et al., 2013; Tian et al., 2013; Yi et al., 2014; Sohrabi et al., 2018) or identifying CNVs based on some certain traits (Zhang et al., 2014; Xu et al., 2017). The evolutionary spectrum of CNVs (e.g., magnitudes, trajectories, and mechanisms) affecting important economic traits of livestock species during their domestication and breed formation remains largely elusive. Therefore, the main aim of the study is not merely yielding an exhaustive CNV map for chicken but also exploring the role of CNVs in evolution and domestication. What's more, the exclusive CNVs of each breed were systematically analyzed to identify the genes contributing to breed-specific characteristics. This study offers a new perspective on the evolutionary spectrum of CNVs under artificial selection during chicken domestication and breed shaping, and the results may help reveal potential genetic mechanisms for some meaningful traits in chicken, accelerate breeding programs, and improve the quality and efficiency of production.

MATERIALS AND METHODS

Sample Preparation and Sequencing

We collected 47 samples from 6 typical breeds, including 4 Chinese indigenous breeds [eight Xinghua (XH) chickens from Guangdong province, eight Luxi Game (LXG) fowl from Shandong province, eight Beijing You (YOU) chickens from Beijing, nine Silkie (SILK) chickens from Jiangxi province], and 2 introduced commercial breeds [eight Recessive White Rock (RW) chickens, six White Leghorn (WL) chickens]. SILK chicken is known for its fluffy plumage, dark blue bones and skin, and five toes on each foot. Besides, the SILK chicken is thought to have medicinal properties. LXG chicken is a famous gamecock breed. YOU chicken is a dual-purpose breed with a special appearance (crest on the head, beard under the lower jaw, feathers on both shanks, and five or more toes on each foot). XH chicken has slow growth, low production, and favorable meat quality. As for the two commercial breeds, WL chicken is famous for high egg production and RW chicken for high meat yield. Samples were randomly collected to avoid genetic affinity. Blood samples from 47 birds were used to construct libraries and were sequenced using Hiseq2000 platform. About 95–117 million clean paired-end reads with a length of 100 bp were produced for each sample. The raw sequence data reported in this paper

have been deposited in the Genome Sequence Archive in the BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under the accession number PRJCA000093, and they are publicly accessible at http://bigd.big.ac.cn/gsa. Data of four wild chickens, Red Jungle Fowl (RJF), were downloaded from the Sequence Read Archive (SRA) database with accession number PRJNA241474 (Wang et al., 2015). In total, 51 individuals from 7 breeds were included for whole-genome analysis (**Supplementary Table S1**).

Profiling of Whole-Genome CNVs

We analyzed the whole-genome CNVs using four programs, namely mrFAST, CNVnator, BreakDancer, and Pindel, which have disparate algorithms. mrFAST algorithm is mainly used to identify duplicated segments and simultaneously predict absolute CNVs for the duplicated segments (Alkan et al., 2009). CNVnator is based on the read depth signals, representing greater challenges when calling retrotransposons, duplications, and balanced CNVs (Abyzov et al., 2011). CNVnator is able to discover CNVs in a vast range of sizes, from a few hundred bases to mega bases in length, in the whole genome. BreakDancer is a bioinformatics tool that relates paired-end read alignments from a test genome to the reference genome for the purpose of comprehensively and accurately detecting various types of structural variation including deletion, insertion, inversion, intra-chromosomal translocation, and inter-chromosomal translocation. While BreakDancer is not suitable for detecting small variants. Pindel is a split read-based pattern growth approach to detect CNVs. Pindel performs better in detecting small insertions and deletions and can only detect a limited number of large structural variations (> 1 kb) (Abyzov et al., 2011; Liu et al., 2020).

First, we used mrFAST and mrCaNaVaR to exploit the whole-genome CNVs. The reference genome was first masked for repeat elements using RepeatMasker (Smit and Green, 2013) and Tandem Repeats Finder (Benson, 1999). mrFAST was implemented to align paired-end reads to the reference genome in a single mode, accommodating five edit distances, which reported all possible aligned locations for each read. mrCaNaVaR operated the whole-genome shotgun sequence algorithm, which utilized three categories of sliding windows to calculate the normalized read depth, identify large segmental duplications and deletions, and calibrate their absolute copy numbers. The raw read depth in each window was normalized based on its GC content via a LOESS-based smoothing technique. Segmental duplication and deletion regions were declared where at least six out of seven consecutive 5-kb non-masked windows with 1 kb sliding steps presented significantly increased and decreased read depths (mean ± 4*standard deviation); their boundaries were subsequently refined by 1 kb non-masked windows having increased and attenuated read depths (mean ± 2*standard deviation). The absolute copy numbers within 1 kb of non-overlapping non-masked windows were predicted based on the normalized read depths.

For CNVnator, BreakDancer, and Pindel, the clean paired-end reads were aligned to the indexed chicken reference genome (galGal4) by performing the maximal exact match algorithm in BWA (Li and Durbin, 2009), where reads were either uniquely mapped or randomly located at a place if they had multiple alignments. After alignment, we exploited BreakDancerMax to first classify the aligned read pairs into six types: normal, deletion, insertion, inversion, intra-chromosomal translocation, and inter-chromosomal translocation based on the separation distances and alignment orientation between the paired reads. Phred-style quality scores were required greater than 30, with sufficient mapping quality, and a confidence score of 90 was required. Only the deletion results were retained for further analysis. Pindel with split-reads approach was used with default parameters except that the cut off value for the number of supporting reads was set to 5. The deletion with length ≥50 bp was selected.

For CNVnator, the read depth signals were calculated as the count of mapped reads within consecutive non-overlapping bins and further corrected according to the corresponding GC content. A mean-shift technology was applied to partition these read depth signals into segments with presumably different underlying copy numbers. Based on these, the putative CNVs were further predicted by performing statistical significance tests, whose exact copy numbers were estimated as normalized read depths. Since the optimal bin size is crucial for CNV calling, it was once recommended as the one at which the ratio of the average read depth signal to its standard deviation is approximately 4–5 (Abyzov et al., 2011). The bin size in our study was determined as 500 bp. In order to avoid false discovery issues, we subsequently filtered out the CNVs with q0 >50% (zero mapping quality). And only the deletion results were retained for further analysis. We finally prepared the CNV dataset using mrFAST duplication result as duplication regions and integrating the deletion results supported by at least two of four softwares as deletion regions.

Population Genetics Analysis

We integrated the CNV into copy number variation region (CNVR) based on 1 bp overlap. If a CNVR was present in the sample, we labeled 1 for the region, and 0 for absent event. For example: sample 1 has a deletion of chr1:50500-51499, sample 2 has a deletion of chr1:50600- 52199, sample 3 has a deletion of chr1:53000-53999. Then the CNVR is chr1:50500-52199 and chr1:53000-53999. For the first CNVR, sample 1 and 2 present, but sample 3 does not. Then we performed principal component analysis (PCA) using princomp () under R environment (Team, 2021).

To explore the CNVR that shows stratification among populations, we calculated Vst among pairwise populations (Redon et al., 2006). We used a copy number of 1 kb repeat-free non-overlapping windows provided by mrFAST in CNVR. Vst was calculated by the following formula:

$$Vst = \frac{\sigma_T^2 - \frac{n_1 \times \sigma_1^2 + n_2 \times \sigma_2^2}{n_1 + n_2}}{\sigma_T^2}$$

where σ_T^2 is the total variance among all unrelated individuals, σ_1^2 and σ_2^2 are the variance within populations 1 and 2, respectively, and n_1 and n_2 are the population sizes.

The Vst value fluctuated between 0 and 1. The region where Vst = 0 exhibits no differentiation, whereas Vst = 1 means

complete differentiation. When comparing one population with the other, if all the Vst values were above 0.4, we selected the region as a candidate breed-specific region.

Gene Annotation and Enrichment Analysis

We downloaded gene annotations from the Ensembl database with version 78. If there was overlap of at least 1 bp with CNV, the gene was annotated as CNV associated gene. We performed GO and KEGG enrichment analyses using DAVID (https://david.ncifcrf.gov/). The ingenuity pathway analysis (IPA, http://www.ingenuity.com) software was used to analyze the enrichment based on the IPA knowledgebase. Categories of disease and disorder, molecular and cellular functions, and physiological system development and functions are developed in IPA. The *p*-value was calculated by right-tailed Fisher's exact tests.

Different Expression Analysis

Nine RNA sequencing data of cerebrum samples from LXG, SILK, RW, and WL male chickens (two for LXG, RW, and WL respectively and three for SILK) (Hou et al., 2020), as well as two transcriptomes of cerebrum samples from RJFs (one male and one female) with accession numbers SRR306710-306711 (Brawand et al., 2011) were downloaded and analyzed together. The clean sequencing data were mapped to the reference genome (galGal4) by TopHat v2.0.13 (Trapnell et al., 2009). The different expression level for the comparison of RJF versus all domesticated chickens was detected by HTSeq and DESeq with filters of Padj <0.01 and fold change (FC) >1.5 (Anders and Huber, 2010). We obtained the average expression level of differentially expressed genes (DEGs) for each breed based on the normalized count value provided by DESeq, and then we calculated the Pearson correlation coefficient with the average copy number of each breed.

RESULTS

Identification of Whole-Genome Copy Number Variation

In this study, 47 birds from four Chinese indigenous breeds and two introduced commercial breeds were randomly selected and whole-genome sequenced (**Supplementary Table S1**). SILK, YOU, LXG, and XH chickens were sampled from Jiangxi, Beijing, Shandong, and Guangdong provinces, respectively. The two commercial breeds were RW and WL, representative of broiler and layer, respectively. Additionally, the genome sequence of four RJFs, which represented the wild chicken, was downloaded from the SRA database (Wang et al., 2015). Herein, whole-genome sequencing data from 51 individuals for seven breeds were included in our analysis. The sequencing data were endowed with an approximate coverage depth of 10× paired-end reads per individual and only autosome data were used (**Supplementary Table S2**).

To complementarily capture disparate aspects of CNVs, we applied multiple approaches that simultaneously characterize distinct properties of CNVs based on different principles. For duplication, the mrFAST method was performed, and an average of 379 duplications for each individual was detected. For deletion, we integrated four methods of mrFAST, CNVnator, BreakDancer, and Pindel, and short variations (< 50 bp) were removed. The average deletion numbers for each bird detected by four different methods were 54, 794, 2,282, and 2,278, respectively (**Supplementary Table S2**). Approximately 35.61% of these deletion events were discovered *via* at least two methods, and only 0.26% were shared by all four methods. For next analysis, only deletions supported by at least two methods were retained, with an average of 1,936 deletions per individual.

Then, we merged the above results of 19,329 duplications and 98,736 deletions to obtain CNVRs. A total of 11,123 CNVRs (length range: 68–2,802,702 bp) were elaborately identified, which represented 7% of the autosome genome, with the mean and median length of 6,748 bp and 986 bp, respectively. Most CNVRs (8,834) were denominated as "deletion," and 1,911 were "duplication," while the other 378 were "complex," which included both "deletion" and "duplication" events (**Supplementary Table S3**). From the viewpoint of frequency spectrum that shared among CNVRs of the 51 individuals, we found 4,515 singletons, 5,121 lowly shared CNVRs with frequencies less than 20% excluding singletons, 1,200 medially shared with a frequency of 20–80%, 135 highly shared with a frequency of 80–99%, and 152 common CNVRs, resulting in a typical skewed distribution (**Supplementary Figure S1**). Meanwhile, the CNVR lengths presented a typical skewed distribution identically with a higher proportion of smaller CNVRs (**Supplementary Figure S1**), generally indicating a selective genetic load on CNVs.

Furthermore, we compared the CNVRs with the Ensembl BioMart v78 database (galGal4) to identify genes contained in those CNVRs and filtered the singleton CNV in the population to ensure a high quality of functional CNVs. A total of 2,636 protein-coding genes overlapped with duplications and deletions in CNVRs. Then, Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway and Gene Ontology (GO) term analysis were performed using the DAVID database. Genes with copy number variations were mainly associated with axon guidance, ATP binding, GTPase activator activity, and phosphatidylinositol binding based on the GO term, and they were involved in vascular smooth muscle contraction, oocyte meiosis, and gap junction according to KEGG pathway analysis (**Supplementary Table S4**). This suggests that the functional shape of CNVs mainly modifies behavior, energy, and reproduction, which are typical target traits in chicken domestication.

Population Characteristics of Whole-Genome Copy Number Variation

PCA was performed on all CNVRs detected in the seven populations to explore the population stratification based on the variations (**Figure 1**). By integrating PC1 and PC2, which interpreted 16% and 10% of the phenotypic variance respectively, the samples were primarily divided into four clusters. The first and second clusters represented the introduced meat dynamo RW and the introduced layer dynamo WL, respectively. The third

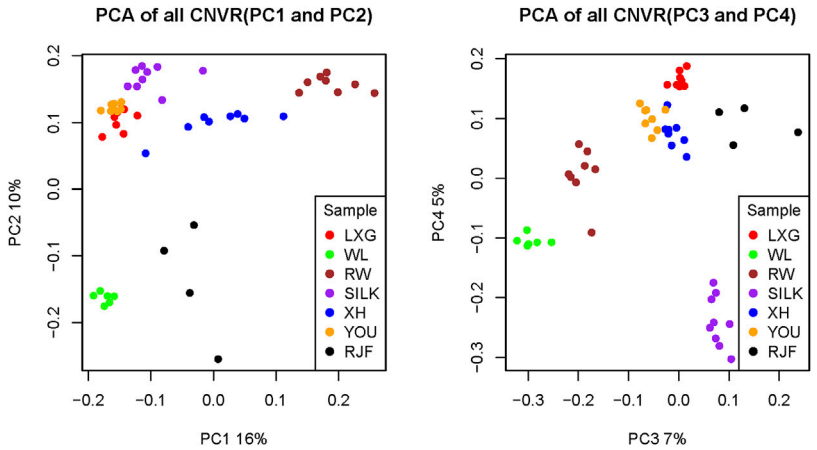

FIGURE 1 | The principal component analysis (PCA) plot of all CNVRs. **(A)** Samples were divided into four groups: commercial layer, commercial broiler, RJFs, and Chinese native breeds. **(B)** Samples were divided more detailly that SILK was further separated from the other Chinese native breeds. XH: Xinghua; LXG: Luxi Game fowl; YOU: Beijing You; SILK: Silkie; RW: Recessive White Rock; WL: White Leghorn; RJF: Red jungle fowl.

cluster was the wild RJF from Yunnan and Hainan provinces of China. The fourth was a cluster of Chinese native breeds, which could also be separated from each other, except for YOU and LXG breeds. Based on PC3 and PC4, which described 7% and 5% of the total variance, respectively, these populations could also be distinguished apparently, and SILK chicken was further separated from other Chinese native breeds. These findings suggested that the CNV spectrum can substantially reflect breed differentiation in light of genetic divergence and artificial selection.

Ubiquitous Copy Number Variations in Chickens

The common copy number variations of chickens included in all the wild and domesticated individuals indicate that the genomic characteristics might be relevant to speciation of chicken or its ancestors. As mentioned, 152 CNVRs were commonly detected among all 51 individuals. Among these regions, 72 were highly represented duplication CNVRs and 172 corresponding genes were involved, with minimum and maximum copy numbers of 3 and 373, respectively. They were predominantly enrichment in feather keratin, sensory perception, metabolic, reproductive processes, muscle contraction, and immunity (**Supplementary Table S5**).

As expected, highly duplicated segments of keratin genes in terms of feather keratin and feather-keratin-like protein were observed in chromosomes 25 and 27, with copy numbers ranging from 3 to 90. *MUC5B*, a member of the mucin family, was discovered to be intensively duplicated with varied copy numbers ranging from 17 to 42. Additionally, four genes, *MRPS35*, *SMARCD3*, *KHDRBS2*, and *ANO3* exhibited a minimum copy number of more than 30 and a maximum of over 60. *NRG3* and *NCAM2*, which are associated with development and differentiation of nervous system, were also included, and their copy numbers ranged from 3 to 42 and 15 to 40, respectively. Besides, 3–5 copy numbers of *VTG2* and 3–6 of *GBE1* were identified.

Copy Number Variations Correlated With Chicken Domestication

The loci with high copy number differentiation between wild and domestic populations are considered as candidates driving adaptive selection. Therefore, the copy number variations between RJF and all the six domestic breeds were compared to identify the potential genes associated with chicken domestication. The method of Vst was used to estimate the population differentiation of copy numbers (Redon et al., 2006). A total of 235 CNVRs containing 255 protein-coding genes were discovered in the windows with 1% toppest Vst (> 0.2042) (**Supplementary Table S6**). The gene enrichment analysis by IPA showed that these copy number differentiated genes are predominantly involved in nervous system development, reproductive system development and function, cellular growth and proliferation, and immunity ($p < 0.01$) (**Supplementary Table S7**).

First, quite a few of genes which were involved in the development and function of nervous system were found to be highly differentiated in gene copy, including *NRG3*, *BICD1*, *ANKRD11*, *SEMA3F*, *UNK*, *EPS15L1*, *KIF3C*, *FBXO41*, *MYT1*, *NR2E1*, *PSMC5*, *SEMA3F*, *SOX2*, *PTPRM*, and *RPS6KB1*. A CNVR in exon 1 of *NRG3* was observed to have an average copy number higher than 30 in RJFs, but the median copy number in domestic individuals was 16 (**Supplementary Figure S2**). Second, some genes associated with immunity or stress response, such as *CBFA2T3*, *IL22RA1*, *HERPUD1*, *ARIH2*, *GATA3*, *TNFAIP8L1*, *CD86*, *CD8A*, *SMARCD2*, and *TNPO3*, were also detected possessing differential copy numbers in genome. Besides, some other valuable genes correlated with energy metabolism, cell growth and proliferation or reproduction functions were also found, for instance, *GBE1*,

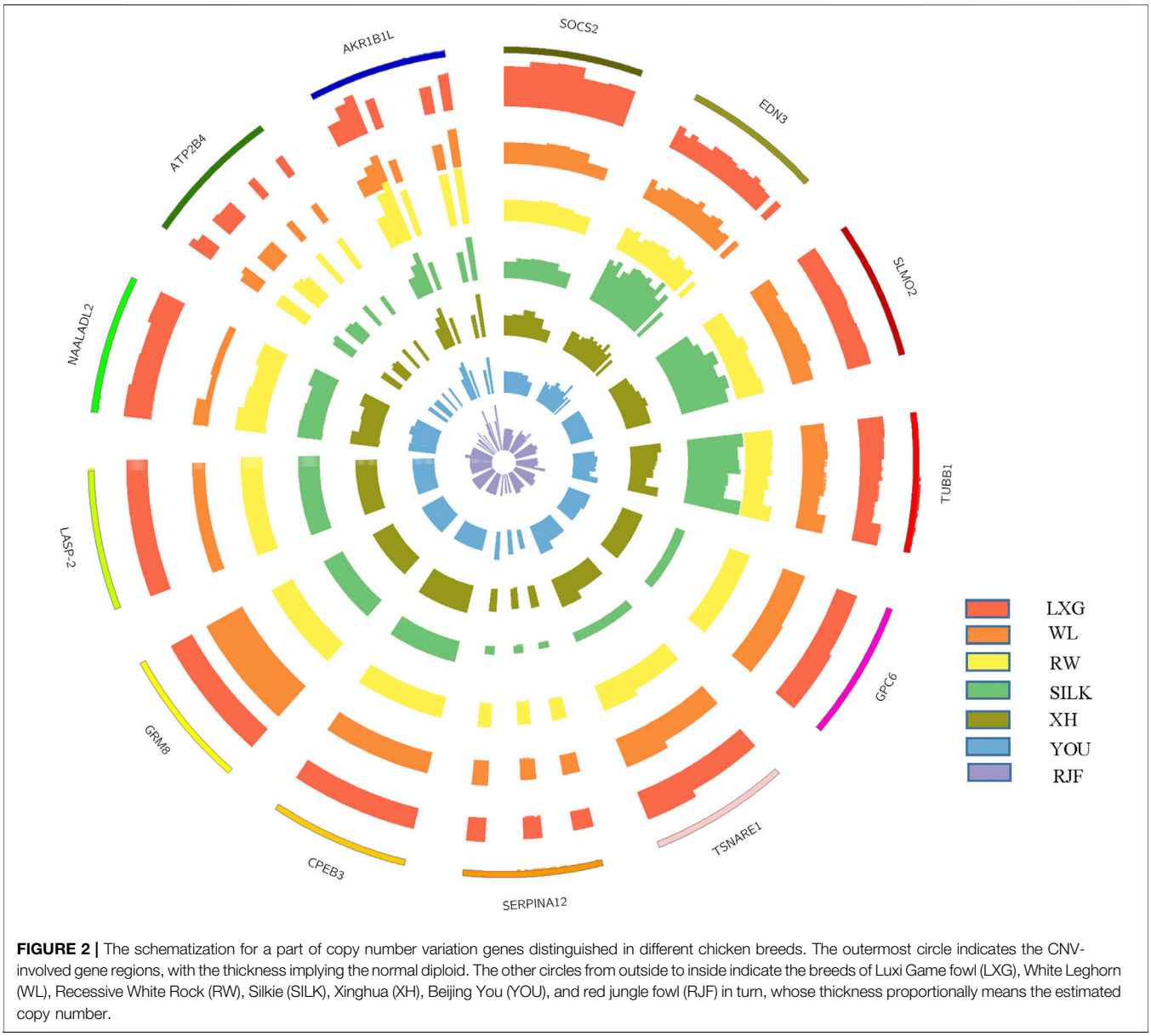

FIGURE 2 | The schematization for a part of copy number variation genes distinguished in different chicken breeds. The outermost circle indicates the CNV-involved gene regions, with the thickness implying the normal diploid. The other circles from outside to inside indicate the breeds of Luxi Game fowl (LXG), White Leghorn (WL), Recessive White Rock (RW), Silkie (SILK), Xinghua (XH), Beijing You (YOU), and red jungle fowl (RJF) in turn, whose thickness proportionally means the estimated copy number.

ACAT, *ACO2*, *HEP21*, *INHA*, *GPC6*, *MEI1*, *PTPRM*, *RPS6KB1*, etc. *GBE1* presented one copy number more in intron 7 in RJF than in all the other six domesticated chicken breeds (**Supplementary Figure S2**).

Breed-Specific Copy Number Variations

Considering that CNV may be correlated with the origin or breeding of different chicken breeds, we further compared each domestic breed with each of the other six breeds to determine the extremely specific regions. If the six Vst values of a region were over 0.4, it was considered as a breed-specific region. In total, 600 CNVRs harboring 90 annotated genes were found to be breed-specific (**Supplementary Table S8**).

RW chickens exhibited 64 breed-specific CNVRs and 39 protein-coding genes, showing the highest number of specific genes across all the six domestic breeds. Genes specific to RW chickens may be candidate genes for growth and muscle traits. For example, *POLR3H*, *MCM9*, *ASF1A*, *MYH1B*, and *DOCK3* may contribute to rapid growth and plump muscle traits. Furthermore, WL chickens possess 63 regions and 19 genes. Many genes, i.e., *GSK3B*, *CSMD1*, *NTRK3*, and *STRBP*, may be associated with reproduction traits.

For the four Chinese native chicken breeds, LXG, SILK, YOU, and XH, we detected 32, 23, 17, and 2 regions, corresponding to 8, 15, 7, and 1 protein-coding gene, respectively. From the schematization of copy numbers for a part of genes and the CNV-involved gene segments that distinguished chicken breeds (**Figure 2**), we could notice the difference clearly. In LXG chickens, *SOCS2* presented a copy number of about 4, which was significantly higher than that of the other breeds. *EDN3* had

around 4 copies exclusively in SILK chickens, and its neighboring genes *SLMO2* and *TUBB1* also had higher copy numbers than other breeds, which have been frequently targeted as tetra- or penta-ploid (Han et al., 2014; Yi et al., 2014). Moreover, SILK chickens have specially experienced one-copy deletion on *GPC6*, *TSNARE1*, and *SERPINA12*. For XH chickens, *CPEB3* was the only detected breed-specific gene, showing one-copy higher particularly. It was reported that *CPEB3* was related to pathway of oocyte meiosis (Wang Y. et al., 2020).

Correlation Between Copy Number Variation and Gene Expression in Cerebrum

Considering that variation in gene copy number is expected to change the RNA expression level, the DEGs between wild RJF and each of the other four domesticated breeds (LXG, SILK, RW, and WL) in cerebrum published by our team (Hou et al., 2020) were further explored. A total of 28 genes were shared among the domestication-related copy number differentiated genes. Additionally, the correlations between the gene expression and copy numbers were intersected (**Supplementary Table S9**). More than half (16/28) of the 28 overlaps presented a significant correlation between copy number and gene expression, with absolute Pearson's correlation coefficients larger than 0.88 ($p < 0.05$). Among the significantly correlated genes, most (9/16) showed negative correlations.

As mentioned, 15 genes participating in the development and function of nervous system showed copy number variation specifically in RJF, among which seven (*BICD1*, *ANKRD11*, *SEMA3F*, *UNK*, *EPS15L1*, *KIF3C*, and *FBXO41*) were significantly differentially expressed in the cerebrum, with one (*ANKRD11*) upregulated and the other six downregulated in domesticated breeds compared with RJF. Meanwhile, except for *BICD1*, six showed negative correlation with copy number.

DISCUSSION

In this study, whole-genome CNVs were explored in RJF and six domestic chicken breeds. The particularity for each breed predominantly contributed to their inner divergences, for example, high egg performance for WL, high meat yield for RW, fluffy plumage, black skin, and medical purpose for SILK, game fighting for LXG, low production and favorable meat quality for XH, and special appearances for YOU chickens. By combining multiple methods, four for deletions and one for duplications, we systematically and exquisitely analyzed the copy number information in 51 individuals and obtained a comprehensive map of genomic variations for chickens. A total of 19,329 duplications and 98,736 deletions were detected, covering 11,123 CNVRs and 7% of the autosome genome.

PCA results with both duplication and deletion CNVRs revealed that the seven chicken populations were well divided into four parts according to the degree of artificial selection pressure and the purpose of domestication: commercial broiler, commercial layer, Chinese native breeds, and their ancestor RJF. These results were consistent with the population stratification pattern based on SNPs (Hou et al., 2020), revealing that CNV is a powerful and effective tool for distinguishing samples and that CNVs may also participate in the driving of chicken domestication.

Genes Relating to Chicken Evolution

As an important genetic source, copy number variation can provide a good perspective for exploring the track of animal evolution. Therefore, 72 high duplication CNVRs (corresponding to 172 genes) detected in all individuals were analyzed to explore chicken evolution, which might be characterized as the genomic features specialized in chicken species.

As expected, feather keratin (*FK* or *F-KER*), olfactory receptor (*OR*), and neuregulin 3 (*NRG3*) were detected CNVs as previous studies reported. Keratin is a main protein that makes up avian feathers (Ramakrishnan et al., 2018). *FK* or *F-KER* was observed with intensive copy number duplication in current study, which is highly consistent with the results of Hillier et al. (2004). *OR* genes, which plays a pivotal role in the sense of smell among vertebrates, exhibited a copy number around 20 in this study. Zhou et al. (2019) reported more than 90 *OR* genes in chicken, which was still much fewer than in mammals like humans (396), dogs (857), cats (865), cattle (881), and pigs (1113) (Bu et al., 2019). The contraction of *OR* genes in chicken genome may be responsible for the relatively poor sense of smell for bird species. *NRG3* was found to be associated with the development and differentiation of nervous system. Abe et al. (2017) also detected copy number variations and frameshift deletion of *NRG3* in chicken.

Besides, some genes, such as *MUC5B*, *NCAM2*, and *VTG2*, were discovered with copy number differences in all individuals. *MUC5B* is a major contributor to the lubricating and viscoelastic properties of whole saliva, normal lung mucus, and cervical mucus (Roy et al., 2014; Käsdorf et al., 2017). The gene *NCAM2* has been proposed to contribute to neurodevelopmental disorders in humans (Winther et al., 2012). Furthermore, we identified that *VTG2*, which is the yolk precursor protein expressed in females of nearly all oviparous organisms, possessed 3–5 copies across all individuals. The presence of multiple copies of *VTG2* in the genome was due to the gene family expansion from common ancestor (Biscotti et al., 2018) and occurred prior to the most recent common ancestor of birds (Brawand et al., 2008; Finn et al., 2009).

Genes Involved in Chicken Domestication

About 150 years ago, Darwin first introduced the concept of domestication syndrome, that domestic animals tended to share some common characteristics, such as more docility and frequent estrus cycles (Darwin, 2010; Wilkins et al., 2014). The domestication genes that cause genetic differentiation between wild and domestic populations appear in the early stages of domestication (Larson et al., 2014). We compared the diversity of CNVs between RJF and six domestic chicken populations to identify genes that may further illuminate the genetic basis of

chicken domestication. A total of 255 protein-coding genes, predominantly involved in the development of nervous, tissue and reproductive system, were detected.

Previous reports demonstrated that neuronal development and behavior modification are under strong selection during long-term domestication progresses (Albert et al., 2008; Sanchez-Villagra et al., 2016; Zhang et al., 2018). More than 15 genes associated with the development and function of the neural system exhibited copy number variations between domestic breeds and RJF. Seven out of these genes were differentially expressed in cerebrum between domestic breeds and RJF. Except for *ANKRD11*, *NRG3*, and *MYT1*, the other genes were detected copy number variation for the first time. *NRG3* showed high copy number deletion in domestic chickens, which is consistent with previous study (Abe et al., 2017). As mentioned, *NRG3* has been implicated in severe neurological disorders with developmental origins (Meier et al., 2013; Li et al., 2020). *ANKRD11* showed copy number deletion in people with autism spectrum disorder compared to healthy persons (Marshall et al., 2008). The knockdown of *ANKRD11* exhibited markedly reduced neurons dendrite outgrowth (Ka and Kim, 2018). The overexpression of *MYT1* was suggested to help reduce anxiety in rats and has been associated with intellectual disability previously (Bahi and Dreyer, 2017). *BICD1*, *UNK*, and *KIF3C* are essential for neuron differentiation during development. *BICD1*, a conserved gene in *Drosophila*, *C. elegans* (Aguirre-Chen et al., 2011), and human (Baens and Marynen, 1997), is highly expressed in the developing central and peripheral nervous systems and plays an important role in neuronal homeostasis as a regulator of neurotrophin signaling (Terenzio et al., 2014). Loss of *UNK* alters the number of neural stem cells and neural progenitors resulting in increased neurogenesis (Maierbrugger et al., 2020). *KIF3C*, one of the kinesin-related motor subunits of the *KIF3* family, is selectively expressed in the nervous system during embryonic development (Navone et al., 2001). *FBXO41* and *EPS15L1* are associated with the defects in central nervous system (Mukherjee et al., 2015; King et al., 2019; Milesi et al., 2019). *SEMA3F*, the class 3 subfamily of semaphorins, was found to contribute to the development of neuronal circuitry related to anxiety and fear responses in *SEMA3F*-knockout mice (Milesi et al., 2019). Besides, the transcription factors of *nuclear receptor TLX* (*NR2E1*) might be regarded as a key regulator of neural development and differentiation (Matsushita et al., 2014; Sun et al., 2017).

The domestication of wild animals is a process of human civilization and food production. Therefore, improving production performances is an important part for chicken domestication. More than nine genes correlated with energy metabolism, reproduction function or cell growth and proliferation were screened. Three representative genes, *GBE1*, *HEP21*, and *INHA*, were first found to have copy number variation between domesticated and wild chickens. *GBE1* is involved in the metabolism of carbohydrate, presenting one copy deletion in all six domesticated chicken breeds. This gene was reported to be related with abdominal fat traits and presented higher expression levels in fast-growing chickens than in slow-growing chickens (Claire D'andre et al., 2013) or in fat chicken line than in lean chickens (Jin et al., 2017). *HEP21*, a gene unique to poultry, has doubled copy number variation in domesticated breeds. The expression of *HEP21* was associated with the development and function of chicken oviduct (Lim and Song, 2014), and SNPs in the *HEP21* gene have been reported to be related with sexual maturity (Chen et al., 2020). *INHA* protein is secreted by the avian granulosa cells of preovulatory follicles, and its gene expression level changed with the follicle development (Cui et al., 2019).

Improving disease resistance and immunocompetence of livestock is a precondition for security and abundance of food in the domestication of animals. At least ten genes were found to participate in the pathways of inflammatory and immune response. *CBFA2T3*, a factor in human acute myeloid leukemia, promotes myeloid differentiation (Steinauer et al., 2020). In cattle, this gene is found under selection pressure, which might be related to physiological adaptations to the environment and disease resistance (Ben-Jemaa et al., 2020). *IL22RA1* is one of the receptors for IL-22, an important cytokine involved in host defense and inflammatory responses (Gaudino et al., 2020). The expression of *ARIH2* is significantly altered in chicken immune organs after vaccine immunization (Wu et al., 2019). The biological function of *TNFAIP8* has been extensively investigated and was reported to play vital roles in modulating inflammation and immunity (Lou and Liu, 2011). *GATA3* was shown to regulate the differentiation and proliferation of T-cell (Ho et al., 2009). And *HERPUD1* is found differentially expressed in chickens under stress (Guo Y. et al., 2020).

Breed-Specific Genes

Obviously, CNV plays a crucial role in chicken domestication and provides us a new perspective for studying the genetic mechanism underlying complex phenotypes. Therefore, copy number information exclusive to each domestic breed was screened to explore the genes contributing to breed characteristics. Some genes were discovered by previous study, such as *EDN3*, *SOCS2*, *HOXB7*, and *HOXB8*, but most genes relating to breed characteristics were identified for the first time. Among the discovered genes, some are already verified contributing to appearance, growth or reproduction traits in chicken, and the others, whose biological function are discussed based on studies in human or other animals, are worthy of further researches.

For RW chickens, the significantly high copy number duplications of *POLR3H*, *MCM9*, *DOCK3*, *AKR1B1L*, *ASF1A*, *ATP2B4*, and *MYH1B* may contribute to the fast-growing traits. Knocking down any of the three genes, that is, *POLR3H*, *MCM9*, and *DOCK3*, would cause delayed development or short growth (Lutzmann et al., 2012; Qin et al., 2015; Iwata-Otsubo et al., 2018; Franca et al., 2019; Guo T. et al., 2020). *AKR1B1L*, a member of the aldo/keto reductase superfamily, was related to body length and fat deposition in cattle (An et al., 2020) and chickens (Claire D'andre et al., 2013). *ASF1A* was reported to be a candidate gene for muscle weight and meat quality (Liu et al., 2013; Zhang et al., 2020). *ATP2B4* is one of ATPase family that regulates intracellular calcium homeostasis (De Koning et al., 2001) and plays an important role in bone development (Wu et al., 2004;

Kim et al., 2012). Furthermore, *MYH1B* was also demonstrated to participate in the early development of breast muscle or regeneration of muscle fibers in broilers (Pampouille et al., 2018).

For WL chickens, the copy number deletion of *CSMD1* and *NTRK3* may be involved in the outstanding reproduction traits, because *NTRK3* and *CSMD1* have been reported to play crucial roles in follicle development, ovarian quality, and infertility in human and chicken (Lachman et al., 2007; Nilsson et al., 2009; Lee et al., 2019).

In SILK chickens, which are famous for their black skin and bone, four genes (*EDN3*, *SLMO2*, *TUBB1*, and *GFPT1*) exhibited a significant copy number increase, which may be responsible for melanin formation and deposition. *EDN3*, a well-known candidate for melanocytic proliferation and maintenance, has been reported to be strongly related to dermal pigmentation in SILK chickens (Shinomiya et al., 2012). *SLMO2* and *TUBB1* have also been discovered with around 4 and 5 copy number duplications, respectively, and they showed significantly high expression in both skin and muscle tissues in SILK chickens (Dorshorst et al., 2011). *GFPT1* mutation has been reported to specifically affect the production of melanin in zebrafish (Yang et al., 2007).

LXG chickens, a typical cockfighting breed characterized by aggressiveness, strong bone strength, and powerful muscle, exclusively exhibited duplications on *SOCS2*, *PRDM5*, *CRADD*, and *SORCS2*. *SOCS2* (Lorentzon et al., 2005) and *PRDM5* (Galli et al., 2012) may contribute to the high bone mineral density. And the increased copy number and gene expression variation of *SOCS2* in gamecocks were validated by Bi et al. (2017). The one more copy number increase of *CRADD* in LXG may be responsible for fast muscle growth trait (Horvat and Medrano, 1998; Smith et al., 2000). Meanwhile, a CNV region (GGA: 80,834,764–80,851,024 bp) overlapping the gene *SORCS2* was particularly detected in LXG. *SORCS2* was suggested to be a candidate gene for the aggressive behavior of LXG, because knockdown of *SORCS2* would significantly decrease the expression of dopamine receptor genes and nerve growth factor (Li et al., 2016).

YOU chicken is a Chinese native breed with special appearance characteristics, including crests, beards, feathered legs, and polydactyly. Recently, Yang et al. (2020) deciphered that the copy number variation in *HOXB7* and *HOXB8* is responsible for the beard trait, which coincides exactly with the CNV detected in our study, especially exhibiting two copy number increase in YOU chicken.

CONCLUSION

This study provides genome-wide CNV information in one wild chicken breed (RJF), two commercial breeds, and four Chinese local breeds to synthetically understand the process of chicken domestication and to exclusively identify the genes contributing to breed characteristics. A total of 11,123 CNVRs with 19,329 duplications and 98,736 deletions were detected, covering 7% of the autosome genome and overlapping 2,636 protein-coding genes. The PCA results revealed that CNV is a powerful method for exploring evolution and domestication. In chicken evolution, genes related to nervous system, sensory, and follicle development were supposed to have pivotal roles. Meanwhile, cognitive and behavior changes, reproduction function, metabolism, and immunity are under strong selection during long-term animal domestication process. What's more, some valuable genes contributing to fast growth, high reproduction, and distinct breed characteristics were also identified in this paper, which deserve further investigations. Taken together, the present study provides numerous copy number information for chickens and it is a valuable resource to facilitate genetic and functional investigation of domestication and economic traits in chickens.

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 authors.

ETHICS STATEMENT

The animal study was reviewed and approved by all the experimental procedures involving animals were conducted in accordance with the animal welfare guidelines of ARRIVE guidelines (Percie Du Sert et al., 2020), and were approved by the Animal Experimental Ethics Board in Beijing Institute of Genomics, Chinese Academy of Science (permit number: 2017-YJ-01). In addition, all methods were carried out in accordance with relevant guidelines and regulations in the declaration.

AUTHOR CONTRIBUTIONS

XB and YH initiated the study and designed the project. XC, XB, ZY, QC, and YH collected the samples. QC, YH, and HL supervised the study. XB, BZ, and YH performed the bioinformatics. XC, QC, and YH summarized the results. XC, XB, QC, and YH drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This work is financially supported by the Beijing Municipal Science and Technology Project (D171100007817003), Innovation Program of Beijing Academy of Agriculture and Forestry Sciences (KJCX20200101), and the China Agriculture Research System of MOF and MARA (CARS-41).

ACKNOWLEDGMENTS

The authors are thankful to Dr. Ying Yu in China Agricultural University, Dr. Guiping Zhao in Institute of Animal Sciences,

Chinese Academy of Agricultural Sciences, Dr. Xiquan Zhang in Huanan Agricultural University, and Dr. Jianbin Li in Shandong Academy of Agricultural Sciences, for their kindly help and donation of chicken samples for the purpose of scientific research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.830393/full#supplementary-material

REFERENCES

Abe, H., Aoya, D., Takeuchi, H.-A., and Inoue-Murayama, M. (2017). Gene Expression Patterns of Chicken Neuregulin 3 in Association with Copy Number Variation and Frameshift Deletion. *BMC Genet.* 18, 69. doi:10.1186/s12863-017-0537-z

Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. (2011). CNVnator: an Approach to Discover, Genotype, and Characterize Typical and Atypical CNVs from Family and Population Genome Sequencing. *Genome Res.* 21, 974–984. doi:10.1101/gr.114876.110

Aguirre-Chen, C., Bülow, H. E., and Kaprielian, Z. (2011). *C. elegans* Bicd-1, Homolog of the Drosophila Dynein Accessory Factor Bicaudal D, Regulates the Branching of PVD Sensory Neuron Dendrites. *Development (Cambridge, England).* 138, 507–518. doi:10.1242/dev.060939

Albert, F. W., Shchepina, O., Winter, C., Römpler, H., Teupser, D., Palme, R., et al. (2008). Phenotypic Differences in Behavior, Physiology and Neurochemistry between Rats Selected for Tameness and for Defensive Aggression towards Humans. *Horm. Behav.* 53, 413–421. doi:10.1016/j.yhbeh.2007.11.010

Alkan, C., Kidd, J. M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., et al. (2009). Personalized Copy Number and Segmental Duplication Maps Using Next-Generation Sequencing. *Nat. Genet.* 41, 1061–1067. doi:10.1038/ng.437

An, B. X., Xu, L., Xia, J. W., Wang, X. Q., Miao, J., Chang, T. P., et al. (2020). Multiple Association Analysis of Loci and Candidate Genes that Regulate Body Size at Three Growth Stages in Simmental Beef Cattle. *Bmc Genet.* 21, 32. doi:10.1186/s12863-020-0837-6

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

Arlt, M. F., Wilson, T. E., and Glover, T. W. (2012). Replication Stress and Mechanisms of CNV Formation. *Curr. Opin. Genet. Development.* 22, 204–210. doi:10.1016/j.gde.2012.01.009

Baens, M., and Marynen, P. (1997). A Human Homologue (BICD1) of theDrosophila Bicaudal-DGene. *Genomics.* 45, 601–606. doi:10.1006/geno.1997.4971

Bahi, A., and Dreyer, J.-L. (2017). Viral-mediated Overexpression of the Myelin Transcription Factor 1 (MyT1) in the Dentate Gyrus Attenuates Anxiety- and Ethanol-Related Behaviors in Rats. *Psychopharmacology.* 234, 1829–1840. doi:10.1007/s00213-017-4588-7

Ben-Jemaa, S., Mastrangelo, S., Lee, S.-H., Lee, J. H., and Boussaha, M. (2020). Genome-wide Scan for Selection Signatures Reveals Novel Insights into the Adaptive Capacity in Local North African Cattle. *Sci. Rep.* 10, 19466. doi:10.1038/s41598-020-76576-3

Benson, G. (1999). Tandem Repeats Finder: a Program to Analyze DNA Sequences. *Nucleic Acids Res.* 27, 573–580. doi:10.1093/nar/27.2.573

Beroukhim, R., Zhang, X., and Meyerson, M. (2016). Copy Number Alterations Unmasked as Enhancer Hijackers. *Nat. Genet.* 49, 5–6. doi:10.1038/ng.3754

Bi, H., Yi, G., and Yang, N. (2017). Increased Copy Number of SOCS2 Gene in Chinese Gamecocks. *Poult. Sci.* 96, 1041–1044. doi:10.3382/ps/pew391

Biscotti, M. A., Barucca, M., Carducci, F., and Canapa, A. (2018). New Perspectives on the Evolutionary History of Vitellogenin Gene Family in Vertebrates. *Genome Biol. Evol.* 10, 2709–2715. doi:10.1093/gbe/evy206

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., et al. (2011). The Evolution of Gene Expression Levels in Mammalian Organs. *Nature.* 478, 343–348. doi:10.1038/nature10532

Brawand, D., Wahli, W., and Kaessmann, H. (2008). Loss of Egg Yolk Genes in Mammals and the Origin of Lactation and Placentation. *Plos Biol.* 6, e63. doi:10.1371/journal.pbio.0060063

Bu, P., Jian, Z., Koshy, J., Shen, Y., Yue, B., and Fan, Z. (2019). The Olfactory Subgenome and Specific Odor Recognition in forest Musk Deer. *Anim. Genet.* 50, 358–366. doi:10.1111/age.12796

Charles Darwin, M. A. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Chen, B., Liang, G., Zhu, X., Tan, Y., Xu, J., Wu, H., et al. (2020). Gene Expression Profiling in Ovaries and Association Analyses Reveal as a Candidate Gene for Sexual Maturity in Chickens. *Anim. : open access J. MDPI.* 10, 181. doi:10.3390/ani10020181

Claire D'andre, H., Paul, W., Shen, X., Jia, X., Zhang, R., Sun, L., et al. (2013). Identification and Characterization of Genes that Control Fat Deposition in Chickens. *J. Anim. Sci. Biotechnol.* 4 (1), 43. doi:10.1186/2049-1891-4-43

Crooijmans, R. P., Fife, M. S., Fitzgerald, T. W., Strickland, S., Cheng, H. H., Kaiser, P., et al. (2013). Large Scale Variation in DNA Copy Number in Chicken Breeds. *BMC Genomics.* 14, 398. doi:10.1186/1471-2164-14-398

Cui, Z., Liu, L., Zhao, X., Ran, J., Wang, Y., Yin, H., et al. (2019). Analysis of Expression and Single Nucleotide Polymorphisms of INHA Gene Associated with Reproductive Traits in Chickens. *Biomed. Res. Int.* 2019, 8572837. doi:10.1155/2019/8572837

Darwin, C. (2010). *The Variation of Animals and Plants under Domestication*. Cambridge University Press. Vol. 2.

De Koning, D. J., Harlizius, B., Rattink, A. P., Groenen, M. A., Brascamp, E. W., and Van Arendonk, J. A. (2001). Detection and Characterization of Quantitative Trait Loci for Meat Quality Traits in Pigs. *J. Anim. Sci.* 79, 2812–2819. doi:10.2527/2001.79112812x

Dorshorst, B., Molin, A.-M., Rubin, C.-J., Johansson, A. M., Strömstedt, L., Pham, M.-H., et al. (2011). A Complex Genomic Rearrangement Involving the Endothelin 3 Locus Causes Dermal Hyperpigmentation in the Chicken. *Plos Genet.* 7, e1002412. doi:10.1371/journal.pgen.1002412

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, 81–84. doi:10.1038/nature10757

Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural Variation in the Human Genome. *Nat. Rev. Genet.* 7, 85–97. doi:10.1038/nrg1767

Finn, R., Kolarevic, J., Kongshaug, H., and Nilsen, F. (2009). Evolution and Differential Expression of a Vertebrate Vitellogenin Gene Cluster. *BMC Evol. Biol.* 9, 2. doi:10.1186/1471-2148-9-2

Franca, M. M., Han, X., Funari, M. F. A., Lerario, A. M., Nishi, M. Y., Fontenele, E. G. P., et al. (2019). Exome Sequencing Reveals the POLR3H Gene as a Novel Cause of Primary Ovarian Insufficiency. *J. Clin. Endocrinol. Metab.* 104, 2827–2841. doi:10.1210/jc.2018-02485

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., et al. (2016). Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications. *Nature.* 538, 265–269. doi:10.1038/nature19800

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., Mccarroll, S. A., Altshuler, D. M., et al. (2006). Copy Number Variation: New Insights in Genome Diversity. *Genome Res.* 16, 949–961. doi:10.1101/gr.3677206

Galli, G. G., Honnens De Lichtenberg, K., Carrara, M., Hans, W., Wuelling, M., Mentz, B., et al. (2012). Prdm5 Regulates Collagen Gene Transcription by Association with RNA Polymerase II in Developing Bone. *Plos Genet.* 8, e1002711. doi:10.1371/journal.pgen.1002711

Gaudino, S. J., Beaupre, M., Lin, X., Joshi, P., Rathi, S., Mclaughlin, P. A., et al. (2020). IL-22 Receptor Signaling in Paneth Cells Is Critical for Their Maturation, Microbiota Colonization, Th17-Related Immune Responses, and Anti-Salmonella Immunity. *Mucosal Immunol.* 14 (2), 389–401. doi:10.1038/s41385-020-00348-5

Giuffra, E., Törnsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J. M. H., et al. (2002). A Large Duplication Associated with Dominant white Color in Pigs Originated by Homologous Recombination between LINE Elements Flanking KIT. *Mamm. Genome.* 13, 569–577. doi:10.1007/s00335-002-2184-5

Guo, T., Zheng, Y., Li, G., Zhao, S., Ma, J., and Qin, Y. (2020). Novel Pathogenic Mutations in Minichromosome Maintenance Complex Component 9 (MCM9)

Responsible for Premature Ovarian Insufficiency. *Fertil. sterility.* 113, 845–852. doi:10.1016/j.fertnstert.2019.11.015

Guo, Y., Jiang, R., Su, A., Tian, H., Zhang, Y., Li, W., et al. (2020). Identification of Genes Related to Effects of Stress on Immune Function in the Spleen in a Chicken Stress Model Using Transcriptome Analysis. *Mol. Immunol.* 124, 180–189. doi:10.1016/j.molimm.2020.06.004

Han, R., Yang, P., Tian, Y., Wang, D., Zhang, Z., Wang, L., et al. (2014). Identification and Functional Characterization of Copy Number Variations in Diverse Chicken Breeds. *BMC Genomics.* 15, 934. doi:10.1186/1471-2164-15-934

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009). Mechanisms of Change in Gene Copy Number. *Nat. Rev. Genet.* 10, 551–564. doi:10.1038/nrg2593

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., et al. (2004). Sequence and Comparative Analysis of the Chicken Genome Provide Unique Perspectives on Vertebrate Evolution. *Nature.* 432, 695–716. doi:10.1038/nature03154

Ho, I.-C., Tai, T.-S., and Pai, S.-Y. (2009). GATA3 and the T-Cell Lineage: Essential Functions before and after T-Helper-2-Cell Differentiation. *Nat. Rev. Immunol.* 9, 125–135. doi:10.1038/nri2476

Horvat, S., and Medrano, J. F. (1998). A 500-kb YAC and BAC Contig Encompassing the High-Growth Deletion in Mouse Chromosome 10 and Identification of the MurineRaidd/CraddGene in the Candidate Region. *Genomics.* 54, 159–164. doi:10.1006/geno.1998.5540

Hou, Y., Qi, F., Bai, X., Ren, T., Shen, X., Chu, Q., et al. (2020). Genome-wide Analysis Reveals Molecular Convergence Underlying Domestication in 7 Bird and Mammals. *BMC Genomics.* 21 (1), 204. doi:10.1186/s12864-020-6613-1

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of Large-Scale Variation in the Human Genome. *Nat. Genet.* 36, 949–951. doi:10.1038/ng1416

International Chicken Genome Sequencing, C (2004). Sequence and Comparative Analysis of the Chicken Genome Provide Unique Perspectives on Vertebrate Evolution. *Nature.* 432, 695–716. doi:10.1038/nature03154

Iwata-Otsubo, A., Ritter, A. L., Weckselbatt, B., Ryan, N. R., Burgess, D., Conlin, L. K., et al. (2018). DOCK3 -related Neurodevelopmental Syndrome: Biallelic Intragenic Deletion of DOCK3 in a Boy with Developmental Delay and Hypotonia. *Am. J. Med. Genet.* 176, 241–245. doi:10.1002/ajmg.a.38517

Jin, P., Wu, X., Xu, S., Zhang, H., Li, Y., Cao, Z., et al. (2017). Differential Expression of Six Genes and Correlation with Fatness Traits in a Unique Broiler Population. *Saudi J. Biol. Sci.* 24, 945–949. doi:10.1016/j.sjbs.2015.04.014

Ka, M., and Kim, W.-Y. (2018). ANKRD11 Associated with Intellectual Disability and Autism Regulates Dendrite Differentiation via the BDNF/TrkB Signaling Pathway. *Neurobiol. Dis.* 111, 138–152. doi:10.1016/j.nbd.2017.12.008

Käsdorf, B. T., Weber, F., Petrou, G., Srivastava, V., Crouzier, T., and Lieleg, O. (2017). Mucin-Inspired Lubrication on Hydrophobic Surfaces. *Biomacromolecules.* 18, 2454–2462. doi:10.1021/acs.biomac.7b00605

Kim, H. J., Prasad, V., Hyung, S.-W., Lee, Z. H., Lee, S.-W., Bhargava, A., et al. (2012). Plasma Membrane Calcium ATPase Regulates Bone Mass by fine-tuning Osteoclast Differentiation and Survival. *J. cell Biol.* 199, 1145–1158. doi:10.1083/jcb.201204067

King, C. R., A. A. Quadros, A. R., Chazeau, A., Saarloos, I., van der Graaf, A. J., Verhage, M., et al. (2019). Fbxo41 Promotes Disassembly of Neuronal Primary Cilia. *Sci. Rep.* 9, 8179. doi:10.1038/s41598-019-44589-2

Lachman, H. M., Pedrosa, E., Petruolo, O. A., Cockerham, M., Papolos, A., Novak, T., et al. (2007). Increase inGSK3β Gene Copy Number Variation in Bipolar Disorder. *Am. J. Med. Genet.* 144B, 259–265. doi:10.1002/ajmg.b.30498

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current Perspectives and the Future of Domestication Studies. *Proc. Natl. Acad. Sci.* 111, 6139–6146. doi:10.1073/pnas.1323964111

Lee, A. S., Rusch, J., Lima, A. C., Usmani, A., Huang, N., Lepamets, M., et al. (2019). Rare Mutations in the Complement Regulatory Gene CSMD1 Are Associated with Male and Female Infertility. *Nat. Commun.* 10, 4626. doi:10.1038/s41467-019-12522-w

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

Li, Z., Liu, L., Lin, W., Zhou, Y., Zhang, G., Du, X., et al. (2020). NRG3 Contributes to Cognitive Deficits in Chronic Patients with Schizophrenia. *Schizophrenia Res.* 215, 134–139. doi:10.1016/j.schres.2019.10.060

Li, Z., Zheng, M., Abdalla, B. A., Zhang, Z., Xu, Z., Ye, Q., et al. (2016). Genome-wide Association Study of Aggressive Behaviour in Chicken. *Sci. Rep.* 6, 30981. doi:10.1038/srep30981

Lim, W., and Song, G. (2014). Pivotal Roles for Hormonally Regulated Expression of the HEP21 Gene in the Reproductive Tract of Chickens for Oviduct Development and in Ovarian Carcinogenesis. *Domest. Anim. Endocrinol.* 48, 136–144. doi:10.1016/j.domaniend.2014.03.003

Liu, R., Sun, Y., Zhao, G., Wang, F., Wu, D., Zheng, M., et al. (2013). Genome-wide Association Study Identifies Loci and Candidate Genes for Body Composition and Meat Quality Traits in Beijing-You Chickens. *PLoS One.* 8, e61172. doi:10.1371/journal.pone.0061172

Liu, Y., Zhang, M., Sun, J., Chang, W., Sun, M., Zhang, S., et al. (2020). Comparison of Multiple Algorithms to Reliably Detect Structural Variants in Pears. *BMC genomics.* 21, 61. doi:10.1186/s12864-020-6455-x

Lorentzon, M., Greenhalgh, C. J., Mohan, S., Alexander, W. S., and Ohlsson, C. (2005). Reduced Bone mineral Density in SOCS-2-Deficient Mice. *Pediatr. Res.* 57, 223–226. doi:10.1203/01.pdr.0000148735.21084.d3

Lou, Y., and Liu, S. (2011). The TIPE (TNFAIP8) Family in Inflammation, Immunity, and Cancer. *Mol. Immunol.* 49, 4–7. doi:10.1016/j.molimm.2011.08.006

Lutzmann, M., Grey, C., Traver, S., Ganier, O., Maya-Mendoza, A., Ranisavljevic, N., et al. (2012). MCM8- and MCM9-Deficient Mice Reveal Gametogenesis Defects and Genome Instability Due to Impaired Homologous Recombination. *Mol. Cell.* 47, 523–534. doi:10.1016/j.molcel.2012.05.048

Maierbrugger, K. T., Sousa-Nunes, R., and Bateman, J. M. (2020). The mTOR Pathway Component Unkempt Regulates Neural Stem Cell and Neural Progenitor Cell Cycle in the Drosophila central Nervous System. *Developmental Biol.* 461, 55–65. doi:10.1016/j.ydbio.2020.01.006

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., et al. (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *Am. J. Hum. Genet.* 82, 477–488. doi:10.1016/j.ajhg.2007.12.009

Matsushita, F., Kameyama, T., Kadokawa, Y., and Marunouchi, T. (2014). Spatiotemporal Expression Pattern of Myt/NZF Family Zinc finger Transcription Factors during Mouse Nervous System Development. *Dev. Dyn.* 243, 588–600. doi:10.1002/dvdy.24091

Meier, S., Strohmaier, J., Breuer, R., Mattheisen, M., Degenhardt, F., Mühleisen, T. W., et al. (2013). Neuregulin 3 Is Associated with Attention Deficits in Schizophrenia and Bipolar Disorder. *Int. J. Neuropsychopharmacol.* 16, 549–556. doi:10.1017/s1461145712000697

Milesi, C., Alberici, P., Pozzi, B., Oldani, A., Beznoussenko, G. V., Raimondi, A., et al. (2019). Redundant and Nonredundant Organismal Functions of EPS15 and EPS15L1. *Life Sci. alliance.* 2 (1), e201800273. doi:10.26508/lsa.201800273

Minias, P., Pikus, E., Whittingham, L. A., and Dunn, P. O. (2019). Evolution of Copy Number at the MHC Varies across the Avian Tree of Life. *Genome Biol. Evol.* 11, 17–28. doi:10.1093/gbe/evy253

Mukherjee, C., Holubowska, A., Schwedhelm-Domeyer, N., Mitkovski, M., Lee, S.-J., Kannan, M., et al. (2015). Loss of the Neuron-specific F-Box Protein FBXO41 Models an Ataxia-like Phenotype in Mice with Neuronal Migration Defects and Degeneration in the Cerebellum. *J. Neurosci.* 35, 8701–8717. doi:10.1523/jneurosci.2133-14.2015

Navone, F., Consalez, G. G., Sardella, M., Caspani, E., Pozzoli, O., Frassoni, C., et al. (2001). Expression of KIF3C Kinesin during Neural Development and *In Vitro* Neuronal Differentiation. *J. Neurochem.* 77, 741–753. doi:10.1046/j.1471-4159.2001.00277.x

Nilsson, E., Dole, G., and Skinner, M. K. (2009). Neurotrophin NT3 Promotes Ovarian Primordial to Primary Follicle Transition. *Reproduction (Cambridge, England).* 138, 697–707. doi:10.1530/REP-09-0179

Pampouille, E., Berri, C., Boitard, S., Hennequet-Antier, C., Beauclercq, S. A., Godet, E., et al. (2018). Mapping QTL for white Striping in Relation to Breast Muscle Yield and Meat Quality Traits in Broiler Chickens. *BMC genomics.* 19, 202. doi:10.1186/s12864-018-4598-9

Pei, S. W., Qin, F., Li, W. H., Li, F. D., and Yue, X. P. (2019). Copy Number Variation of ZNF280AY across 21 Cattle Breeds and its Association with the Reproductive Traits of Holstein and Simmental Bulls. *J. Dairy Sci.* 102, 7226–7236. doi:10.3168/jds.2018-16063

Percie Du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., et al. (2020). The ARRIVE Guidelines 2.0: Updated Guidelines for Reporting Animal Research. *Plos Biol.* 18, e3000410. doi:10.1371/journal.pbio.3000410

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., et al. (2007). Diet and the Evolution of Human Amylase Gene Copy Number Variation. *Nat. Genet.* 39, 1256–1260. doi:10.1038/ng2123

Peters, J., Lebrasseur, O., Deng, H., and Larson, G. (2016). Holocene Cultural History of Red Jungle Fowl (Gallus gallus) and its Domestic Descendant in East Asia. *Quat. Sci. Rev.* 142, 102–119. doi:10.1016/j.quascirev.2016.04.004

Qin, Y., Jiao, X., Simpson, J. L., and Chen, Z.-J. (2015). Genetics of Primary Ovarian Insufficiency: New Developments and Opportunities. *Hum. Reprod. Update.* 21, 787–808. doi:10.1093/humupd/dmv036

Ramakrishnan, N., Sharma, S., Gupta, A., and Alashwal, B. Y. (2018). Keratin Based Bioplastic Film from Chicken Feathers and its Characterization. *Int. J. Biol. macromolecules.* 111, 352–358. doi:10.1016/j.ijbiomac.2018.01.037

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global Variation in Copy Number in the Human Genome. *Nature.* 444, 444–454. doi:10.1038/nature05329

Rice, A. M., and Mclysaght, A. (2017). Dosage Sensitivity Is a Major Determinant of Human Copy Number Variant Pathogenicity. *Nat. Commun.* 8, 14366. doi:10.1038/ncomms14366

Roy, M. G., Livraghi-Butrico, A., Fletcher, A. A., Mcelwee, M. M., Evans, S. E., Boerner, R. M., et al. (2014). Muc5b Is Required for Airway Defence. *Nature.* 505, 412–416. doi:10.1038/nature12807

Sánchez-Villagra, M. R., Geiger, M., and Schneider, R. A. (2016). The Taming of the Neural Crest: a Developmental Perspective on the Origins of Morphological Covariation in Domesticated Mammals. *R. Soc. Open Sci.* 3, 160107. doi:10.1098/rsos.160107

Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., et al. (2012). Insights into Hominid Evolution from the gorilla Genome Sequence. *Nature.* 483, 169–175. doi:10.1038/nature10842

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al. (2004). Large-scale Copy Number Polymorphism in the Human Genome. *Science.* 305, 525–528. doi:10.1126/science.1098918

Shinomiya, A., Kayashima, Y., Kinoshita, K., Mizutani, M., Namikawa, T., Matsuda, Y., et al. (2012). Gene Duplication of Endothelin 3 Is Closely Correlated with the Hyperpigmentation of the Internal Organs (Fibromelanosis) in Silky Chickens. *Genetics.* 190, 627–638. doi:10.1534/genetics.111.136705

Smit, A., and Green, P. (2013). RepeatMasker Open-4.0. Available at: http://www.repeatmasker.org.

Smith, J., Paton, I. R., Horvat, S., Medrano, J. F., and Burt, D. W. (2000). Mapping the RAIDD Gene of Chicken (Gallus gallus): Identification of a Region Homologous to the Mouse High-Growth Region. *Mamm. Genome.* 11, 706–709. doi:10.1007/s003350010140

Sohrabi, S. S., Mohammadabadi, M., Wu, D.-D., and Esmailizadeh, A. (2018). Detection of Breed-specific Copy Number Variations in Domestic Chicken Genome. *Genome.* 61, 7–14. doi:10.1139/gen-2017-0016

Steinauer, N., Guo, C., and Zhang, J. (2020). The Transcriptional Corepressor CBFA2T3 Inhibits All-Trans-Retinoic Acid-Induced Myeloid Gene Expression and Differentiation in Acute Myeloid Leukemia. *J. Biol. Chem.* 295, 8887–8900. doi:10.1074/jbc.RA120.013042

Sun, G., Cui, Q., and Shi, Y. (2017). Nuclear Receptor TLX in Development and Diseases. *Curr. Top. Dev. Biol.* 125, 257–273. doi:10.1016/bs.ctdb.2016.12.003

Team, R. C. (2021). R: A Language and Environment for Statistical Computing. *R. Found. Stat. Comput*

Terenzio, M., Golding, M., Russell, M. R. G., Wicher, K. B., Rosewell, I., Spencer-Dene, B., et al. (2014). Bicaudal- D 1 Regulates the Intracellular Sorting and Signalling of Neurotrophin Receptors. *Embo J.* 33, 1582–1598. doi:10.15252/embj.201387579

Tian, M., Wang, Y., Gu, X., Feng, C., Fang, S., Hu, X., et al. (2013). Copy Number Variants in Locally Raised Chinese Chicken Genomes Determined Using Array Comparative Genomic Hybridization. *BMC genomics.* 14, 262. doi:10.1186/1471-2164-14-262

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: Discovering Splice Junctions with RNA-Seq. *Bioinformatics (Oxford, England).* 25, 1105–1111. doi:10.1093/bioinformatics/btp120

Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., et al. (2015). Genomic Analyses Reveal Potential Independent Adaptation to High Altitude in Tibetan Chickens. *Mol. Biol. Evol.* 32, 1880–1889. doi:10.1093/molbev/msv071

Wang, M.-S., Thakur, M., Peng, M.-S., Jiang, Y., Frantz, L. A. F., Li, M., et al. (2020). 863 Genomes Reveal the Origin and Domestication of Chicken. *Cell Res.* 30, 693–701. doi:10.1038/s41422-020-0349-y

Wang, Y., Chen, C. Z., Fu, X. H., Liu, J. B., Peng, Y. X., Wang, Y. J., et al. (2020). CPEB3 Regulates the Proliferation and Apoptosis of Bovine Cumulus Cells. *Anim. Sci. J.* 91, e13416. doi:10.1111/asj.13416

Wilkins, A. S., Wrangham, R. W., and Fitch, W. T. (2014). The "domestication Syndrome" in Mammals: a Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics.* 197, 795–808. doi:10.1534/genetics.114.165423

Winther, M., Berezin, V., and Walmod, P. S. (2012). NCAM2/OCAM/RNCAM: Cell Adhesion Molecule with a Role in Neuronal Compartmentalization. *Int. J. Biochem. Cell Biol.* 44, 441–446. doi:10.1016/j.biocel.2011.11.020

Wright, D., Boije, H., Meadows, J. R. S., Bed'hom, B., Gourichon, D., Vieaud, A., et al. (2009). Copy Number Variation in Intron 1 of SOX5 Causes the Pea-Comb Phenotype in Chickens. *Plos Genet.* 5, e1000512. doi:10.1371/journal.pgen.1000512

Wu, G., Xu, S., Zhang, W., Liu, Y., Wang, Q., and Man, C. (2019). Arih2 Gene Influences Immune Response and Tissue Development in Chicken. *Biosci. Rep.* 39 (10), BSR20190933. doi:10.1042/BSR20190933

Wu, X., Zhu, Z., Yerle, M., Wang, H. L., Wang, H., Gu, M., et al. (2004). Radiation Hybrid Mapping of Four Genes (MYBPC1,LUM,ZRF1andATP2B4) Expressed in Embryo Skeleton Muscle to Pig Chromosomes 5 and 9. *Anim. Genet.* 35, 472–473. doi:10.1111/j.1365-2052.2004.01195.x

Xu, L., He, Y., Ding, Y., Sun, G., Carrillo, J. A., Li, Y., et al. (2017). Characterization of Copy Number Variation's Potential Role in Marek's Disease. *Int. J. Mol. Sci.* 18, 1020. doi:10.3390/ijms18051020

Yang, C.-T., Hindes, A. E., Hultman, K. A., and Johnson, S. L. (2007). Mutations in Gfpt1 and Skiv2l2 Cause Distinct Stage-specific Defects in Larval Melanocyte Regeneration in Zebrafish. *Plos Genet.* 3, e88. doi:10.1371/journal.pgen.0030088

Yang, K. X., Zhou, H., Ding, J. M., He, C., Niu, Q., Gu, C. J., et al. (2020). Copy Number Variation in HOXB7 and HOXB8 Involves in the Formation of Beard Trait in Chickens. *Anim. Genet.* 51, 958–963. doi:10.1111/age.13011

Yi, G., Qu, L., Liu, J., Yan, Y., Xu, G., and Yang, N. (2014). Genome-wide Patterns of Copy Number Variation in the Diversified Chicken Genomes Using Next-Generation Sequencing. *BMC Genomics.* 15, 962. doi:10.1186/1471-2164-15-962

Zhang, H., Du, Z.-Q., Dong, J.-Q., Wang, H.-X., Shi, H.-Y., Wang, N., et al. (2014). Detection of Genome-wide Copy Number Variations in Two Chicken Lines Divergently Selected for Abdominal Fat Content. *BMC Genomics.* 15, 517. doi:10.1186/1471-2164-15-517

Zhang, H., Shen, L.-Y., Xu, Z.-C., Kramer, L. M., Yu, J.-Q., Zhang, X.-Y., et al. (2020). Haplotype-based Genome-wide Association Studies for Carcass and Growth Traits in Chicken. *Poult. Sci.* 99, 2349–2361. doi:10.1016/j.psj.2020.01.009

Zhang, Z., Jia, Y., Almeida, P., Mank, J. E., Van Tuinen, M., Wang, Q., et al. (2018). Whole-genome Resequencing Reveals Signatures of Selection and Timing of Duck Domestication. *Gigascience.* 7 (4), giy027. doi:10.1093/gigascience/giy027

Zhou, C., Tu, H. M., Yu, H. R., Zheng, S., Dai, B., Price, M., et al. (2019). The Draft Genome of the Endangered Sichuan Partridge (*Arborophila rufipectus*) with Evolutionary Implications. *Genes.* 10. doi:10.3390/genes10090677

Zhou, J., Lemos, B., Dopman, E. B., and Hartl, D. L. (2011). Copy-Number Variation: The Balance between Gene Dosage and Expression in *Drosophila melanogaster*. *Genome Biol. Evol.* 3, 1014. doi:10.1093/gbe/evr023

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fgene.2022.860669

Population Genetic Structure and Selection Signature Analysis of Beijing Black Pig

Wenjing Yang[1†], Zhen Liu[1†], Qiqi Zhao[1†], Heng Du[1], Jian Yu[1], Hongwei Wang[2], Xiance Liu[2], Hai Liu[2], Xitao Jing[2], Hongping Yang[2], Guohua Shi[2], Lei Zhou[1] and Jianfeng Liu[1]**

[1]College of Animal Science and Technology, China Agricultural University, Beijing, China, [2]Beijing Heiliu Stockbreeding Technology Co.,Ltd, Beijing, China

Beijing Black pig is an excellent cultivated black pig breed in China, with desirable body shape, tender meat quality and robust disease resistance. To explore the level of admixture and selection signatures of Beijing Black pigs, a total number of 90 individuals covering nine pig breeds were used in our study, including Beijing Black pig, Large White, Landrace, Duroc, Lantang pig, Luchuan pig, Mashen pig, Huainan pig and Min pig. These animals were resequenced with 18.19 folds mapped read depth on average. Generally, we found that Beijing Black pig was genetically closer to commercial pig breeds by population genetic structure and genetic diversity analysis, and was also affected by Chinese domestic breeds Huainan pig and Min pig. These results are consistent with the cross-breeding history of Beijing Black pig. Selection signal detections were performed on three pig breeds, Beijing Black pig, Duroc and Large White, using three complementary methods (F_{ST}, $\theta\pi$, and XP-EHH). In total, 1,167 significant selected regions and 392 candidate genes were identified. Functional annotations were enriched to pathways related to immune processes and meat and lipid metabolism. Finally, potential candidate genes, influencing meat quality (*GPHA2, EHD1*, *HNF1A*, *C12orf43*, *GLTP*, *TRPV4*, *MVK*, and *MMAB*), reproduction (*PPP2R5B* and *MAP9*), and disease resistance (*OASL*, *ANKRD13A*, and *GIT2*), were further detected by gene annotation analysis. Our results advanced the understanding of the genetic mechanism behind artificial selection of Beijing Black pigs, and provided theoretical basis for the subsequent breeding and genetic research of this breed.

Keywords: whole-genome sequencing, beijing black pig, genetic diversity, selection regions, candidate genes

OPEN ACCESS

Edited by:
Lingyang Xu,
Institute of Animal Sciences (CAAS), China

Reviewed by:
Shi-Yi Chen,
Sichuan Agricultural University, China
Qianjun Zhao,
Institute of Animal Sciences (CAAS), China

****Correspondence:***
Lei Zhou
leiz@cau.edu.cn
Jianfeng Liu
liujf@cau.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: *23 January 2022*
Accepted: *08 March 2022*
Published: *24 March 2022*

Citation:
Yang W, Liu Z, Zhao Q, Du H, Yu J, Wang H, Liu X, Liu H, Jing X, Yang H, Shi G, Zhou L and Liu J (2022) Population Genetic Structure and Selection Signature Analysis of Beijing Black Pig. Front. Genet. 13:860669. doi: 10.3389/fgene.2022.860669

INTRODUCTION

From early domestication to modern breeding practices, artificial selection for agricultural economic traits has shaped the genomes of domestic pigs and led to many breeds and populations worldwide (Gouveia et al., 2014). Under positive selection pressure, the frequencies of favorable alleles would increase, and it would show an unusual long-range linkage disequilibrium (LD) with a high population frequency (Smith and Haigh, 1974). Signatures of selection in the genome have been used frequently to understand the relationships between genotype and phenotype in pigs. For instance, strong selection signatures were found at three loci which were related to morphological changes in the domestic pigs using whole-genome resequencing (Rubin et al., 2012); evidence of artificial

selection of lean muscle mass, fertility and immunization traits were revealed in Duroc pigs (Ma et al., 2018).

Beijing Black pig (BJB), which is a typical composite black pig breed in China, is best known for its perfect combination of the characters of Chinese native pig breeds (i.e., superior meat quality, strong resistance to disease and desirable reproduction performance) and commercial pig breeds (i.e., fast growth rate, high lean meat rate and feed conversion efficiency). For example, this breed is renowned for its meat quality and high intramuscular fat content in pork with an average value of ~3.11% as compared to less than 2% of commercial pig breeds (Zhang et al., 2018).

Beijing Black pigs have a wide range of breed compositions. According to the breeding history of BJB, it was cross-breeded from multiple Chinese native pigs and foreign commercial pig breeds, including Chinese Northern pig breeds, Chinese Southern pig breeds, Huanghuaihai pigs, Large White, and Berkshine, etc. After many years of systematic genetic improvement programs, BJB was certificated as a new pig breed by the Ministry of Agriculture of China in 1982. In recent years, consumers are increasing focus on health and food quality, and prefer pork with higher quality and better flavors. Beijing Black pig has undoubtedly contributed to offering high-quality pork in the present-day market. However, to date, only a few studies have focused on BJB. A genome-wide association study of vertebral and teat number was performed on 891 BJB by Illumina Porcine 50 K BeadChip, and several quantitative trait loci (QTL) were identified (Niu et al., 2021). Candidate genes for skeletal muscle growth and meat quality of Beijing Black pigs were found via the RNA-seq method (Hou et al., 2021). However, the genetic basis of the characteristics of BJB, particularly at the genomic level, remains largely unknown. Specifically speaking, it is still unknown to what extent the breed contribution of Beijing Black pigs is influenced by commercial breeds and Chinese local breeds. Also, it is essential to detect selection signatures and genes related to the biological processes for the economic important traits of Beijing Black pigs, which have undergone decades of intense artificial selection.

In this study, we used whole-genome resequencing data of BJB, together with eight additional pig breeds, representing potential breed origin of BJB. These eight pig breeds including Huainan pig (HN) and Mashen (MS) pig (from Huanghuaihai district); Lantang pig (LT) and Luchuan pig (LC) (from southern China); Min pig (MIN) (from northeast China); Duroc (DU), Landrace (LD), and Large White (LW) (representing three major commercial pig breeds). We conducted a comprehensive analysis of phylogenetic relationships and genetic diversity among these pig breeds, then three selection signature detection methods were performed to identify genomic regions under selection and potential candidate genes in BJB. Our findings enable to better understand the Beijing Black pig's genome characteristics, and provide novel insights for developing breeding strategies and germplasm conservation in the near future.

MATERIALS AND METHODS

Sample Collection and Sequencing

A total number of 59 samples, including 22 Beijing Black pigs, 22 Duroc, five Lantang pigs, five Mashen pigs, and five Huainan pigs, were sequenced in this study (**Supplementary Table S1**). Genomic DNA was extracted from pig ear tissue using the Qiagen DNeasy Tissue Kit (Qiagen, Germany). Then, DNA integrity and purity were verified using agarose gel electrophoresis and A260/280 ratio. All samples were constructed from Illumina DNA library (paired end, 2 × 150 bp) and sequenced using Illumina HiSeq 2000 sequencing system. In addition, genomic data of 5 Min pigs, five Luchuan pigs, five Landrace, and 16 Large White pigs were downloaded from public domain (**Supplementary Table S1**). In total, whole genome sequencing data of 90 pig samples of nine breeds were analyzed in this study. The raw resequencing reads were filtered using fastp v0.20.1 (Chen et al., 2018), by removing reads containing adapters, low-quality reads with >30% base (quality value ≤ 20, or N bases) and low-quality 3' end reads with base quality scores of ≤20.

Variant Calling and Annotation

Clean sequencing reads were subsequently mapped to the pig reference genome Sus scrofa 11.1 using Burrows-Wheeler Aligner (BWA) v0.7.17 (Li and Durbin, 2009) with default settings. Genome Analysis Toolkit (GATK) v3.8 (Mckenna et al., 2010) was used for SNP calling. HaplotypeCaller and GenotypeGVCFs modules in GATK were jointly used to call variation. Intermediate genomic (gVCF) files were generated using the "-ERC GVCF" mode in HaplotypeCaller. Then, joint genotypes were determined using GenotypeGVCFs. High-quality SNPs, with Quality score>30, MQ RMS mapping quality >20, DP > 5, coverage >30% and minor allele frequency (MAF) > 0.01, were kept with vcftools v0.1.16 for the following analysis (Danecek et al., 2011). Then, the dbSNP database (Sherry et al., 2001) was used to identify the novel genetic variations. Finally, ANNOVAR v2019Oct24 was used to conduct gene-based or region-based annotation processing for the filtered variants (Wang et al., 2010), and the corresponding gene annotation file was downloaded from the Ensembl database (https://asia.ensembl.org/index.html).

Population Structure Analysis

A common subset of 22,112,606 SNPs resulted from the above procedures was used to infer the genetic structure of these nine pig breeds. Genetic distances between individuals and breeds were calculated using an identity-by-state (IBS) similarity matrix via Plink v1.9 software (Purcell et al., 2007). Neighbor-joining (NJ) phylogenetic tree was built based on IBS distance matrix using the PHYLIP v3.698 (Felsenstein, 1993). After that, the NJ tree was visualized via Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Principal component analysis (PCA) was performed by GCTA v1.91.7 (Yang et al., 2011), in which the genetics matrix was firstly

generated using "—make-grm" option, and then the first three principal components were calculated with "—pca3" option. Finally, the PCA plot was plotted using the R language. Moreover, the population ancestry of these breeds was inferred by ADMIXTURE v1.3.0 (Alexander et al., 2009). The optimum number of ancestral clusters K was estimated with a five-fold cross validation procedure. The ancestral clusters number K were tested from 2 to 9, using the plots of ancestry compositions of the tested breeds by a R package "Pophelper" (Francis, 2017).

Analysis of Genome Diversity and Inbreeding

The F-statistic (F_{ST}) is a population differentiation index based on genetic polymorphism data (Holsinger and Weir, 2009). Vcftools v0.1.16 was implemented to evaluate F_{ST}, and an F_{ST} matrix of size 9 × 9 were further obtained, which was graphically represented by heatmap *via* corrplot R package (Wei et al., 2017). Running of homozygosity (ROH) fragments of each individual were determined using Plink v1.9. Then, the average number of ROH fragments per breed was classified into seven categories: 100–200 kb, 200–300 kb, 300–400 kb, 400–500 kb, 500–600 kb, 600–700 kb, and >700 Kb. ROH-based inbreeding coefficient (F_{ROH}) was measured by the ratio of the total length of ROH to the length of autosomes (2.27 Gb in this study) (Mcquillan et al., 2008).

Selection Signature Detection of the Beijing Black Pig

Three methods for genomic selection signature detection, including population differentiation coefficient (F_{ST}), polymorphism levels statistic ($\theta\pi$) and cross-population extended haplotype homozygosity (XP-EHH), were performed to detect the genomic regions under selection in Beijing Black pigs by comparing with Duroc and Large White. The F_{ST} and $\theta\pi$ (DU or LW/BJB) between the Beijing Black pigs and the other two breeds were calculated with vcftools v0.1.16 using 100 kb windows with 10 kb steps among genomes. Then, $\theta\pi$ (DU or LW/BJB) was log2-transformed. The XP-EHH statistic was designed to detect ongoing or nearly fixed selection signatures by comparing haplotypes from two populations (Sabeti et al., 2007). XP-EHH values were estimated using Selscan v1.3.0 (Szpiech and Hernandez, 2014). Then, the XP-EHH value of each SNP was normalized. Extremely high values in the 5% right-tail of each method were empirically selected as potential candidate regions under positive selection.

We also estimated allele frequencies of single-nucleotide variants (SNV) with a genome scan for each pig population, and measured the absolute allele frequency difference (ΔAF) for comparing different breeds. The ΔAF per SNV between the Beijing Black population and the other two populations was calculated using the formula: $\Delta AF = abs(AltAF_{BJB}\text{-}mean(AltAF_{Du} + AltAF_{LW}))$ (Zhao et al., 2018).

Genome Annotation and Enrichment Analysis

Genes in these selection regions were identified through the ensembl database Sus scrofa 11.1 assembly.

(http://ftp.ensembl.org/pub/current_gtf/sus_scrofa/Sus_scrofa.Sscrofa11.1.105. gtf.gz). To further explore the potential biological significance of genes within these candidate regions, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried using KOBAS (Xie et al., 2011). Finally, published pig QTLs were downloaded from pig QTLdb database (Hu et al., 2022) to identify the QTL overlap with the candidate regions in our study.

RESULTS

Genomic Variant Identification in Beijing Black Pig

Whole-genome resequencing of 59 individuals from five pig breeds (BJB, DU, HN, LT, and MS) generated a total size of 2,196 Gb raw paired-end reads. To fully explore the breed origin of BJB and accurately detect genomic footprints left by selection, whole genome resequencing data of another 31 samples of LD, MIN, LU and LW pigs (**Supplementary Table S1**) were downloaded from the public available database. After quality control, genomes of the above 90 pigs were aligned against the sus scrofa 11.1 reference genome using the BWA v0.7.17, resulting an average depth of 18.19 folds (**Supplementary Table S1**).

After variants calling and subsequent stringent quality control, a total number of 15,157,735 SNVs were identified in the BJB population with high quality (**Figure 1A**), of which 2,458,021 SNVs (16%) were considered as novel based on their absence in the pig dbSNP database (**Figure 1B**). Then, all detected SNVs in BJB were annotated using the gene annotation file downloaded from the Ensembl database. As it was expected, the largest number of SNPs was found in intergenic regions (46.42%) and introns (43.07%). Only 0.70% of them were located in exonic regions, including 61,674 synonymous and 42,859 non-synonymous mutations (**Figure 1C**, **Supplementary Table S2**). These potential functional SNPs provide valuable genetic resources for exploring the genetic structure and selective characteristics of BJB.

Phylogenetic Relationships and Population Structure Analysis

To assess the phylogenetic relationship among the nine pig breeds, an identical-by-state (IBS)-derived NJ tree were conducted for all tested individuals (**Figure 2A**). The tree showed that individuals of the same breed generally were clustered together, which signified that they possessed unique breed identities. The NJ tree revealed a clear divergence between modern commercial breeds and Chinese indigenous breeds, as the three modern commercial breeds (DU, LD, LW) formed a separate cluster while all the Chinese local breeds defined a large

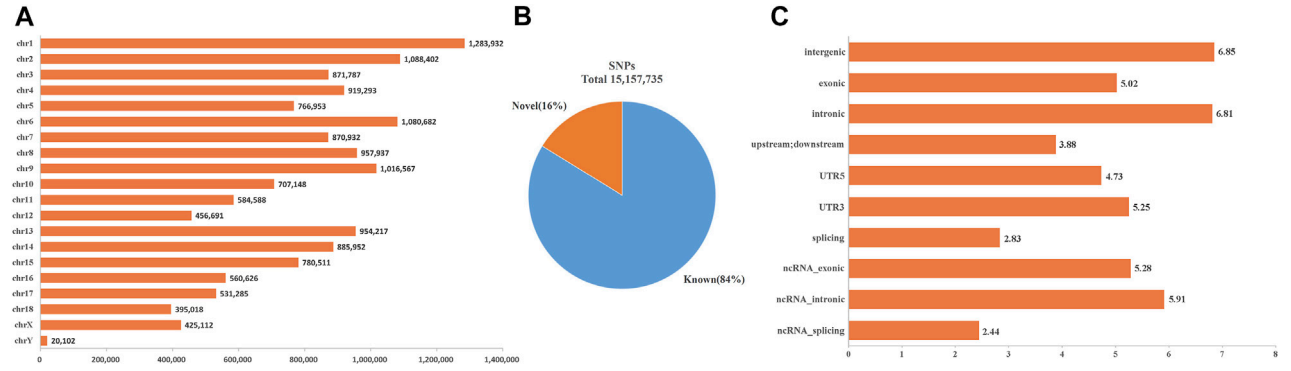

FIGURE 1 | SNV characteristics of Beijing Black population. **(A)** Genome-wide distribution of detected SNVs on 19 chromosomes for the Beijing Black pigs. *X*-axis represents the number of SNVs. *Y*-axis represents 18 autosomes and X chromosome. **(B)** The percentage of Beijing Black SNVs within the dbSNP database. **(C)** Genome-wide annotation of Beijing Black genetic variations. *X*-axis represents the number of genetic variations (log10) within various functional regions. *Y*-axis represents various functional regions.

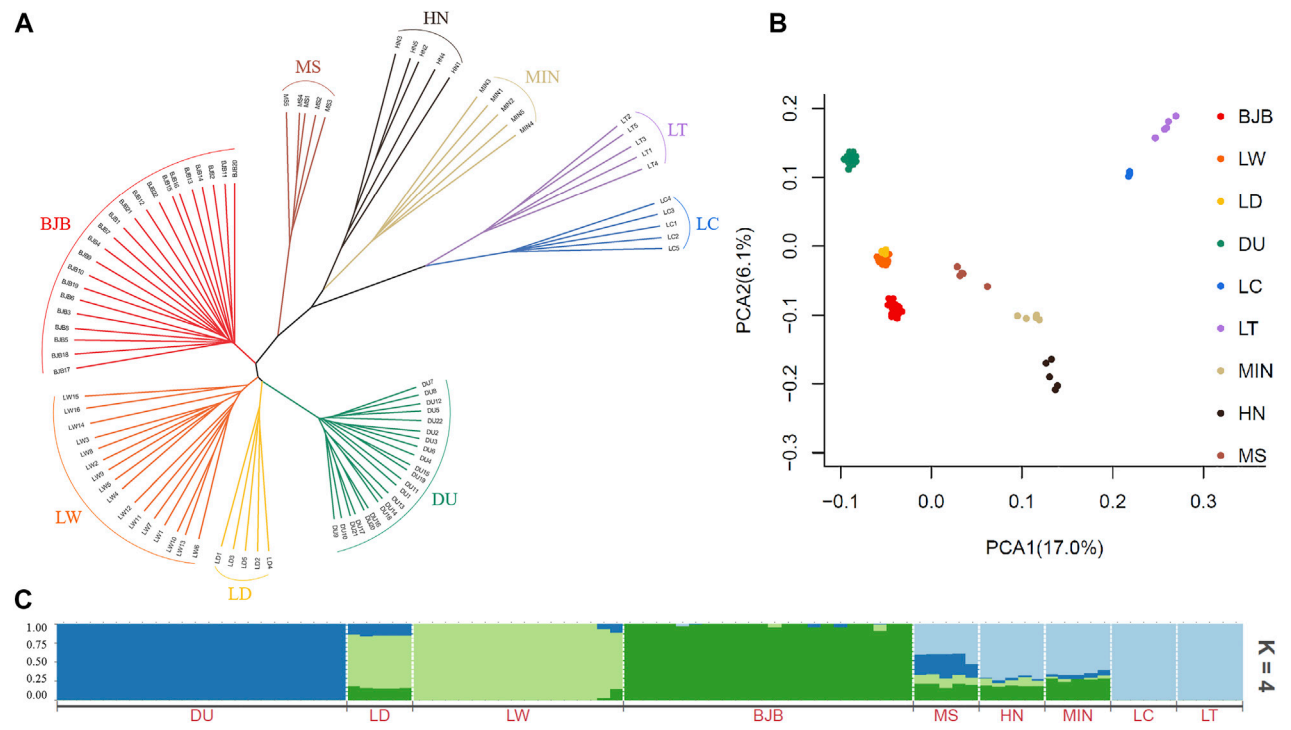

FIGURE 2 | Phylogenetic relationship and population structure of Beijing Black and other eight breeds tested in this study. BJB, Beijing Black pig; HN, Huainan pig; MS, Mashen pig; LT, Lantang pig; LC, Luchuan pig; MIN, Min pig; DU, Duroc; LD, Landrace; LW, Large White. **(A)** Neighbor-joining phylogenetic tree constructed from SNV data among nine populations. **(B)** Principle component analysis for the first two PCs of 90 pigs. **(C)** ADMIXTURE analysis with four presumed ancestral groups (K = 4).

new clade. Interestingly, we noted that BJB were located at intermediate positions between these two major clades, which was consistent with its breeding history.

The result of PCA analysis was consistent with the above NJ clustering pattern (**Figure 2B**). The first two PCs explained 17.0% and 6.1% of the total variation, respectively. It was obvious that BJB had closest genetic relationships to the commercial breeds including DU, LD and LW pigs, followed by Chinese indigenous pigs MS, MIN and HN, whereas LT and LC pigs were least related to BJB pigs. This may indicate that in the breeding history of BJB, more Western breeds such as Large White pigs were more heavily used in the crossbreeding of BJB than the Chinese indigenous pigs.

To investigate admixture levels among all tested breeds, we performed the ADMIXTURE analysis assuming ancestral

number K from 2 to 9 (**Figure 2C**, **Supplementary Figure S1**). According to our results, K = 4 represented the optimal number of assumed ancestors by cross-validation error test (**Supplementary Figure S1**). In this scenario, Beijing Black pigs and Western commercial breeds were differentiated, and a certain proportion of Western ancestries and Chinese native pig breeds were still evidenced in BJB. BJB has formed a unique genetic structure after multiple generations of breeding, which indicates that it can be used as an independent genetic resource.

Genetic Diversity of Beijing Black Pig

The genetic diversity between pairs of pig populations was investigated with the F_{ST} index (**Figure 3A**). The genetic differentiation between BJB and other lean commercial pig breeds (ranged from 0.08 to 0.16) were less than that between BJB and Chinese local pig breeds (ranged from 0.18 to 0.27), and the genetic differentiation between BJB and Large White was the lowest (0.08).

We calculated the inbreeding coefficients F_{ROH} based on ROH. In general, F_{ROH} of commercial pig breeds (ranging from 0.2696 to 0.3464) were higher than that of Chinese native pig breeds (ranging from 0.1825 to 0.2974). F_{ROH} of BJB (0.1906) was the fourth lowest among these nine breeds. The results showed that the existing breeding program could effectively avoid inbreeding of Beijing Black pigs to a certain extent. ROH fragments were then subdivided into seven categories (**Figure 3B**). We found that Beijing Black pigs had the highest number of ROHs, and ROHs of 100-200 Kb accounted for more than 60% of total ROHs in BJB. This may suggest that inbreeding events did not occur in recent generations in BJB or there may be a high proportion of inbreeding events in the first few generations of BJB.

Selection Signature Detection of the Beijing Black Pig Population

In order to explore the genomic evidence related to breed features of Beijing Black pigs, we further compared the genomic signatures of BJB with two typical commercial pig breeds at population level (i.e., 22 Duroc and 16 Large White pigs). Three complementary methods (i.e., genetic differentiation (F_{ST}); polymorphism level ($\theta\pi$); cross-population extended haplotype homozygosity (XP-EHH)) were used to investigate genome-wide selection signals. In order to reduce false positive candidate regions, regions meeting the top 5% threshold in at least two methods were selected as selected regions. The genome distributions of candidate regions detected by different methods of BJB to Duroc and Large White pigs were shown in **Figure 4**. There were 5,278 selected regions (threshold, 5%; F_{ST}, 0.508214; θπ ratio, 0.599414; XPEHH: 1.384220, **Supplementary Table S4–6**, **Supplementary Figure S2**) between BJB and Duroc population, whereas 4,887 selected regions (threshold, 5%; F_{ST}, 0.328,283; $\theta\pi$ ratio, 1.234,365; XPEHH: 1.610,046, **Supplementary Table S7–9**, **Supplementary Figure S2**) between BJB and Large White. Meanwhile, there were 1,167 candidate regions overlapped between the above two groups of selected regions. (**Figure 5A**, **Supplementary Table S10**).

Since highly differentiated SNVs across populations are more likely to occur in the vicinity of the selected regions (Carneiro et al., 2014), we further compared the alternate allele frequency of all identified SNVs in BJB with those in the other two populations. Then, the absolute allele frequency difference ΔAF = abs ($AltAF_{BJB}$-mean ($AltAF_{DU}$ + $AltAF_{LW}$)) were calculated to assess the potential selective sweeps of BJB. Significant enrichment of high-ΔAF SNVs (>0.8) within the identified sweep regions were observed, particularly the overlapping

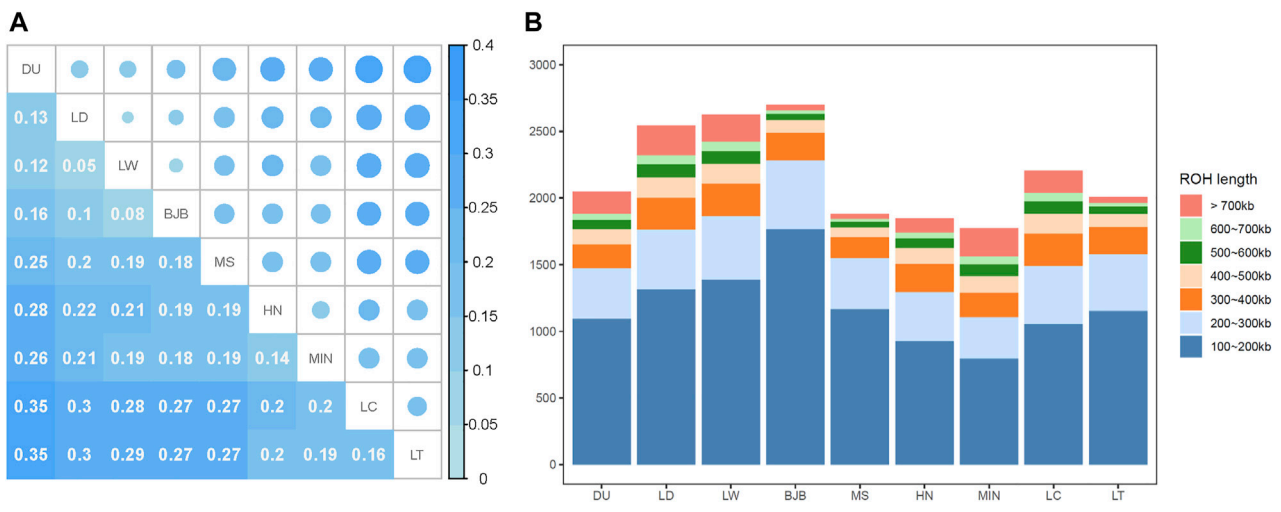

FIGURE 3 | Genetic diversity of Beijing Black pigs. **(A)** Heatmap of F_{ST} distance between breeds. **(B)** The average number of ROH fragments of each breed. The number of ROH was classified into seven categories.

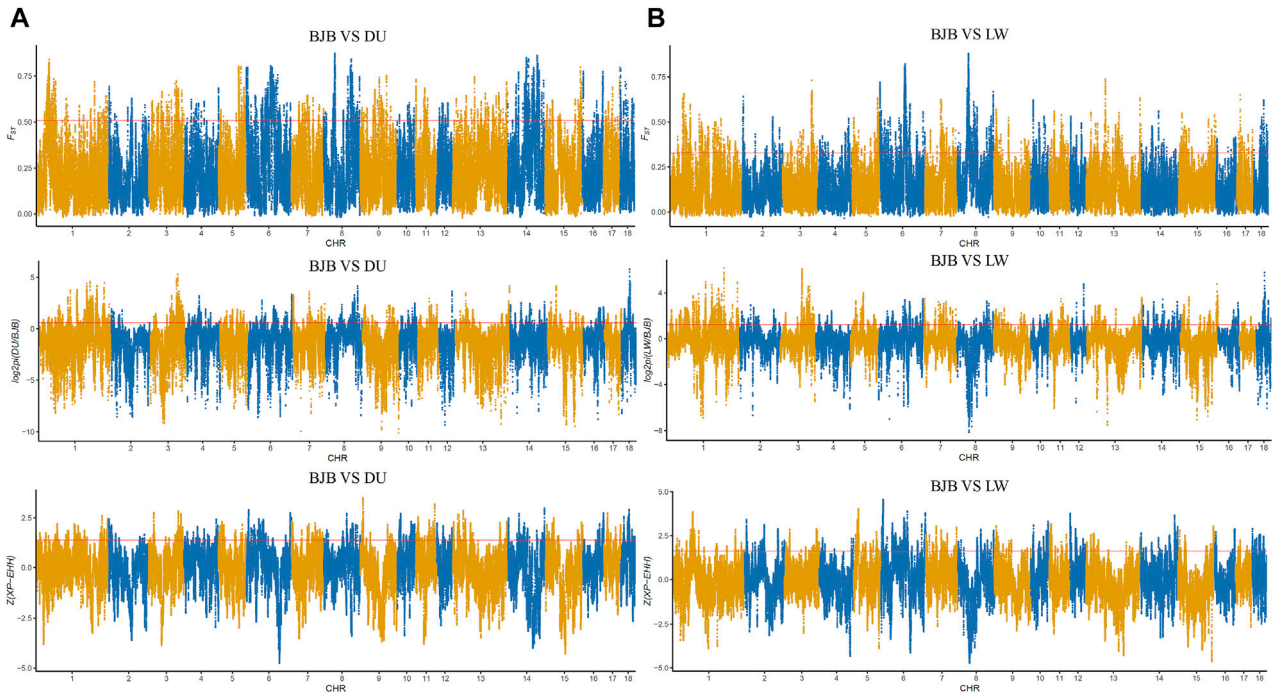

FIGURE 4 | Genome-wide distribution of selection signatures detected by F_{ST}, $\theta\pi$, and XP-EHH on 18 chromosomes from top to bottom. *X*-axis represents 18 autosomes, and *Y*-axis represents statistic values of each method. The $\theta\pi$ values are log2 normalized, and the XP-EHH values are standardized. Red line displays the threshold level of 5%. **(A)** Global distribution of statistic values of three methods between Beijing Black pigs and Duroc. **(B)** Global distribution of statistic values of three methods between Beijing Black pigs and Large White.

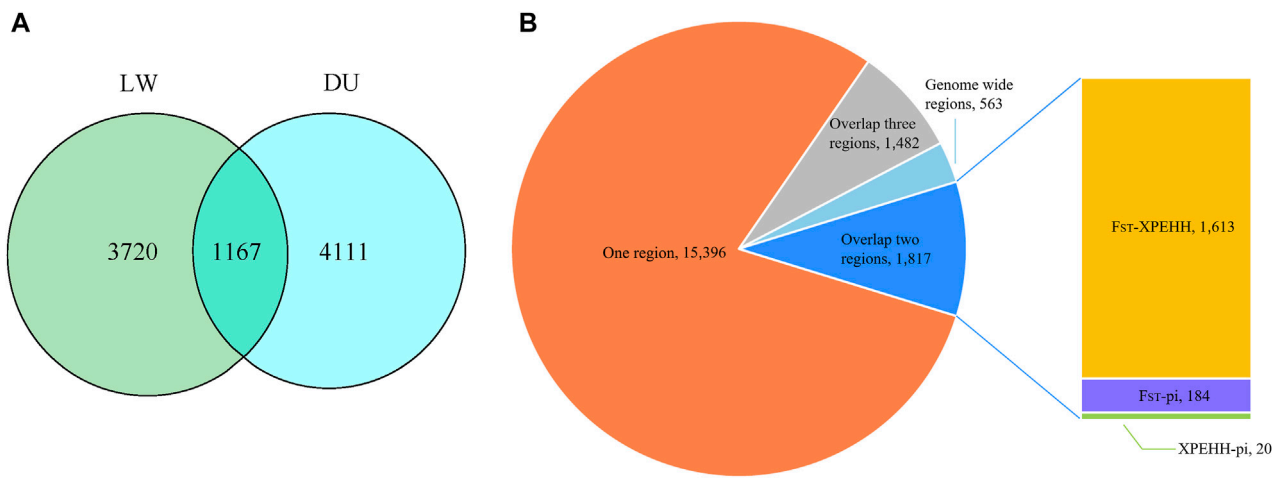

FIGURE 5 | (A) Venn diagram shows the overlap in the number of candidate regions detected by different populations. The left circle represents the numbers of candidate regions detected by more than two methods between Beijing Black pig and Large White. The right circle represents the numbers of candidate regions detected by more than two methods between Beijing Black pig and Duroc. **(B)** The number of SNPs with high ΔAF in each region between Beijing Black pig and other two breeds.

regions identified using both F_{ST} and XPEHH methods (**Figure 5B**). This reflected the fact that the highly differentiated population SNVs were actually associated with artificial and natural selection.

Gene Annotation and Functional Analysis

We annotated 392 genes in the 1,167 candidate regions (**Supplementary Table S11**). Considering the deficiency of functional annotation of pig genome, candidate genes were

transformed into human homologous genes using the Ensembl database. The gene annotation and pathways analysis showed that, genes were found to be significantly (FDR<0.05) enriched in 110 GO terms and 15 KEGG pathways (**Supplementary Table S12**). Some important pathways were found in the GO analysis, including lipid binding, glycerol metabolic process, adaptive thermogenesis, etc. In the KEGG analysis, most of significant pathways were related to disease (9 out of 15) and metabolic (2 out of 15), including papillomavirus infection, PI3K-Akt signaling pathway, and fat digestion and absorption, etc.

To identify publicly reported QTLs overlapped with these candidate regions, a total number of 34,342 QTLs of 708 different traits were downloaded from the Pig QTLdb database (Release 46, 27 Dec 2021). In total, there were 1,960 porcine QTLs (**Supplementary Table S13**) which were identified to be located within or overlapping with these 1,167 candidate regions. Notably, 1,287 (65.7%) QTLs were associated with meat and carcass traits, suggesting selection for meat traits during the breeding of BJB.

To further narrow down the candidate genes, we selected the 43 candidate regions that were detected in all three methods and overlapped in BJB compared with both Large White and Duroc (**Supplementary Table S14**). Twenty-five functionally important genes relevant to the excellent phenotype of BJB were identified, such as meat quality (*GPHA2, EHD1*, *HNF1A*, *C12orf43*, *GLTP*, *TRPV4*, *MVK*, and *MMAB*), reproduction (*PPP2R5B*), and disease resistance (*OASL*, *ANKRD13A*, *GIT2*). These genes were located on chromosomes 2, 6, and 14 respectively (**Supplementary Table S15**).

Through comparison of gene frequencies between BJB and the other two representative breeds, a total of 22 candidate BJB-specific SNVs were identified in exonic regions with the criteria of $AF_{BJB} > 95\%$ and $AF_{non\text{-}BJB} < 5\%$, including seven nonsynonymous SNVs found in three genes (*ENSSSCG00000008892*, *ENSSSCG00000036417*, and *MAP9*, **Supplementary Table S16**). The three genes are located in the 42–43 Mb region of chromosome eight and overlapped with the F_{ST} candidate region of BJB compared to Large White and Duroc pigs. Therefore, it is speculated that this region of BJB is a selection region compared with the other two breeds.

DISCUSSION

In this study, the whole genome resequencing of 22 Beijing Black pigs, 25 Chinese local pigs and 43 commercial pigs representing nine breeds were performed. Phylogenetic analysis, principal component analysis and population structure analysis showed that resequencing data could effectively distinguish BJB from commercial pig breeds and Chinese local pig breeds. Long-term and intensive artificial selection resulted in great differences in genome and breed specificity of Beijing Black pigs. Population genetic differentiation (F_{ST}) of Beijing Black pigs and other populations ranged from 0.10 to 0.27, which showed that Beijing Black pigs were more genetically similar to the commercial pig breeds than Chinese local pigs. However, Beijing Black pig still retains a small amount of genetic components of Huainan pig and Min pig (**Supplementary Figure S1**), which is consistent with the breeding history of Beijing Black pig.

Studies have demonstrated that inbreeding coefficient estimated by F_{ROH} is more accurate than that estimated by pedigree (Purfield et al., 2012), so we calculated inbreeding coefficient F_{ROH} in different pig populations. The results showed that the inbreeding coefficient of different pig populations ranged from 0.1825 to 0.3464, among which the inbreeding coefficient of BJB was 0.1906, which was relatively low among all tested populations. This indicates that although the core population size of BJB is small, the existing breeding programs effectively avoid inbreeding to a certain extent. The length of the homozygous fragment depends on the generational distance between the two individuals to a common ancestor. The shorter the homozygous fragment is, the farther the common ancestor is (Purfield et al., 2012). We found that the length of ROH fragments of Beijing Black pigs mainly concentrated in 100 KB ~ 200 KB. It is speculated that there was a high proportion of inbreeding behavior in the early generations of breed formation (Ceballos et al., 2018).

Beijing Black pigs have characteristics of excellent meat quality, early puberty, and great disease resistance due to the intensive artificial selection for many years. Therefore, there must be selection signatures on the Beijing Black genome. Windows with simultaneously high F_{ST} values, significantly high $\theta\pi$ ratios and high XPEHH values (5% right tail) were selected among populations. In total, there were 1,167 selected regions detected by the across-breeds comparisons, and 382 candidate genes were further identified within these regions. Functional enrichment analyses revealed that these selected genes may play an important role in meat quality, reproduction, and immune process.

We detected a list of genes putatively under selection that are functionally related to BJB breed features, such as *GPHA2*, *EHD1*, *HNF1A*, *MVK*, and *MMAB* for meat quality. *GPHA2* is a cystine forming polypeptide and a subunit of the dimer glycoprotein hormone family. The region near *GPHA2* was significantly associated with the meat tenderness in the genome-wide association analysis of Australian beef cattle (Bolormaa et al., 2011). *EHD1* regulates the recycling of various receptors from the endocytic recycling compartment to the plasma membrane (Kieken et al., 2007). GWAS in pigs found *EHD1* was significantly associated meat to fat ratio (MFR) trait (Falker-Gieske et al., 2019), and *EHD1* knockout mice demonstrated that *EHD1* regulates cholesterol homeostasis and lipid droplet storage (Naslavsky et al., 2007). *HNF1A* encodes the protein which is a transcription factor required to express several liver-specific genes and plays an important role in glucose activation, insulin secretion regulation, and lipid metabolism (Pearson et al., 2007). Studies have found that polymorphism of *HNF1A* is significantly correlated with psoas muscle area, backfat thickness, fat content and muscle glycogen metabolism in Yorkshire and Berkshire pigs (Murphy et al., 2007; Fan et al., 2010). *HNF1A* is linked to several Quantitative Trait Loci (QTL) regions related to meat quality on chromosome 14, and is considered as an important candidate gene for meat quality traits (Uemoto et al., 2012; Kayan et al., 2013). Besides, there is a non-synonymous mutation in the *HNF1A* (**Supplementary Table S17**).

MVK gene regulates cholesterol biosynthesis and terpenoid skeleton biosynthesis through SREBP. It was found that the expression of this gene is related to backfat thickness in pig in previous analyses (Kumar et al., 2019). Interestingly, five non-synonymous mutations were found within the *MVK* (**Supplementary Table S17**). *MVK* and adjacent *MMAB* gene is enriched in the same metabolic pathway. A study in mice reported that *MMAB* and *MVK* share a conserved promoter region and are both affected by sterol regulatory factor binding protein 2, indicating the two genes may share the same function.

Several candidate genes relating to reproduction were also detected, including *PPP2R5B* and *MAP9*. *PPP2R5B*, a regulatory subunit of PP2A is responsible for the dephosphorylation and inactivation of Akt protein (Beg et al., 2016). Studies have shown that *PPP2R5B* played a significant role in maintaining the fertility of boars at high temperatures (Hu et al., 2019), and this gene was expressed stably throughout lactation in sows (Tramontana et al., 2008). *MAP9* involved in mitotic spindle formation (Saffin et al., 2005). *MAP9* is a key factor in the early stage of development in zebra fish studies. Knockdown or overexpression of *MAP9* gene in zebra fish would lead to defects in early embryo development (Venoux et al., 2008).

The function of *ANKRD13A* was speculated to be immune-related. It regulated the K63 ubiquitination form of epidermal growth factor receptor (Tanno et al., 2012) and the endocytosis of B cell antigen receptor by interacting with endocytosis in humans (Satpathy et al., 2015). *OASL*, known inducers of antiviral activity, was found up-regulated of pigs infected with Classical swine fever virus (CSFV). Functional annotations showed that two non-synonymous mutations occurred within *OASL* (**Supplementary Table S17**). Therefore, our data might provide insight into the role of candidate genes in the immunity of BJB.

Overall, we comprehensively evaluated the genetic relationship and genetic diversity of Beijing Black pigs with commercial pig breeds and Chinese local pigs, providing new insights into the historical contribution of Western and Chinese ancestry to Beijing Black pigs. These findings enable us to propose a reliable and sustainable strategy for the conservation and improvement of Beijing Black pigs.

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 Institutional Animal Care and Use Committee of China Agricultural University.

AUTHOR CONTRIBUTIONS

JL and LZ conceived and designed the research. HW, XL, HL, XJ, HY, and GS provided data. WY, ZL, QZ, HD, and JY analyzed the data. WY and ZL wrote the manuscript. JL and LZ provided substantial comments and revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded by Beijing Municipal Commission of Science and Technology (Nos. Z191100004019009 and Z211100004621005), China Agriculture Research System of MOF and MARA, and National Natural Science Foundations of China (31661143013).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.860669/full#supplementary-material

REFERENCES

Beg, M., Srivastava, A., Shankar, K., Varshney, S., Rajan, S., Gupta, A., et al. (2016). PPP2R5B, a Regulatory Subunit of PP2A, Contributes to Adipocyte Insulin Resistance. *Mol. Cell. Endocrinol.* 437, 97–107. doi:10.1016/j.mce.2016.08.016

Bolormaa, S., Neto, L. R. P., Zhang, Y. D., Bunch, R. J., Harrison, B. E., Goddard, M. E., et al. (2011). A Genome-wide Association Study of Meat and Carcass Traits in Australian Cattle1. *J. Anim. Sci.* 89, 2297–2309. doi:10.2527/jas.2010-3138

Carneiro, M., Rubin, C.-J., Di Palma, F., Albert, F. W., Alföldi, J., Barrio, A. M., et al. (2014). Rabbit Genome Analysis Reveals a Polygenic Basis for Phenotypic Change during Domestication. *Science* 345, 1074–1079. doi:10.1126/science.1253714

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, 220–234. doi:10.1038/nrg.2017.109

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Falker-Gieske, C., Blaj, I., Preuß, S., Bennewitz, J., Thaller, G., and Tetens, J. (2019). GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3: Genes, Genomes, Genet.* 9, 2823–2834. doi:10.1534/g3.119.400452

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, 1619–1627. doi:10.1017/S175173111000087X

Felsenstein, J. (1993). *PHYLIP (Phylogeny Inference Package)*.

Francis, R. M. (2017). Pophelper: an R Package and Web App to Analyse and Visualize Population Structure. *Mol. Ecol. Resour.* 17, 27–32. doi:10.1111/1755-0998.12509

Gouveia, J. J. d. S., Silva, M. V. G. B. d., Paiva, S. R., and Oliveira, S. M. P. d. (2014). Identification of Selection Signatures in Livestock Species. *Genet. Mol. Biol.* 37, 330–342. doi:10.1590/S1415-47572014000300004

Holsinger, K. E., and Weir, B. S. (2009). Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting FST. *Nat. Rev. Genet.* 10, 639–650. doi:10.1038/nrg2611

Hou, X., Wang, L., Zhao, F., Liu, X., Gao, H., Shi, L., et al. (2021). Genome-Wide Expression Profiling of mRNAs, lncRNAs and circRNAs in Skeletal Muscle of Two Different Pig Breeds. *Animals* 11, 3169. doi:10.3390/ani11113169

Hu, Y., Deng, J., Tian, K., Yang, W. R., Luo, N. J., Lian, Y., et al. (2019). MiR-8-3p Regulates Hyperthermia-induced Lactate Secretion by Targeting PPP2R5B in Boar Sertoli Cells. *Mol. Reprod. Dev.* 86, 1720–1730. doi:10.1002/mrd.23265

Hu, Z.-L., Park, C. A., and Reecy, J. M. (2022). Bringing the Animal QTLdb and CorrDB into the Future: Meeting New Challenges and Providing Updated Services. *Nucleic Acids Res.* 50, D956–D961. doi:10.1093/nar/gkab1116

Kayan, A., Uddin, M. J., Kocamis, H., Tesfaye, D., Looft, C., Tholen, E., et al. (2013). Association and Expression Analysis of Porcine HNF1A Gene Related to Meat and Carcass Quality Traits. *Meat Sci.* 94, 474–479. doi:10.1016/j.meatsci.2013.04.015

Kieken, F., Jović, M., Naslavsky, N., Caplan, S., and Sorgen, P. L. (2007). EH Domain of EHD1. *J. Biomol. NMR* 39, 323–329. doi:10.1007/s10858-007-9196-0

Kumar, H., Srikanth, K., Park, W., Lee, S.-H., Choi, B.-H., Kim, H., et al. (2019). Transcriptome Analysis to Identify Long Non Coding RNA (lncRNA) and Characterize Their Functional Role in Back Fat Tissue of Pig. *Gene* 703, 71–82. doi:10.1016/j.gene.2019.04.014

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

Ma, Y., Zhang, S., Zhang, K., Fang, C., Xie, S., Du, X., et al. (2018). Genomic Analysis to Identify Signatures of Artificial Selection and Loci Associated with Important Economic Traits in Duroc Pigs. *G3: Genes, Genomes, Genet.* 8, 3617–3625. doi:10.1534/g3.118.200665

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

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

Murphy, C., Murray, A. M., Meaney, S., and Gåfvels, M. (2007). Regulation by SREBP-2 Defines a Potential Link between Isoprenoid and Adenosylcobalamin Metabolism. *Biochem. biophysical Res. Commun.* 355, 359–364. doi:10.1016/j.bbrc.2007.01.155

Naslavsky, N., Rahajeng, J., Rapaport, D., Horowitz, M., and Caplan, S. (2007). EHD1 Regulates Cholesterol Homeostasis and Lipid Droplet Storage. *Biochem. biophysical Res. Commun.* 357, 792–799. doi:10.1016/j.bbrc.2007.04.022

Niu, N., Wang, H., Shi, G., Liu, X., Liu, H., Liu, Q., et al. (2021). Genome Scanning Reveals Novel Candidate Genes for Vertebral and Teat Number in the Beijing Black Pig. *Anim. Genet.* 52, 734–738. doi:10.1111/age.13111

Pearson, E. R., Boj, S. F., Steele, A. M., Barrett, T., Stals, K., Shield, J. P., et al. (2007). Macrosomia and Hyperinsulinaemic Hypoglycaemia in Patients with Heterozygous Mutations in the HNF4A Gene. *Plos Med.* 4, e118. doi:10.1371/journal.pmed.0040118

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, 1–11. doi:10.1186/1471-2156-13-70

Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong Signatures of Selection in the Domestic Pig Genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi:10.1073/pnas.1217149109

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

Saffin, J.-M., Venoux, M., Prigent, C., Espeut, J., Poulat, F., Giorgi, D., et al. (2005). ASAP, a Human Microtubule-Associated Protein Required for Bipolar Spindle Assembly and Cytokinesis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11302–11307. doi:10.1073/pnas.0500964102

Satpathy, S., Wagner, S. A., Beli, P., Gupta, R., Kristiansen, T. A., Malinova, D., et al. (2015). Systems-wide Analysis of BCR Signalosomes and Downstream Phosphorylation and Ubiquitylation. *Mol. Syst. Biol.* 11, 810. doi:10.15252/msb.20145880

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI Database of Genetic Variation. *Nucleic Acids Res.* 29, 308–311. doi:10.1093/nar/29.1.308

Smith, J. M., and Haigh, J. (1974). The Hitch-Hiking Effect of a Favourable Gene. *Genet. Res.* 23, 23–35. doi:10.1017/S0016672300014634

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: an Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31, 2824–2827. doi:10.1093/molbev/msu211

Tanno, H., Yamaguchi, T., Goto, E., Ishido, S., and Komada, M. (2012). The Ankrd 13 Family of UIM-Bearing Proteins Regulates EGF Receptor Endocytosis from the Plasma Membrane. *MBoC* 23, 1343–1353. doi:10.1091/mbc.e11-09-0817

Tramontana, S., Bionaz, M., Sharma, A., Graugnard, D. E., Cutler, E. A., Ajmone-Marsan, P., et al. (2008). Internal Controls for Quantitative Polymerase Chain Reaction of Swine Mammary Glands during Pregnancy and Lactation. *J. Dairy Sci.* 91, 3057–3066. doi:10.3168/jds.2008-1164

Uemoto, Y., Nakano, H., Kikuchi, T., Sato, S., Ishida, M., Shibata, T., et al. (2012). Fine Mapping of Porcine SSC14 QTL and SCD Gene Effects on Fatty Acid Composition and Melting point of Fat in a Duroc Purebred Population. *Anim. Genet.* 43, 225–228. doi:10.1111/j.1365-2052.2011.02236.x

Venoux, M., Delmouly, K., Milhavet, O., Vidal-Eychenié, S., Giorgi, D., and Rouquier, S. (2008). Gene Organization, Evolution and Expression of the Microtubule-Associated Protein ASAP (MAP9). *BMC genomics* 9, 1–22. doi:10.1186/1471-2164-9-406

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2017). Package 'corrplot. *Statistician* 56, e24.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a Web Server for Annotation and Identification of Enriched Pathways and Diseases. *Nucleic Acids Res.* 39, W316–W322. doi:10.1093/nar/gkr483

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88, 76–82. doi:10.1016/j.ajhg.2010.11.011

Zhang, J., Chai, J., Luo, Z., He, H., Chen, L., Liu, X., et al. (2018). Meat and Nutritional Quality Comparison of Purebred and Crossbred Pigs. *Anim. Sci. J.* 89, 202–210. doi:10.1111/asj.12878

Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., et al. (2018). Evidence of Evolutionary History and Selective Sweeps in the Genome of Meishan Pig Reveals its Genetic and Phenotypic Characterization. *Gigascience* 7, giy058. doi:10.1093/gigascience/giy058

Conflict of Interest: HW, XL, HL, XJ, HY, and GS were employed by Beijing Heiliu Stockbreeding Technology Co.,Ltd.

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 March 2022
doi: 10.3389/fgene.2022.833475

Whole-Genome Sequencing of Endangered Dengchuan Cattle Reveals Its Genomic Diversity and Selection Signatures

Liangliang Jin[1,2], Kaixing Qu[3], Quratulain Hanif[4], Jicai Zhang[1], Jianyong Liu[1], Ningbo Chen[2], Quji Suolang[5], Chuzhao Lei[2] and Bizhi Huang[1]**

[1]Yunnan Academy of Grassland and Animal Science, Kunming, China, [2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, [3]Academy of Science and Technology, Chuxiong Normal University, Chuxiong, China, [4]National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, [5]Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Science, Lhasa, China

OPEN ACCESS

Edited by:
Tiago do Prado Paim, Goiano Federal Institute (IFGOIANO), Brazil

Reviewed by:
Ranjit Singh Kataria, National Bureau of Animal Genetic Resources (NBAGR), India
Qianjun Zhao, Institute of Animal Sciences (CAAS), China

***Correspondence:**
Chuzhao Lei
leichuzhao1118@126.com
Bizhi Huang
hbz@ynbp.cn

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

Received: 11 December 2021
Accepted: 04 March 2022
Published: 29 March 2022

Citation:
Jin L, Qu K, Hanif Q, Zhang J, Liu J, Chen N, Suolang Q, Lei C and Huang B (2022) Whole-Genome Sequencing of Endangered Dengchuan Cattle Reveals Its Genomic Diversity and Selection Signatures. Front. Genet. 13:833475. doi: 10.3389/fgene.2022.833475

Dengchuan cattle are the only dairy yellow cattle and endangered cattle among Yunnan native cattle breeds. However, its genetic background remains unclear. Here, we performed whole-genome sequencing of ten Dengchuan cattle. Integrating our data with the publicly available data, Dengchuan cattle were observed to be highly interbred than other cattle in the dataset. Furthermore, the positive selective signals were mainly manifested in candidate genes and pathways related to milk production, disease resistance, growth and development, and heat tolerance. Notably, five genes (*KRT39*, *PGR*, *KRT40*, *ESR2,* and *PRKACB*) were significantly enriched in the estrogen signaling pathway. Moreover, the missense mutation in the *PGR* gene (c.190T > C, p.Ser64Pro) showed a homozygous mutation pattern with higher frequency (83.3%) in Dengchuan cattle. In addition, a large number of strong candidate regions matched genes and QTLs related to milk yield and composition. Our research provides a theoretical basis for analyzing the genetic mechanism underlying Dengchuan cattle with excellent lactation and adaptability, crude feed tolerance, good immune performance, and small body size and also laid a foundation for genetic breeding research of Dengchuan cattle in the future.

Keywords: whole-genome resequencing, Chinese cattle, genetic diversity, population structure, selection signatures, lactation function

INTRODUCTION

Chinese domestic cattle breeds have a broad genetic base and abundant genetic variation, generally consisting of *Bos taurus* and *Bos indicus* lineages. *Bos taurus* mainly originated from cattle in Europe and is distributed in northern China, whereas *Bos indicus* originated from cattle in South Asia and is mainly distributed in southern China (Zhang, 2011). Because of the intensive selection, *Bos taurus* have advantages on beef and milk production; however, it is not adapted to tropical environments and thus cannot make use of its full potential for production in hot and humid areas in southern China (Crouse et al., 1989; Buchanan, 2002; Pegorer et al., 2007; Satrapa et al., 2011). Compared to *Bos taurus*, *Bos indicus* is able to tolerate heat and crude feed. One of the typical and distinctive physical features includes a hump on its back (Utsunomiya et al., 2019; Zhang et al., 2021). Scientific

research indicates hybridization of these sub-species can combine the strengths of *Bos taurus* and *Bos indicus*, being one of the oldest and truest ways to balance productivity with the environmental adaptability (Negussie et al., 1999; Schatz et al., 2020).

Yunnan has been one of the core regions for the migration of Indian indicine into the Chinese territory. Amongst various hybrid cattle, Dengchuan cattle are the only local dairy yellow cattle breed in China. Dengchuan shares a long history of selective breeding, dating back to the Han Dynasty (206 BC-220). Local people paid attention to selecting cattle with longer lactation periods with high milk yield, to breed their offspring, and using them to make milk fats, which is the local specialty dairy product (Zhang, 2011). With the improvement and promotion of artificial insemination and frozen semen technology of fresh semen at room temperature, native people massively introduced Holstein cattle for hybrid improvement. According to statistics data, from 1981 to 1989, the 305-day milk yield of Dengchuan cattle increased from 838.3 kg to 1,066.6 kg, with a milk fat rate of 6.89% and a dry milk matter of 13.52%, whereas the 305-day milk yield of hybrid F3 (Dengchuan × Holstein) increased from 2,111.1 kg to 3,094.4 kg, with a milk fat percentage of 4.09% and a dry milk matter of 12.0% (Zhang and Ma, 1987; Ma, 1988). However, blind hybridization and the lack of breed conservation planning caused the threat of breed degradation in Dengchuan cattle. According to a recent survey, only 212 Dengchuan cattle (206 cows and six bulls) remained, among which the original breed of Dengchuan cattle is extremely rare and endangered (Yang et al., 2021).

With the development of the next-generation sequencing technology and the enrichment of re-sequencing databases, genome-wide genetic analysis plays an increasingly significant role in the investigation and selection of germplasm resources of landraces (Shen et al., 2020; Jiang et al., 2021; Xia et al., 2021; Zhang et al., 2021). A recent study on Dengchuan cattle showed that Dengchuan cattle are a taurine–indicine mixed breed (Foissac et al., 2019). However, there are no previous studies using whole-genome sequencing data to identify genes related to milk production and disease resistance in Dengchuan cattle.

In this study, we performed whole-genome sequencing of ten individuals of Dengchuan cattle to explore the genetic diversity and population genetic structure of the autosomal genome. In order to further explore the genetic potential of Dengchuan cattle, single nucleotide polymorphisms (SNPs) of Dengchuan cattle were compared with those of commercial and native breeds previously collected from around the world.

MATERIALS AND METHODS

Sample Collection and Sequencing

Ten samples of Dengchuan cattle were collected from the ear tissue samples in the Dengchuan area of Yunnan province, China. To explore the ancestry proportions of Dengchuan cattle and compare the genetic diversity with worldwide cattle breeds, additional 68 samples were collected from the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra/) (Leinonen et al., 2011), including European cattle breeds [Angus (n = 9), Simmental (n = 8), and Holstein (n = 8)]; northeast Asia breed (Hanwoo, n = 10); southwest Chinese breeds [Dengchuan (n = 2), Dianzhong (n = 6), and Wenshan (n = 6)]; southeast Chinese breeds [Guangfeng (n = 4) and Wannan (n = 5)]; and India–Pakistan zebu cattle (*Bos indicus*) breeds (n = 10) (**Supplementary Table S1**). A total of 78 samples were used in this study.

Sequencing, Alignment, and Variant Identification

Genomic DNA was extracted using the standard phenol–chloroform method (Sterky et al., 2017). Paired-end libraries with the average insert size of 500 bp were constructed for each individual, with an average read length of 150 bp and an average sequence coverage of ~10.7×. Sequencing was performed using Illumina NovaSeq instruments at Novogene Bioinformatics Institute, Beijing, China. Raw reads data of fastq format were quality trimmed using trimmomatic (SLIDINGWINDOW:3:15 MINLEN:35 TRAILING:20 LEADING:20 AVGQUAL:20 TOPHRED33) (Bolger et al., 2014) to remove adapters and low-quality bases. The Burrows–Wheeler Aligner BWA-MEM (v0.7.15-r1140) with default parameters (Li and Durbin, 2009) was used to align the clean reads to the *Bos taurus* reference assembly ARS-UCD1.2. The Picard tools (http://broadinstitute.github.io/picard) were used to filter potential duplicate reads. We used "Haplotype Caller," "Genotype GVCFs," and "Select Variants" modules of the Genome Analysis Toolkit (GATK, version 3.8-1-0-gf15c1c3ef) (McKenna et al., 2010) to call the SNP. The filtration of raw SNPs was conducted by using "variant Filtration" modules with the parameters "QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, and SOR > 3.0" and the mean sequencing depth of variants (all individuals) "<1/3× and >3×". Based on the *Bos taurus* reference assembly ARS-UCD1.2, SNPs were functionally annotated by ANNOVAR (Wang et al., 2010).

Population Genomic Analysis

SNPs of 78 samples were pruned in high levels of pairwise LD by PLINK v1.90b3.40 software (Purcell et al., 2007), excluding SNPs in strong LD (r2 > 0.2) within a sliding window of 50 SNPs advanced by five SNPs at the time. Principal component analysis (PCA) was carried out using the smartpca program of the EIGENSOFT v5.0 package (Patterson et al., 2006). Population structure analysis was carried out by ADMIXTURE v1.3.0 (Alexander and Lange, 2011). Based on the pairwise distance matrix, the NJ tree was constructed by MEGA v10.2.6 (Saitou and Nei, 1987; Kumar et al., 2018).

Runs of homozygosity (ROHs) were calculated by PLINK software (Purcell et al., 2007). SNPs with minor allele frequencies (MAF) < 0.05 were excluded due to instability. PLINK uses a sliding window of a minimum of 50 SNPs across the genome to identify ROHs, allowing for two missing SNPs and one heterozygous site per window. The minimum number of continuous homozygous SNPs constituting an ROH was set to 100. The minimum SNP density coverage was set to at least 50 SNPs per Kb, allowing for centromeric and SNP-poor regions to be algorithmically excluded from the analysis. The maximum gap

between two consecutive homozygous SNPs was set at 100 Kb. The number and length of ROHs for each breed were estimated, and the length of ROH was divided into three categories: 0.5–1 Mb, 1–2 Mb, and 2–4 Mb, reflecting ancient, historical, and recent inbreeding, respectively (Kirin et al., 2010; Bhati et al., 2020).

Nucleotide diversity of each breed was investigated by VCFtools (Danecek et al., 2011) with the size of 50-kb non-overlapping window. The output of the --het function by VCFtools is a summary for each individual of the observed number of homozygous sites (O(hom)) and the expected number of homozygous sites (E(hom)). It also includes the total number of sites that the individual has data for and the inbreeding coefficient F, which is the canonical estimate of genomic F based on excess SNP homozygosity (Keller et al., 2011).

Linkage disequilibrium (LD) decay with the physical distance between SNPs was calculated and visualized by PopLDdecay software (Zhang et al., 2019) with default parameters.

Selective Sweep Identification

Only SNPs with less than 10% missing were used for selective sweep scanning. The nucleotide diversity ($\theta\pi$) and the composite likelihood ratio (CLR) test (Nielsen et al., 2005) were used to detect the selection signatures in Dengchuan cattle and Holstein cattle. The $\theta\pi$ was estimated based on a sliding window of size 50 kb and a step of size 20 kb by VCFtools (Danecek et al., 2011). The CLR was calculated for sites in non-overlapping 50-kb windows by SweepFinder2 (DeGiorgio et al., 2016), reflecting the likelihood of observing SNP data under the assumption of a sweep.

We also performed the genetic differentiation (F_{ST}) and cross-population composite likelihood ratio test (XP-CLR) (Chen et al., 2010) to identify the difference in potential areas between different cattle breeds. F_{ST} analysis was estimated based on a sliding window of 50 kb and a step of size 20 kb by VCFtools (Danecek et al., 2011). XP-CLR is a likelihood method for detecting selective sweeps by jointly modeling the multilocus allele frequency differentiation between the two groups (Chen et al., 2010). The overlap of the top 1% window in each method was considered as candidate signatures of selection, and genes in those window regions were defined as potential candidate genes.

Enrichment Analyses of Candidate Genes Under Selection

Due to the complexity of biological data-mining situations, enrichment analysis was conducted to identify the possibility of biological processes associated with Dengchuan cattle (Huang et al., 2009). Online Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analyses were conducted by KOBAS 3.0 (Bu et al., 2021). Genes at $p < 0.05$ were considered to be significantly enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and GO (Gene Ontology) annotations.

Aligning Candidate Regions to the Quantitative Trait Loci Database

Biological processes of Dengchuan cattle could be analyzed through genes annotated from candidate regions; however, there were various candidate regions in the non-annotated genic regions, although they showed strong selective signals. We used QTLs to identify possible traits in these candidate regions. The cattle QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) contains 163,725 QTLs. The chromosome information is annotated to the cattle QTLdb to identify the regions of interest detected by selective sweep methods contained or overlapped across the QTLs. The function and information of candidate regions were determined after annotation.

RESULTS

Analysis of the Population Structure and Genetic Diversity

Seventy-eight cattle, representing five geographically diverse cattle populations, namely, East Asian taurine, European taurine, Chinese indicine, Indian indicine (Xu et al., 2018), and local hybrid populations in Yunnan were selected for genome re-sequencing analysis (**Figure 1A**, **Supplementary Table S1**). After quality control, 147,397,064 bi-allelic autosomal SNPs (**Supplementary Table S2**) were used to construct genetic relationships using a neighbor-joining maximum likelihood method and PCA. Both methods revealed that these populations of regions, except Yunnan, clustered into three major genetic groups: *Bos taurus*, Indian indicine, and China indicine (**Figure 1B, C**). It was clear that Dengchuan cattle and Dianzhong cattle showed a certain degree of hybridization. Admixture analysis showed that the cattle breeds separate into *Bos taurus* and *Bos indicus* ancestries (**Figure 1D**, K = 2). When the number of clusters (K) was set to 4, East Asian taurine and European taurine were clearly separated. Dengchuan cattle depicted clear evidence of genetic heterogeneity with its shared genome ancestry with East Asian taurine (Hanwoo), European taurine (Angus, Simmental, and Holstein), Chinese indicine (Wannan and Guangfeng), and Indian indicine. It is rather remarkable that only half the Dengchuan cattle had a European taurine ancestry (**Figure 1D**, K = 4).

Patterns of Genomic Variation

ROH analysis revealed that the vast majority of ROHs identified in all breeds were between 0.5–1 Mb in length, but European commercial breeds (Angus, Holstein, and Simmental) had medium (1–2 Mb) and long ROHs (2–4 Mb). Besides, the total lengths of ROHs in Dengchuan cattle were much longer than those of the other two cattle in Yunnan (**Figure 2A**). This could indicate that European commercial breeds and Dengchuan cattle had undergone artificial selection for a long time. Similarly, the inbreeding coefficient based on genome heterozygosity was the highest in Angus (0.67) and lowest in Chinese indicine (−0.22) (**Figure 2B**). The average nucleotide diversity among Dengchuan cattle and other cattle groups revealed that Chinese indicine was the highest (3.10e−3), followed by Wenshan cattle (2.81e−3), Dianzhong cattle (2.72e−3), and Dengchuan cattle (2.58e−3). In comparison to *Bos indicus*, it was concluded that *Bos taurus* possessed a low level and high density of nucleotide diversity (**Figure 2C**). In contrast, the lowest average genome-wide LD was observed in Dengchuan cattle and Indian indicine. Besides, the

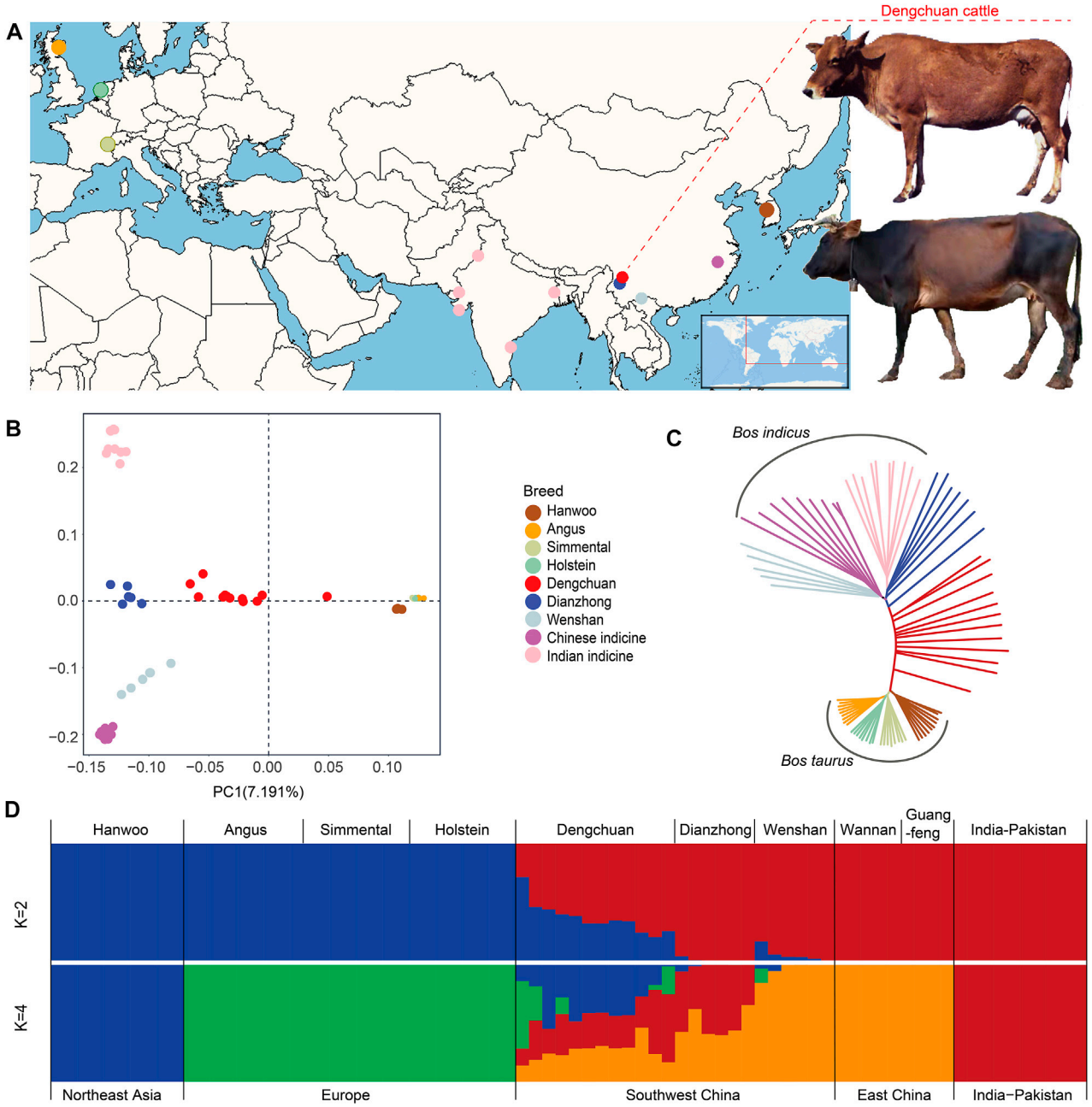

FIGURE 1 | Population genetic analysis of Dengchuan cattle. **(A)** Geographic map indicating the origins of Dengchuan cattle and other cattle analyzed in this study. **(B)** Principal component analysis of cattle with PC1 against PC2. **(C)** Neighbor-joining tree of the 78 domesticated cattle. **(D)** Admixture plot (K = 2, 4) for the 78 cattle individuals. Each individual is shown as a vertical bar divided into K colors.

LD decay in *Bos indicus* was faster than *Bos taurus* when the physical distance of SNP was less than 10 KB (**Figure 2D**).

Positive Selective Signature

A total of 217 candidate genes were detected by both θπ and the CLR test in Dengchuan cattle (**Figure 3A**, **Supplementary Tables S3, S4**). Some positively selected genes were reported to be associated with lactation function and disease resistance, such as the butterfat rate [*PPARGC1A* (Weikard et al., 2005; Schennink et al., 2009)], milk production [*B4GALT1* (Asadollahpour Nanaei et al., 2020; Valsalan et al., 2021)], immunity [*IL2* (McCoard et al., 2019) and *NFATC3* (Hu et al., 2018)], and mastitis resistance [*ITSN2* (Miles and Huson, 2020)]. In particular, *ITSN2* was located at the strongest selection signal on BTA11 (11:74700001-74900000). The result of strong positive selection was further verified by Tajima's D and nucleotide diversity analysis (**Figure 3C**). Moreover, 217 candidate genes were compared with 224 candidate genes detected by both θπ and the CLR test in Holstein cattle (**Supplementary Tables S5, S6**),

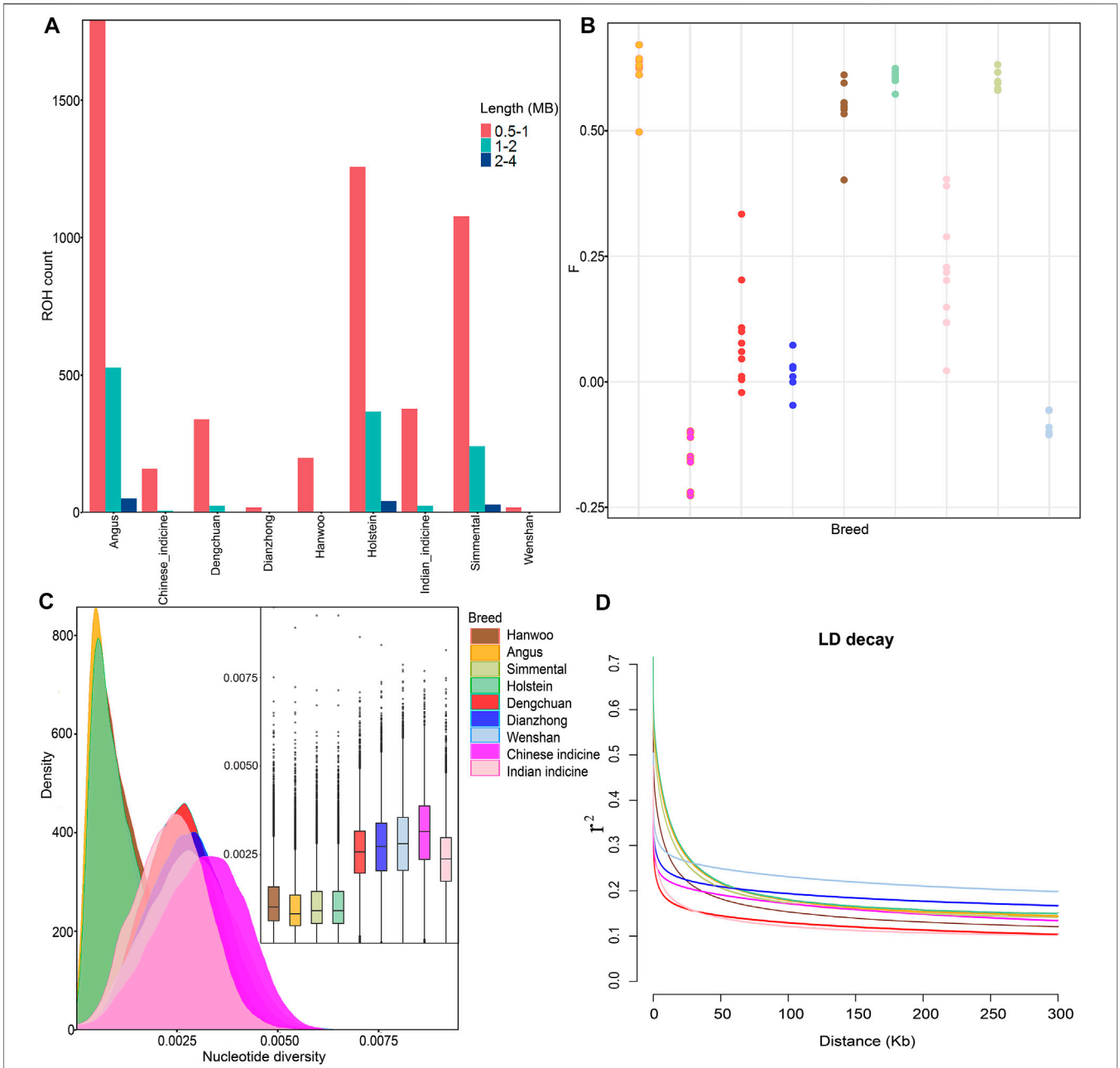

FIGURE 2 | Genetic diversity among 78 samples from nine populations. **(A)** Estimation of the total number of ROH for each group. The three categories of ROH length: 0.5–1 Mb, 1–2 Mb, and 2–4 Mb, reflecting ancient, historical, and recent inbreeding, respectively. **(B)** Inbreeding coefficient for each individual. **(C)** Density plots and Box plots of the nucleotide diversity for each group. **(D)** Genome-wide average LD decay estimated from each group.

whereas seven genes (*RERE*, *SLC45A1*, *RAB11FIP2*, *PTDSS1*, *MTERF3*, *KDM4C*, and *COL27A1*) were shared in both Dengchuan cattle and Holstein cattle.

Biological Process and Pathway Between Dengchuan Cattle and Angus Cattle

F_{ST} and XP-CLR tests were performed to detect the positive selection signatures between Dengchuan and Angus cattle (**Figure 3A**, **Supplementary Tables S7, S8**). A total of 384 genes were overlapped by both methods, which were enriched using GO annotation and KEGG pathway terms to further analyze their biological functions. The results represented significant enrichment of 220 GO terms and 29 KEGG pathways ($p < 0.05$; **Supplementary Tables S9, S10**). Gene list analysis revealed the involvement of various genes in protein synthesis (GO:0042802), endoplasmic reticulum (GO:0005783), temperature homeostasis (GO:0001659), and neutral amino acid

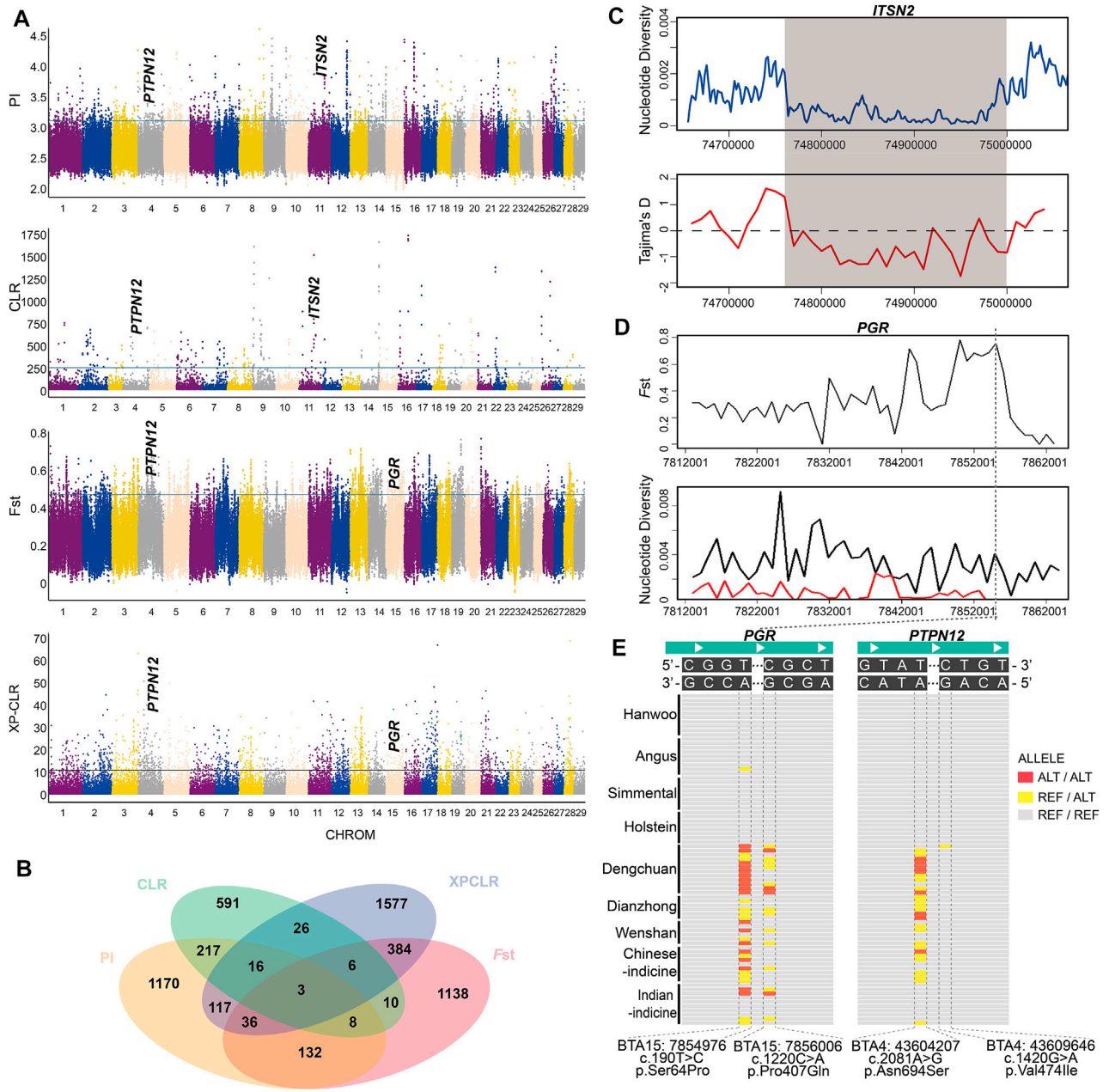

FIGURE 3 | Analysis of the signatures of positive selection in the genome of Dengchuan cattle. **(A)** Manhattan plot of selective sweeps in Dengchuan cattle. **(B)** Venn diagram showing the gene overlaps among θπ, CLR, *F*st, and XP-CLR. **(C)** Nucleotide diversity and Tajima's D plots at the *ITSN2* gene region. **(D)** *F*st and Tajima's D plots at the *PGR* gene region. **(E)** Regional highlight of the missense mutation of genes enriched in the estrogen signaling pathway. *PGR* gene showed a high-frequency homozygous mutation site (c.190T > C) in Dengchuan cattle.

transport (GO:0015804). Similarly, the enriched KEGG pathways included the estrogen signaling pathway, protein processing in the endoplasmic reticulum, biosynthesis of amino acids, glycan biosynthesis, and metabolic pathways. Moreover, *PGR*, a gene enriched in the estrogen signaling pathway, showed strong positive selection in Dengchuan cattle (**Figure 3D**).

It is worth noting that three overlapped genes (*PTPN12*, *KIAA1109*, and *ADAD1*) were detected among the four mentioned selection methods (**Figure 3B**), indicating that these genes were strongly selected in Dengchuan cattle. We checked mutations of eight genes (five estrogen signaling pathway enrichment genes (*KRT39*, *PGR*, *KRT40*, *ESR2*, and *PRKACB*) and three overlapped genes (*PTPN12*, *KIAA1109*, and *ADAD1*)) in Dengchuan cattle, two missense mutations in *PGR* (c.190T > C, p.Ser64Pro; c.1220C > A, p.Pro407Gln), and two missense mutations in *PTPN12* (c.2081A > G, p.Asn694Ser;

c.1420G > A, p.Val474Ile), which showed distinct allelic patterns in Dengchuan cattle (**Figure 3E**).

QTLs Based on Identified Regions

QTLs and selection signatures at the same location indicated that phenotypes and traits were influenced by the joint action of large numbers of polygenes and environmental effects (Georges et al., 1995). Thus, the top 50 candidate regions were studied for each method of scanning in Dengchuan cattle. These candidate regions also included regions that annotated gene failures and were used to extract relevant QTLs from the cattle QTLdb. Since the candidate region might overlap with several QTLs associated with different traits, one result with the most consistent chromosome fragment for each candidate region was picked. As shown in the **Supplementary Table S11**, 94 genomic regions for 100 candidate regions (θπ and CLR) overlapped with QTLs: 49 candidate regions overlapped milk, 21 candidate regions overlapped reproduction and production, 18 candidate regions overlapped health, and six candidate regions overlapped meat and carcass. Simultaneously, 99 genomic regions for 100 candidate regions (F_{ST} and XP-CLR) overlapped with QTLs: 50 candidate regions overlapped milk, 38 candidate regions overlapped reproduction and production, six candidate regions overlapped meat and carcass, four candidate regions overlapped health, and one candidate region overlapped exterior conformation (**Supplementary Table S12**).

DISCUSSION

Genomic information is the instruction of life construction. Here, we have conducted the whole-genome sequence-based study for the genomic diversity and selective signatures in Dengchuan cattle. The ancestral contributions of Dengchuan cattle came from East Asian taurine (~34%), Chinese indicine (~22%), European taurine (10%), and Indian indicine (~34%). It is worth noting that Dengchuan cattle have not been able to cluster completely, showing some differences amongst its individuals (**Figure 1B**). Similar outliers can be seen for the inbreeding coefficient (F) based on ROH, which may be from hybrid lineages or the introduction of crossbreeding. In addition, the ROH distribution and nucleotide diversity of Dengchuan cattle were basically consistent with other native Yunnan breeds (Foissac et al., 2019; Zhang et al., 2021). The LD decay pattern of Dengchuan cattle was similar to that of Indian indicine, confirming the high genetic diversity of Dengchuan cattle.

Dairy cows in hot and humid areas are naturally more prone to environmental mastitis due to bacterial growth (Alain et al., 2009). *ITSN2* is a member of a family of proteins involved in clathrin-mediated endocytosis that encodes a cytoplasmic protein which contains SH3 domains. *ITSN2* is thought to regulate the formation of clathrin-coated vesicles and may also function in the induction of T-cell antigen receptor (TCR) endocytosis (National Center for Biotechnology Information, 2017). Furthermore, *PTPN12* is a protein tyrosine phosphatase that contributes to the stable 3D acinar formation of mammary epithelial cells (Sun et al., 2011). *PGR* promotes alveologenesis in the pregnant mammary gland for milk production (Aikawa et al., 2020). In addition, among the seven genes overlapped between Dengchuan cattle and Holstein cattle, we examined the scanning signal of *PTDSS1* and found that the CLR value in Dengchuan cattle (~638) was much higher than that in Holstein cattle (~431). *PTDSS1* is used for the catalytic synthesis of lecithin. The remaining six genes were also reported to be highly associated with milk production: *RERE* and *SLC45A1* are reported to be associated with milk production (Buaban et al., 2022); *MTERF3* is associated with milk fatty acid composition (Palombo et al., 2018); *KDM4C* is associated with breast cancer (Garcia and Lizcano, 2016); *RAB11FIP2* is associated with transcytosis (Ducharme et al., 2007); and *COL27A1* is associated with the sternum (Maddirevula et al., 2019). The strong selection of these genes may be the reason why Dengchuan cattle are a good dairy breed in hot and humid climates.

The comparative analysis of genetic differentiation between Dengchuan cattle and other breeds revealed the estimated value of Dengchuan cattle and Angus cattle to be the highest (~0.22), which was suitable for subsequent analysis (**Supplementary Table S13**). Interestingly, the KEGG pathway with significant enrichment of differential signals included the estrogen signaling pathway. Activation of the estrogen signaling pathway results in prolonged lactation and high milk yield (Lawrence, 2022). Additionally, in the current study, the missense mutations in PGR (c.190T > C, p.Ser64Pro) (**Figure 3E**) likely play an important role in elevated milk production in Dengchuan cattle.

Milk yield and composition are typical polygenic traits (Georges et al., 1995). For a more comprehensive explanation of the function in strong candidate regions, the most promising QTLs were matched for two-hundred candidate regions. Because of the window size of scanning methods and the different fragment sizes of different genes and QTLs, some regions were separated and calculated several times and were matched to the same QTL. This results in a large number of milk-related QTLs matching our candidate regions. Overall, about half of the candidate regions matched QTLs associated with milk yield and its composition. Furthermore, the candidate regions corresponded with the health QTLs, which were mainly related to the somatic cell score and heat tolerance. Most of the QTLs related to reproduction were associated with the body size in our result. For example, some growth-related QTLs have been annotated on chromosome 14 (**Supplementary Table S11**). Similarly, the *PLAG1* gene on chromosome 14 has been shown to be associated with the body size (Karim et al., 2011). *F*st results showed that there exists significant genetic differentiation of the *PLAG1* gene between Dengchuan cattle and Angus cattle (**Supplementary Table S7**). It is worth noting that Dengchuan cattle are small in body stature, approximately 105 cm tall and weighing 225 kg in adulthood (Zhang, 2011).

In conclusion, this study provides a theoretical basis for analyzing the genetic mechanism of Dengchuan cattle with excellent lactation and adaptability, crude feed tolerance, good immune performance, and small body size, which also lays a foundation for genetic breeding research of Dengchuan cattle in the future.

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 study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University following the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China. Specific consent procedures were not required for this study following the recommendations of the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

AUTHOR CONTRIBUTIONS

NC and CL conceived and designed the experiments. LJ and KQ performed the experiments. JZ, JL, and QS contributed analysis tools. LJ wrote the manuscript. QH and CL revised the manuscript and provided suggestions. KQ, CL, and BH contributed to the funding for the research. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the China Agriculture Research System of MOF and MARA (CARS-37), the Program of Yunling Scholar, the Young and Middle-aged Academic Technology Leader Backup Talent Cultivation Program in Yunnan Province, China (No. 2018HB045), and Yunnan Provincial Major S&T Project (No. 2019ZG007), and the Doctoral Startup Project of Chuxiong Normal University (No. BSQD2101).

ACKNOWLEDGMENTS

Thanks to high-performance computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.833475/full#supplementary-material

REFERENCES

Aikawa, S., Yuan, J., Dewar, A., Sun, X., and Dey, S. K. (2020). Scribble Promotes Alveologenesis in the Pregnant Mammary Gland for Milk Production. *Reproduction* 159, 719–731. doi:10.1530/rep-20-0108

Alain, K., Karrow, N. A., Thibault, C., St-Pierre, J., Lessard, M., and Bissonnette, N. (2009). Osteopontin: an Early Innate Immune Marker of *Escherichia coli* Mastitis Harbors Genetic Polymorphisms with Possible Links with Resistance to Mastitis. *BMC Genomics* 10, 444. doi:10.1186/1471-2164-10-444

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinformatics* 12, 246. doi:10.1186/1471-2105-12-246

Asadollahpour Nanaei, H., Dehghani Qanatqestani, M., and Esmailizadeh, A. (2020). Whole-genome Resequencing Reveals Selection Signatures Associated with Milk Production Traits in African Kenana Dairy Zebu Cattle. *Genomics* 112, 880–885. doi:10.1016/j.ygeno.2019.06.002

Bhati, M., Kadri, N. K., Crysnanto, D., and Pausch, H. (2020). Assessing Genomic Diversity and Signatures of Selection in Original Braunvieh Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 21, 27. doi:10.1186/s12864-020-6446-y

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

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

Buaban, S., Lengnudum, K., Boonkum, W., and Phakdeedindan, P. (2022). Genome-wide Association Study on Milk Production and Somatic Cell Score for Thai Dairy Cattle Using Weighted Single-step Approach with Random Regression Test-Day Model. *J. Dairy Sci.* 105, 468–494. doi:10.3168/jds.2020-19826

Buchanan, D. S. (2002). "Animals that Produce Dairy Foods | Major Bos Taurus Breeds," in *Encyclopedia of Dairy Sciences*. Editor J. W. Fuquay. Second Edition (San Diego: Academic Press), 284–292. doi:10.1016/b978-0-12-374407-4.00032-7

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

Crouse, J. D., Cundiff, L. V., Koch, R. M., Koohmaraie, M., and Seideman, S. C. (1989). Comparisons of and Inheritance for Carcass Beef Characteristics and Meat Palatability. *J. Anim.* 67, 2661. doi:10.2527/jas1989.67102661x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., and Nielsen, R. (2016). SweepFinder2: Increased Sensitivity, Robustness and Flexibility. *Bioinformatics* 32, 1895–1897. doi:10.1093/bioinformatics/btw051

Ducharme, N. A., Williams, J. A., Oztan, A., Apodaca, G., Lapierre, L. A., and Goldenring, J. R. (2007). Rab11-FIP2 Regulates Differentiable Steps in Transcytosis. *Am. J. Physiol. Cel Physiol.* 293 (3), C1059–C1072. doi:10.1152/ajpcell.00078.2007

Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., et al. (2019). Multi-species Annotation of Transcriptome and Chromatin Structure in Domesticated Animals. *BMC Biol.* 17, 108. doi:10.1186/s12915-019-0726-5

Garcia, J., and Lizcano, F. (2016). KDM4C Activity Modulates Cell Proliferation and Chromosome Segregation in Triple-Negative Breast Cancer. *Breast Cancer (Auckl)* 10, 169–175. doi:10.4137/BCBCR.S40182

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., et al. (1995). Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. *Genetics* 139, 907–920. doi:10.1093/genetics/139.2.907

Hu, L., He, F., Huang, M., Peng, M., Zhou, Z., Liu, F., et al. (2018). NFATc3 Deficiency Reduces the Classical Activation of Adipose Tissue Macrophages. *J. Mol. Endocrinol.* 61, 79–89. doi:10.1530/jme-18-0070

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* 37, 1–13. doi:10.1093/nar/gkn923

Jiang, L., Kon, T., Chen, C., Ichikawa, R., Zheng, Q., Pei, L., et al. (2021). Whole-genome Sequencing of Endangered Zhoushan Cattle Suggests its Origin and the Association of MC1R with Black Coat Colour. *Sci. Rep.* 11, 17359. doi:10.1038/s41598-021-96896-2

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A. C., Baurain, D., et al. (2011). Variants Modulating the Expression of a Chromosome Domain Encompassing PLAG1 Influence Bovine Stature. *Nat. Genet.* 43, 405–413. doi:10.1038/ng.814

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

Kirin, M., Mcquillan, R., Franklin, C. S., Campbell, H., Mckeigue, P. M., and Wilson, J. F. (2010). Genomic Runs of Homozygosity Record Population History and Consanguinity. *PloS One* 5, e13996. doi:10.1371/journal.pone.0013996

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096

Lawrence, R. A. (2022). "3-Physiology of Lactation," in *Breastfeeding*. Editors R. A. Lawrence, and R. M. Lawrence. Ninth Edition (Philadelphia: Elsevier), 58–92. doi:10.1016/b978-0-323-68013-4.00003-1

Leinonen, R., Sugawara, H., and Shumway, M, International Nucleotide Sequence Database Collaboration (2011). The Sequence Read Archive. *Nucleic Acids Res.* 39, D19–D21. doi:10.1093/nar/gkq1019

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

Ma, Y. (1988). Crossing Improvement Effect of Holstein Cattle and Dengchuan Cattle (In Chinese). *Yunnan J. Anim. Sci. Vet. Med.* 03, 12–13.

Maddirevula, S., Alzahrani, F., Al-Owain, M., Al Muhaizea, M. A., Kayyali, H. R., Alhashem, A., et al. (2019). Autozygome and High Throughput Confirmation of Disease Genes Candidacy. *Genet. Med.* 21 (3), 736–742. doi:10.1038/s41436-018-0138-x

McCoard, S., Heiser, A., Lowe, K., Molenaar, A., Maclean, P., Johnstone, P., et al. (2019). Effect of Weaning Age on Growth, Mammary Gland Development, and Immune Function in Holstein Friesian Calves Fed Conserved Alfalfa (FiberStart). *J. Dairy Sci.* 102, 6076–6087. doi:10.3168/jds.2018-15615

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Miles, A. M., and Huson, H. J. (2020). Time- and Population-dependent Genetic Patterns Underlie Bovine Milk Somatic Cell Count. *J. Dairy Sci.* 103, 8292–8304. doi:10.3168/jds.2020-18322

National Center for Biotechnology Information (2017). Gene: ITSN2 Intersectin 2. Available at: https://www.ncbi.nlm.nih.gov/gene/50618 (Accessed November, 2021).

Negussie, E., Brannang, E., and Rottmann, O. J. (1999). Reproductive Performance and Herd Life of Dairy Cattle at Asella Livestock Farm, Arsi, Ethiopia. II. Crossbreds with 50, 75 and 87.5% European Inheritance. *J. Anim. Breed. Genet.* 116 (3), 225–234. doi:10.1046/j.1439-0388.1999.00191.x

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic Scans for Selective Sweeps Using SNP Data. *Genome Res.* 15, 1566–1575. doi:10.1101/gr.4252305

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., et al. (2018). Genome-wide Association Study of Milk Fatty Acid Composition in Italian Simmental and Italian Holstein Cows Using Single Nucleotide Polymorphism Arrays. *J. Dairy Sci.* 101, 11004–11019. doi:10.3168/jds.2018-14413

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *Plos Genet.* 2, e190. doi:10.1371/journal.pgen.0020190

Pegorer, M. F., Vasconcelos, J. L. M., Trinca, L. A., Hansen, P. J., and Barros, C. M. (2007). Influence of Sire and Sire Breed (Gyr versus Holstein) on Establishment of Pregnancy and Embryonic Loss in Lactating Holstein Cows during Summer Heat Stress. *Theriogenology* 67, 692–697. doi:10.1016/j.theriogenology.2006.09.042

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

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: a New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Satrapa, R. A., Nabhan, T., Silva, C. F., Simões, R. A. L., Razza, E. M., Puelker, R. Z., et al. (2011). Influence of Sire Breed (*Bos indicus* versus *Bos taurus*) and Interval from slaughter to Oocyte Aspiration on Heat Stress Tolerance of In Vitro-produced Bovine Embryos. *Theriogenology* 76, 1162–1167. doi:10.1016/j.theriogenology.2011.05.026

Schatz, T., Thomas, S., Reed, S., and Hearnden, M. (2020). Crossbreeding with a Tropically Adapted *Bos taurus* Breed (Senepol) to Improve Meat Quality and Production from Brahman Herds in Northern Australia. 1. Steer Performance. *Anim. Prod. Sci.* 60, 487–491. doi:10.1071/AN18609

Schennink, A., Bovenhuis, H., Léon-Kloosterziel, K. M., Van Arendonk, J. A. M., and Visker, M. H. P. W. (2009). Effect of Polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5 Agenes on Bovine Milk-Fat Composition. *Anim. Genet.* 40, 909–916. doi:10.1111/j.1365-2052.2009.01940.x

Shen, J., Hanif, Q., Cao, Y., Yu, Y., Lei, C., Zhang, G., et al. (2020). Whole Genome Scan and Selection Signatures for Climate Adaption in Yanbian Cattle. *Front. Genet.* 11, 94. doi:10.3389/fgene.2020.00094

Sterky, F. H., Trotter, J. H., Lee, S.-J., Recktenwald, C. V., Du, X., Zhou, B., et al. (2017). Carbonic Anhydrase-Related Protein CA10 Is an Evolutionarily Conserved Pan-Neurexin Ligand. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1253–E1262. doi:10.1073/pnas.1621321114

Sun, T., Aceto, N., Meerbrey, K. L., Kessler, J. D., Zhou, C., Migliaccio, I., et al. (2011). Activation of Multiple Proto-Oncogenic Tyrosine Kinases in Breast Cancer via Loss of the PTPN12 Phosphatase. *Cell* 144, 703–718. doi:10.1016/j.cell.2011.02.003

Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., et al. (2019). Genomic Clues of the Evolutionary History of *Bos indicus* Cattle. *Anim. Genet.* 50, 557–568. doi:10.1111/age.12836

Valsalan, J., Sadan, T., Venkatachalapathy, T., Anilkumar, K., and Aravindakshan, T. V. (2021). Identification of Novel Single-Nucleotide Polymorphism at Exon1 and 2 Region of B4GALT1 Gene and its Association with Milk Production Traits in Crossbred Cattle of Kerala, India. *Anim. Biotechnol.* 3, 1–9. doi:10.1080/10495398.2020.1866591

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Weikard, R., Kühn, C., Goldammer, T., Freyer, G., and Schwerin, M. (2005). The Bovine PPARGC1A Gene: Molecular Characterization and Association of an SNP with Variation of Milk Fat Synthesis. *Physiol. Genomics* 21, 1–13. doi:10.1152/physiolgenomics.00103.2004

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22, 43. doi:10.1186/s12864-020-07340-0

Xu, J., Dong, Q., Yu, Y., Niu, B., Ji, D., Li, M., et al. (2018). Mass Spider Silk Production through Targeted Gene Replacement in *Bombyx mori*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8757–8762. doi:10.1073/pnas.1806805115

Yang, L., Su, Z., Li, H., and Zhao, T. (2021). Investigation and Analysis of the Present Situation of Dengchuan Cattle. (In Chinese). *Heilongjiang Anim. Sci. Vet. Med.* 13, 58–61. doi:10.13881/j.cnki.hljxmsy.2020.01.0281

Zhang, G., and Ma, Y. (1987). Benefit of Progressive Crossing of Dungchuan Cattle (In Chinese). *Chin. J. Anim. Sci.* 02, 31–32.

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: a Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35, 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, X., Qu, K., Jia, P., Zhang, J., Liu, J., Lei, C., et al. (2021). Assessing Genomic Diversity and Productivity Signatures in Dianzhong Cattle by Whole-Genome Scanning. *Front. Genet.* 12. doi:10.3389/fgene.2021.719215

Zhang, Y. (2011). *Animal Genetic Resources in China Bovines*. (in Chinese). Beijing: China Agriculture Press.

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: 11 April 2022
doi: 10.3389/fgene.2022.846449

Genetic Basis of Dorper Sheep (*Ovis aries*) Revealed by Long-Read *De Novo* Genome Assembly

Guoyan Qiao[1,2†], Pan Xu[3†], Tingting Guo[1,2], Yi Wu[1,2], Xiaofang Lu[4,5], Qingfeng Zhang[4,5], Xue He[1], Shaohua Zhu[1,2], Hongchang Zhao[1,2], Zhihui Lei[1,2], Weibo Sun[1,2], Bohui Yang[1,2] and Yaojing Yue[1,2]**

[1]Lanzhou Institute of Husbandry and Pharmaceutical Sciences of Chinese Academy of Agricultural Sciences, Lanzhou, China, [2]State Key Laboratory of Grassland Agro-Ecosystems, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, [3]Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, [4]Tianjin Aoqun Animal Husbandry Pty., Ltd., Tianjin, China, [5]The Enterprises Key Laboratory of Tianjin Meat-Type-Sheep Genetics and Breeding, Tianjin, China

OPEN ACCESS

Edited by:
El Hamidi Hay, United States Department of Agriculture, United States

Reviewed by:
Guillermo Giovambattista, CONICET Institute of Veterinary Genetics (IGEVET), Argentina
Emily Louise Clark, University of Edinburgh, United Kingdom

***Correspondence:**
Yaojing Yue
yueyaojing@126.com
Bohui Yang
yangbh2004@163.com

†These authors have contributed equally to this work and share first authorship

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

Received: *31 December 2021*
Accepted: *25 March 2022*
Published: *11 April 2022*

Citation:
Qiao G, Xu P, Guo T, Wu Y, Lu X, Zhang Q, He X, Zhu S, Zhao H, Lei Z, Sun W, Yang B and Yue Y (2022) Genetic Basis of Dorper Sheep (Ovis aries) Revealed by Long-Read De Novo Genome Assembly. Front. Genet. 13:846449. doi: 10.3389/fgene.2022.846449

Dorper sheep (*Ovis aries*) (DPS), developed in the 1930s by crossing Dorset Horn and Blackhead Persian sheep in South Africa, is a world-famous composite breed for mutton production. The genetic basis underlying this breed is yet to be elucidated. Here, we report the sequencing and assembly of a highly contiguous Dorper sheep genome via integration of Oxford Nanopore Technology (ONT) sequencing and Hi-C (chromatin conformation capture) approaches. The assembled genome was around 2.64 Gb with a contig N50 of 73.33 Mb and 140 contigs in total. More than 99.5% of the assembled sequences could be anchored to 27 chromosomes and they were annotated with 20,450 protein-coding genes. Allele-specific expression (ASE) genes of Dorper sheep were revealed through ASE analysis and they were involved in the immune system, lipid metabolism, and environmental adaptation. A total of 5,701 and 456 allelic sites were observed in the SNP and indels loci identified from relevant whole-genome resequencing data. These allelic SNP and INDEL sites were annotated in 1,002 and 294 genes, respectively. Moreover, we calculated the number of variant sites and related genes derived from the maternal and paternal ancestors, revealing the genetic basis of outstanding phenotypic performance of Dorper sheep. In conclusion, this study reports the first reference genome of Dorper sheep and reveals its genetic basis through ASE. This study also provides a pipeline for mining genetic information of composite breeds, which has an implication for future hybrid-breeding practices.

Keywords: dorper sheep, reference genome, genetic basis, composite breed, allele-specific expression (ASE)

1 INTRODUCTION

The sheep (*Ovis aries*) was one of the first animals domesticated for agricultural purposes (Hiendleder, Kaupe, Wassmuth, & Janke, 2002). Currently, the total number of domesticated sheep in the world exceeds one billion. They are an important source of meat, wool, and dairy products, and play an important role in the global agricultural economy. Sheep are widely distributed in the cold zone, in the tropics, and at high elevations due to their rich phenotypic variations of breeds for different production

targets (Alberto et al., 2018). High quality genomes of sheep are the basis for systematically exploring their evolution and analyzing the unique biological traits, which is of great significance for the conservation and utilization of genetic resources and the mining of genetic characteristics (Roth, 2019).

The Dorper sheep is an easy-care, fast-growing, meat-producing hair sheep, that was developed and became the second largest breed in South Africa by crossing Dorset Horn with Blackhead Persian in the 1930s, then spread to many other countries throughout the world (Cloete, Snyman, & Herselman, 2000). Hair sheep are widely used to breed heat-tolerant lambs, which is particularly important given the current and future pressures on sheep breeding that are associated with climate change. Unlike wool sheep, Dorper sheep don't require shearing, crutching, or mulesing, and they are much less prone to flystrike (Pollott, 2011). They have high fertility and maternal instinct, combined with a high growth rate and hardiness. They reputably do well in various environmental and feeding conditions, particularly intensive feeding system (Tesema et al., 2020). However, only limited genomic information is available for this important breed. With the rapid development and lowering cost of sequencing technologies, the use of genomics in mining livestock genetic diversity is becoming more widespread (van Dijk, Jaszczyszyn, Naquin, & Thermes, 2018). Up to now, the genomes of at least 15 domestic sheep breeds have been assembled (Jiang et al., 2014; Li et al., 2021; Davenport et al., 2022). There are various breeds of sheep being commercially bred, but Dorper holds significance due to its special characteristics. Therefore, an annotated complete genome is of great significance for the study of this sheep breed.

As mentioned earlier, Dorper sheep is a composite breed whose ancestors were Blackhead Persian sheep and Dorset Horn sheep. Blackhead Persian sheep locally adapted to warm and dry environments are likely to carry genes that could resist the negative effects of global warming and increased aridity. The Dorset Horn sheep is one of the best meat sheep breeds in the world and yields heavily muscled carcasses, best known for its ability to produce a lamb crop any time of the year (Porter, Alderson, Hall, & Sponenberg, 2016). In a population structure research of African sheep breeds, Dorper sheep were clustered in between the Dorset and Blackhead Persian clusters based on PCA and ADMIXTURE analysis. The research confirmed the relationship between Doper sheep and their ancestors (Dzomba et al., 2020). Given the breeding and genomic background of Doper sheep, we can determine that this breed inherits desirable traits from the Dorset Horn and Blackhead Persian sheep. The divergence of gene expression is an essential source of phenotypic diversity and the co-expression of alleles at the parent locus underpins certain traits of diploid hybrids (Knight 2004). Numerous studies have suggested that only one allele is expressed in heterozygotes, and monoallelic expression or an imbalance in heterozygote allelic expression has been studied in depth in humans and other mammals (Song et al., 2013). A study with tissues from goats' hybrids uncovered multiple genes exhibiting allele-specific expression (Cao et al., 2019). Wang et al. investigated the global allele-specific expression and splicing across adipose and muscle from 15 adult crossbred sheep successfully identified ASE genes with a potential role in muscle growth and fat deposition (Wang et al., 2022). These studies have demonstrated that ASE (allele-specific expression) could contribute to explaining the genetic basis of composite breeds. In this study, we tried to establish a workflow to investigate the ancestral genomic components of Dorper sheep by ASE analysis.

The selection of reference genomes is also crucial to complete ASE analysis. In essence, a reference assembly is an attempt at a complete representation of the nucleotide sequence of an individual genome. This reference assembly allows for a shortcut when sequencing future samples/individuals as they can be mapped to the reference, instead of building a new assembly.

In this study, we need to compare the transcriptome data of Dorper sheep with the appropriate reference genome to detect SNP and INDEL that can represent the Population of the Dorper sheep. There are many different versions of the sheep genome, however, the genomes of different sheep breeds are not exactly the same. Compared with the genome of Doper sheep, some regions may be missing in others, and some regions may be very different. Therefore, the genome of other sheep breeds may induce biases in variant calling, and SNPs and INDEL in those missing regions cannot be detected. That is to say, using the Doper sheep itself as a reference genome can greatly reduce the genetic differences caused by different breeds, thus, we can ensure the pertinence and accuracy of the SNP and INDEL that we found. It is for these reasons that the assembly of the Dorper sheep reference genome is necessary.

In the present work, we completed two tasks. First, a high-quality chromosome-level reference genome was assembled using Oxford Nanopore Technology (ONT) sequencing and chromatin conformation capture (Hi-C) technology. This research fills a gap in the reference genome of hair sheep. Second, we provided a pipeline for mining genetic information for composite breeds using allele-specific expression analysis. Using the assembled Dorper sheep genome as the reference genome, a data set of ancestor-specific differential SNP and INDEL loci was obtained using ancestral parent population genome re-sequencing data. Based on this dataset, Dorper sheep transcriptome data were used to verify allele specific expression sites. We then annotated these sites to identify the ASE genes. The study identified 21,289,550 SNPs in the ancestors, from which 5,701 ancestral unique SNP genotypes were obtained by genotyping. Similarly, the research detected 2,388,815 indels, which contained 476 ancestral genotypes. Based on outcomes from genotyping, 1,002 and 292 genes were identified from these SNPs and indels, respectively. These results contribute to our understanding of the genomic architecture of Dorper sheep.

2 MATERIALS AND METHODS

2.1 Genome Assembly and Annotation

2.1.1 Ethics Statement and Sample Collection

All animal experimental procedures in this study were approved by the Ethics Committee of the Lanzhou Institute of Husbandry

and Pharmaceutical Sciences of the Chinese Academy of Agricultural Sciences. For whole-genome sequencing using the PromethION and Illumina sequencer instrument, blood sample of a Dorper sheep was collected from a healthy male from the Tianjin Aoqun Animal Husbandry Pty., Ltd. The blood was used for genomic DNA (gDNA) extraction, sequencing, and Hi-C library construction.

2.1.2 DNA Extraction and Sequencing

High molecular weight genomic DNA was prepared from the blood using the SDS method followed by purification with QIAGEN® Genomic kit (Cat#13343, QIAGEN) according to the standard operating procedure provided by the manufacturer. The degradation and contamination of the extracted DNA were monitored on 1% agarose gel. DNA purity was then detected using a NanoDrop™ UV-Vis spectrophotometer (Thermo Fisher Scientific, United States), with an OD260/280 ranging from 1.8 to 2.0 and OD 260/230 between 2.0 and 2.2. DNA concentration was further measured using Qubit® 3.0 Fluorometer (Invitrogen, United States).

A total of 2 μg DNA per sample was used as input material for the ONT library preparation. After qualification, a size-selection of long DNA fragments was performed using the BluePippin system (Sage Science, United States). Next, the ends of DNA fragments were repaired, and A-ligation reaction was conducted using a NEBNext Ultra II End Repair/dA-Tailing Module (Cat# E7546, NEB). The adapter in the LSK109 kit (SQK-LSK109, Oxford Nanopore Technologies, United Kingdom) was used for further ligation reaction, and the Qubit® 3.0 Fluorometer was used to quantify the size of library fragments. Sequencing was then performed on a PromethION sequencer instrument (Oxford Nanopore Technologies, United Kingdom), using Nextomics.

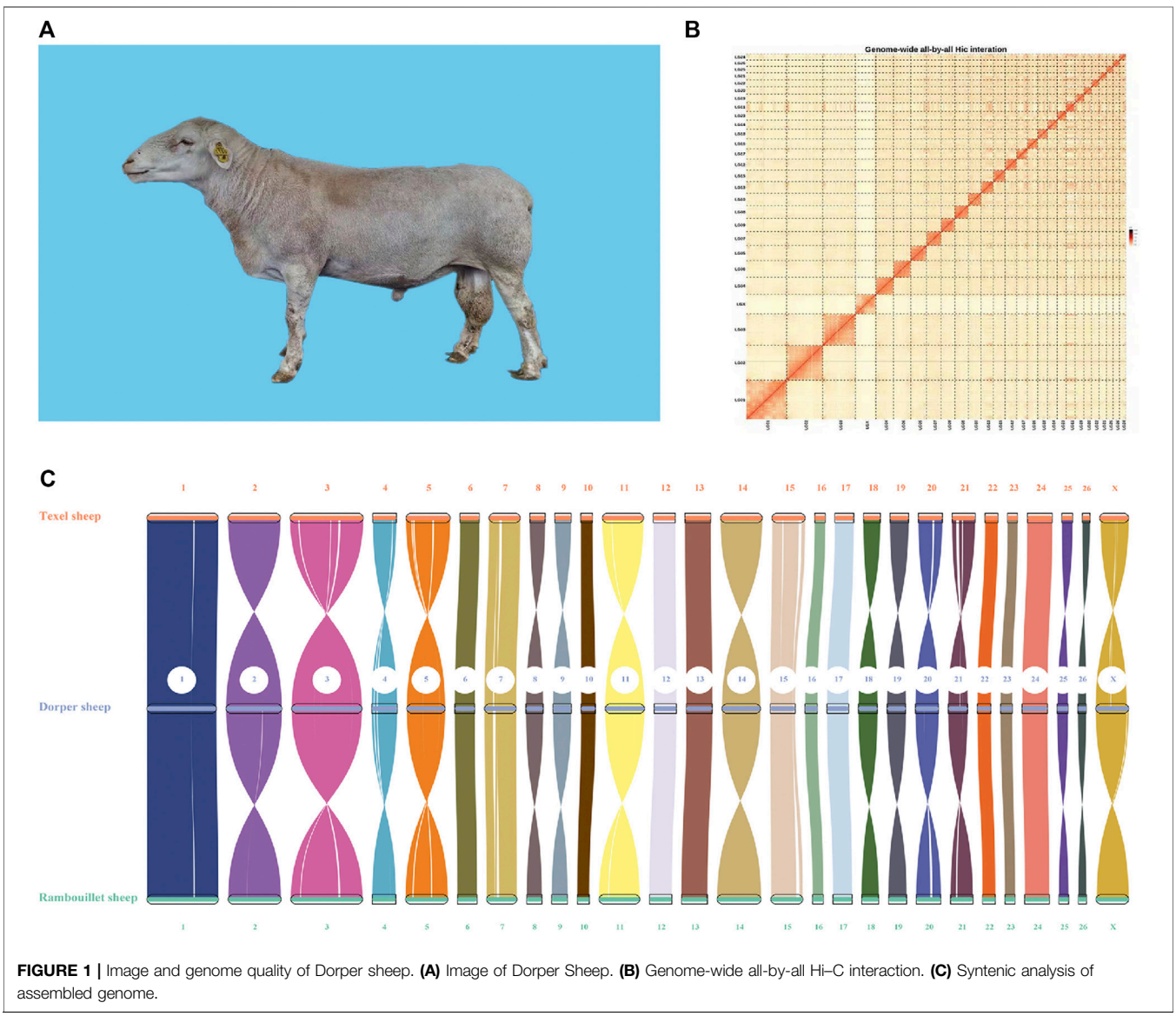

FIGURE 1 | Image and genome quality of Dorper sheep. **(A)** Image of Dorper Sheep. **(B)** Genome-wide all-by-all Hi–C interaction. **(C)** Syntenic analysis of assembled genome.

2.1.3 Data Quality Control

Base-calling was performed to convert the FAST5 files to fastq format with Guppy (V 3.2.2 + 9fe0a78) (https://github.com/nanoporetech/taiyaki). The raw reads of fastq format with a mean_qscore_template <7 were then filtered, resulting in pass reads.

2.1.4 Genome Size and Heterozygosity Estimation

The k-mer analysis was performed using Illumina short-read data prior to genome assembly, to estimate genome size and heterozygosity. Briefly, quality-filtered reads were subjected to 17-mer frequency distribution analysis using the Jellyfish (v2.3.0) (Marçais & Kingsford, 2011). We estimated the genome size of the Dorper with the following equation: G = K-num/K-depth (where K-num is the total number of 17-mers, K-depth denotes the k-mer depth, and G represents the genome size). Further combination of the simulation data resulted from Arabidopsis with different heterozygosity and the frequency peak distribution of 17-mers was done to estimate the heterozygosity and repeat content of the Dorper genome.

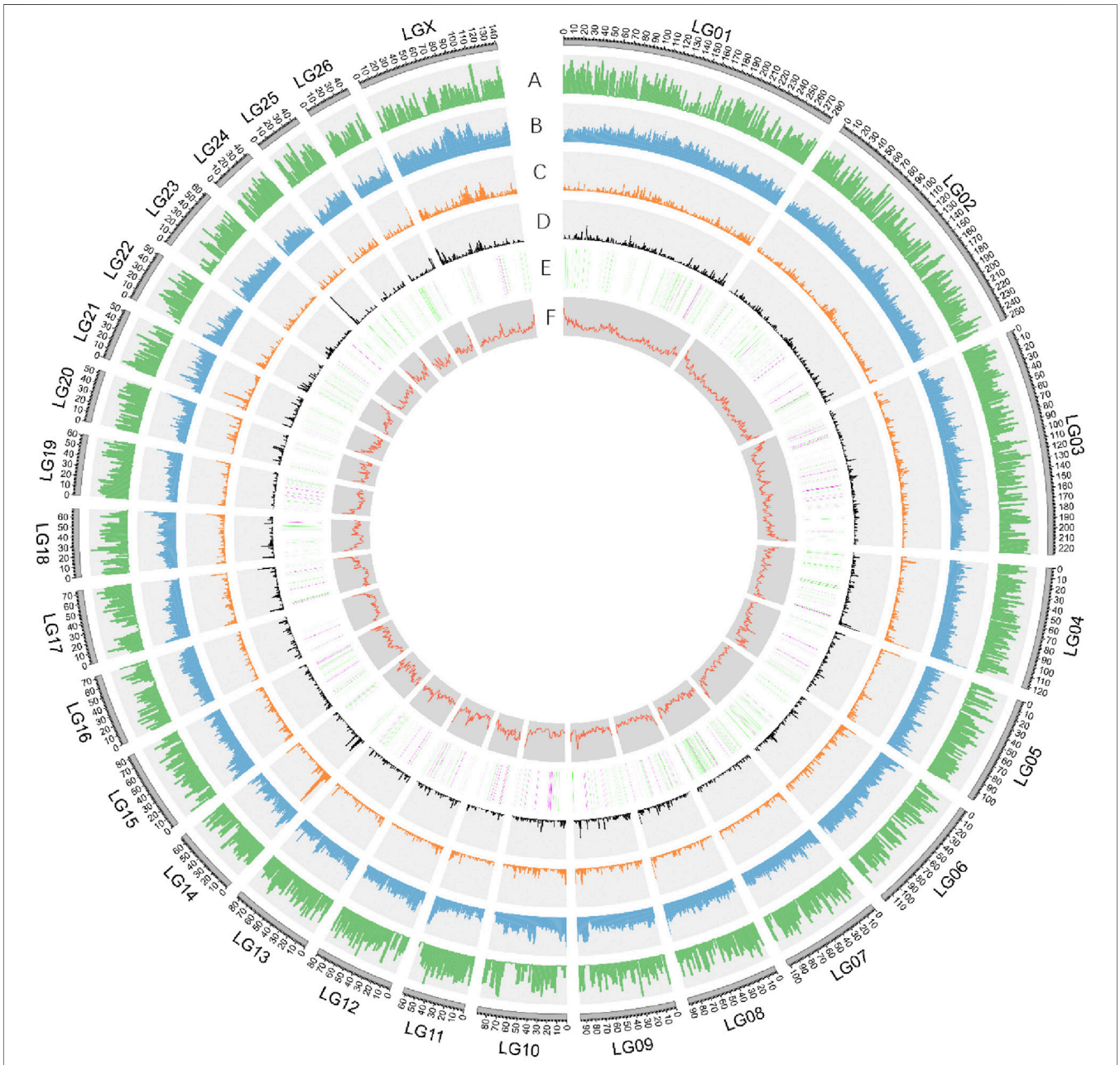

FIGURE 2 | Characterization of Dorper sheep genome. Circos diagram showing (outer to inner): A. Gene density (genes per 1 Mb window). B. Repeat density (repetitive sequence per in 1 Mb window). C. Gypsy-type LTR density. D. Copia-type LTR density. E. The numbers and distribution of ASE genes derived from DSS and PSS were represented in green and purple respectively. F. GC content in sliding windows of 1 Mb across each chromosome.

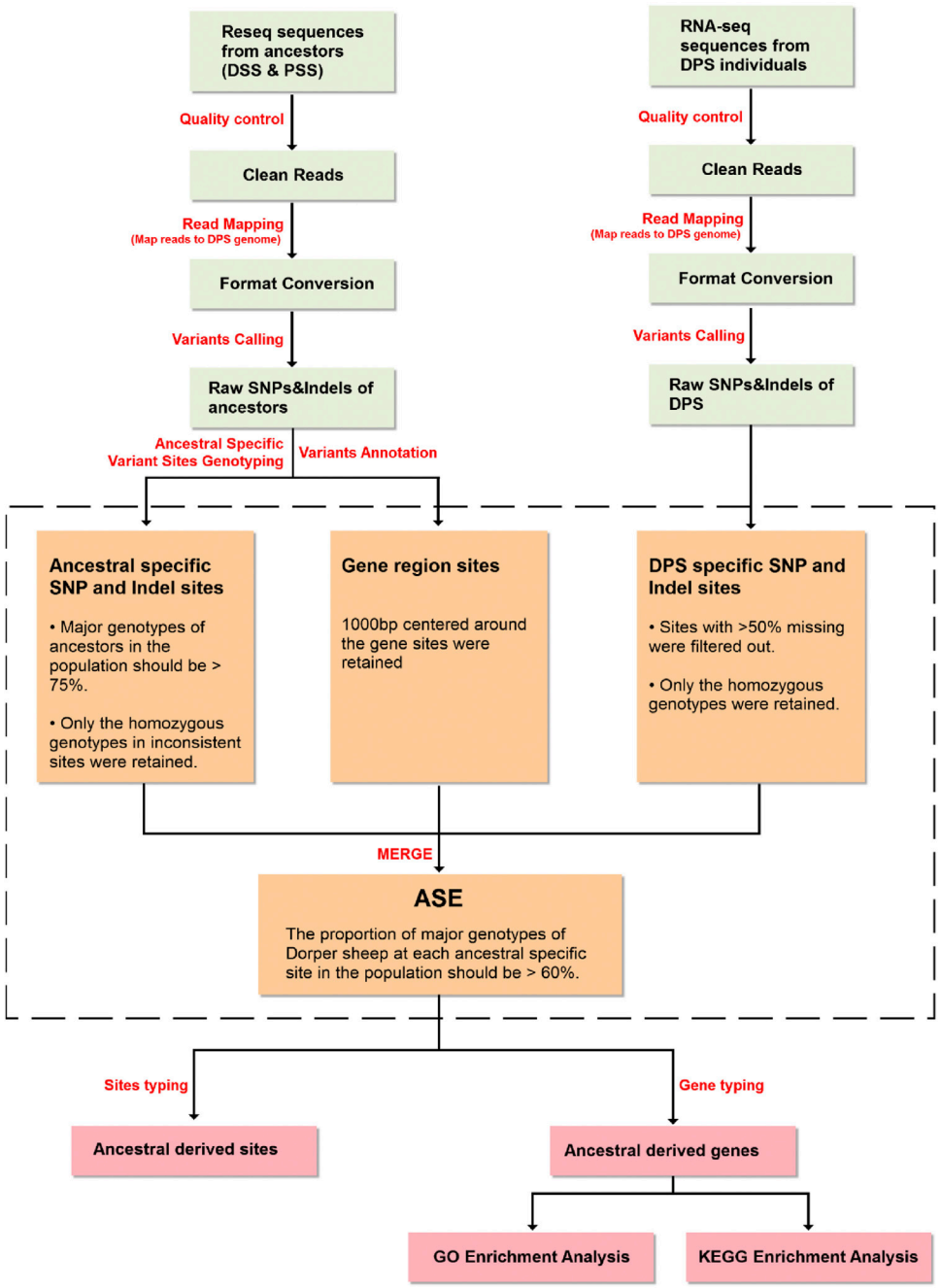

FIGURE 3 | Bioinformatics pipeline for allelic genes expression estimation.

2.1.5 *De Novo* Assembly

For *de novo* genome assembly, ONT-only assembly was constructed using a string graph method with NextDenovo (v2.3.1) (https://github.com/Nextomics/NextDenovo.git). Considering the high error rate of ONT raw reads, the original subreads were first self-corrected using the NextCorrect module, resulting in consistent sequences (CNS reads). Comparison of the CNS reads was then performed with the NextGraph module to capture correlations of the CNS. Based on the correlation of CNS, the preliminary genome was assembled. To improve the accuracy of the assembly, the contigs were refined with Racon (v1.3.1) (https://github.com/isovic/racon.git) using ONT long reads and polished with Nextpolish (v1.3.0) (https://github.com/Nextomics/NextPolish.git) using Illumina short reads with default parameters. To discard possibly redundant contigs and

TABLE 1 | Genome assembly statistics of Dorper sheep.

Statistic	Contig length (bp)	Contig number
N50	73,326,320	13
N50	64,997,665	17
N50	43,243,940	21
N50	37,997,576	28
N50	22,706,521	37
Longest	158,282,255	1
Total	2,648,309,365	140
Length >= 1 kb	2,648,309,365	140
Length >= 2 kb	2,648,309,365	140
Length >= 5 kb	2,648,309,365	140

generate a final assembly, similarity searches were performed using Redundans (Pryszcz & Gabaldón, 2016) with the parameters "-identity 0.9-overlap 0.9".

The completeness of genome assembly was assessed using BUSCO (v4.0.5) (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). To evaluate the accuracy of the assembly, all the Illumina paired-ended reads were mapped to the assembled genome using BWA (Burrows-Wheeler Aligner) (Li & Durbin, 2009) and the mapping rate, as well as genome coverage of sequencing reads, was assessed using SAMtools (v0.1.1855) (Kurtz et al., 2004). The base accuracy of the assembly was calculated with bcftools (v1.8.0) (http://samtools.github.io/bcftools/).

The coverage of expressed genes of the assembly was examined by aligning all the RNA-seq reads against the assembly using Hisat2 (v2.1.0) (http://ccb.jhu.edu/software/hisat/index.shtml) with default parameters. To avoid the inclusion of mitochondrial DNA sequences in the assembly, the draft genome assembly was submitted to the NT library to check contamination.

2.1.6 Chromosome Assembly Based on Hi-C Technology

To anchor hybrid contigs onto the chromosome, genomic DNA was extracted from the Dorper male for the Hi-C library construction and sequencing *via* the Illumina Novaseq/MGI-2000 platform. In total, 370 million paired-end reads were generated from the libraries. Then, quality controlling of Hi-C raw data was performed using Hi-C-Pro (v2.8.1) (Servant et al., 2015). First, low-quality sequences (quality scores < 20), adaptor sequences, and sequences shorter than 30 bp were filtered out using fastp v0.19.8 (https://github.com/OpenGene/fastp), and then the clean paired-end reads were mapped to the draft assembled sequences using bowtie2 (v2.3.2) (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) to get the unique mapped paired-end reads. For further analysis, valid interaction paired reads were identified and retained by the HiC-Pro from uniquely mapped paired-end reads. Invalid read pairs, including dangling-end, self-cycle, re-ligation, and dumped products, were filtered by the HiC-Pro. The contigs were further clustered, ordered, and oriented onto chromosomes using LACHESIS (Burton et al., 2013). Finally, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

2.1.7 Repetitive Sequence Detection

We first annotated tandem repeats using GMATA (v2.2) (https://sourceforge.net/projects/gmata/?source=navbar) and Tandem Repeats Finder (TRF) (v4.07b) (http://tandem.bu.edu/trf/trf.html) where GMATA identified simple repeat sequences (SSRs) and TRF recognized all tandem repeat elements in the whole genome. Transposable elements (TE) in the Dorper genome were then identified using a combination of ab initio and homology-based methods. Briefly, a ab initio repeat library for Dorper was first predicted using MITE-hunter (https://github.com/jburnette/MITE-Hunter) and Repeat Modeler (v open-1.0.11) (https://github.com/Dfam-consortium/RepeatModeler) with default parameters. The resultant library was then aligned to TEclass RepBase (http://www.girinst.org/repbase) to classify the type of each repeat family. For further identification of the repeats throughout the genome, RepeatMasker (Chen, 2004) was applied to search for known and novel TEs by mapping sequences against the *de novo* repeat library and RepBase (Bao, Kojima, & Kohany, 2015) TE library. Overlapping transposable elements belonging to the same repeat class were collated and combined.

2.1.8 Gene Prediction and Annotation

Three independent approaches, including ab initio prediction, homology search, and reference guided transcriptome assembly, were used for gene prediction in a repeat-masked genome. GeMoMa (v1.6.1) (Keilwagen et al., 2016) was used to align the homologous peptides from related species to the assembly to obtain the gene structure information, which was the homolog prediction. For RNA-seq-based gene prediction, filtered mRNA-seq reads were aligned to the reference genome using STAR (vSTAR-2.7.3a) (https://github.com/alexdobin/STAR). The

TABLE 2 | Assembly quality statistics comparison.

Assembly statistic	ASM1914517V1	ARS-UI_Ramb_v2.0	Oar_rambouillet_v1.0	Oar_v4.0
Total Length (Mb)	2,648.31	2,628.15	2,869.91	2,615.52
Contig No.	140	226	7,486	48,482
Contig N50 (bp)	73326320	43,178,051	2,850,956	145,655
Contig L50 (No. of contigs)	13	24	263	5,206
Complete, single-copy BUSCOs (%)	91.58	93.9	93.0	91.2
Complete, duplicated BUSCOs (%)	1.6	2.1	2.6	1.6
Percent of fragmented BUSCOs	2.22	0.9	1.1	2.4
Percent of missing BUSCOs	6.2	3.1	3.3	4.8

TABLE 3 | The top ten genes from Persian sheep

	Gene symbol	Full name	CHR
1	*SOCS2*	Gene—Suppressor Of Cytokine Signaling 2	3
2	*MYCB2*	MYC Binding Protein 2	11
3	*ARFGEF2*	ADP Ribosylation Factor Guanine Nucleotide Exchange Factor 2	12
4	*SEC31*	SEC31 Homolog A, COPII Coat Complex Component	6
5	*EXT2*	Exostosin Glycosyltransferase 2	13
6	*ITPR1*	Inositol 1,4,5-Trisphosphate Receptor Type 1	21
7	*ARIH1*	Ariadne RBR E3 Ubiquitin Protein Ligase 1	8
8	*LRIG1*	Leucine Rich Repeats And Immunoglobulin Like Domains 1	21
9	*UCHL4*	Ubiquitin C-Terminal Hydrolase L4	11
10	*UBP22*	UBIQUITIN-SPECIFIC PROTEASE 22	20

transcripts were then assembled using Stringtie (v1.3.4) (Pertea et al., 2015) and Open Reading Frames (ORFs) were predicted using TransDecoder (v2.0) (https://sourceforge.net/projects/transdecoder/) and PASA (v2.3.3) (Haas et al., 2003). For the *de novo* prediction, Augustus (v3.3.1) (Stanke & Waack, 2003), Genscan (v.3.1) (Burge & Karlin, 1997), GeneID (v.1.4) (Blanco, Parra, & Guigó, 2007), GlimmerHMM (v.1.2) (Majoros, Pertea, & Salzberg, 2004), GeneMarkS-T (v 4) (Besemer, Lomsadze, & Borodovsky, 2001), and SNAP (v.2006- 07-28) (Korf, 2004) were used with the default parameters. Finally, EVidenceModeler (v1.1.1) (Haas et al., 2008) was used to produce an integrated gene set from which genes with TEs were removed using the TransposonPSI package (http://transposonpsi.sourceforge.net/) and the miscoded genes were further filtered out.

Gene function information, motifs, and domains of the proteins were assigned by comparing them with public databases, including SwissProt, NR, KEGG, KOG, and Gene Ontology. The putative domains and GO terms of genes were identified using InterProScan (https://github.com/ebi-pf-team/interproscan/wiki) with default parameters. For the other four databases, BLASTP (v2.7.1) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to compare the Evidence Modeler-integrated protein sequences against the four well-known public protein databases with an E value cutoff of 1e−05; the results with the hit having the lowest E value were retained. Results from the five database searches were concatenated.

2.1.9 Annotation of Non-Coding RNAs

To obtain the ncRNA (non-coding RNA), two strategies were used: searching against a database and prediction with a model. Transfer RNAs (tRNAs) were predicted using tRNAscan-SEM (v2.0) (Lowe & Eddy, 1997) with eukaryote parameters. MicroRNA, rRNA, small nuclear RNA, and small nucleolar RNA were detected using Infernal (v1.1.2) (http://eddylab.org/infernal/) to search the Rfam (Griffiths-Jones et al., 2005) database. The rRNAs and their subunits were predicted using RNAmmer (v1.2) (http://www.cbs.dtu.dk/services/RNAmmer/).

2.2 Ancestral Genomic Components Excavation

2.2.1 Data Quality Control and Mapping

Genome resequencing data of ancestral breeds were downloaded from NCBI. The project numbers of three Persian sheep and 18 Dorset sheep were PRJEB39179 and PRJNA675420, respectively. The RNA-Seq data of six Dorper sheep were downloaded from NCBI BioProjects PRJNA631066.

Before mapping to Dorper reference genome, the data were processed to filter out low-quality reads. The Fastp was used to filter the original data with default parameters. The filtering conditions were as follows: 1) removing adapters of reads; 2) removing reads containing more than 10% of unknown nucleotides; 3) removing low-quality reads containing more than 50% of low-quality (Q-value ≤10) bases; and 4) removing reads for which the average base quality value was less than 20. The filtered reads were then mapping to the Dorper sheep reference genome with BWA v0.19.8 (https://github.com/lh3/bwa).

TABLE 4 | The top ten genes from Dorset sheep.

	Gene symbol	Full name	CHR
1	*UB2E2*	Ubiquitin Conjugating Enzyme E2 E2	26
2	*KCNQ5*	Potassium Voltage-Gated Channel Subfamily Q Member 5	9
3	*PCNX1*	Pecanex 1	8
4	*ARHGAP24*	Rho GTPase Activating Protein 24	6
5	*STX8*	Syntaxin 8	20
6	*KIAA0586*	KIAA0586	8
7	*MRPL42*	Mitochondrial Ribosomal Protein L42	3
8	*KTN1*	Kinectin 1	8
9	*DCAF5*	DDB1 And CUL4 Associated Factor 5	8
10	*UTRN*	Utrophin	10

2.2.2 Variants Detection and Annotation

SNPs and indels were called and filtered using GATK v4.0 (https://gatk.broadinstitute.org/hc/en-us) and VCFtools v0.1.13 (http://vcftools.sourceforge.net/). First, GATK quality value and density were used for filtering. SNPs were filtered with the following parameters: --filter Expression "QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 4.0" -cluster 3 -window 10. The indels were filtered with the following parameters: --filter Expression QUAL < 30.0 || QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQ < 40.0. Then VCFtools was used to filter loci by allele frequency and depth with the following parameters: --min-alleles 2 --max-alleles 2 --min-meanDP 5 --maf 0.05 --max-missing 0.5. ANNOVAR (https://annovar.openbioinformatics.org/en/latest/) was used for the annotation of SNPs and indels.

2.2.3 Allele-Specific Expression Analysis

Allele counts at SNP and indel positions were retrieved using an in-house Python script. The screening conditions for specific SNPs and indels in ancestors (Dorset and Persian sheep) were as follows: 1) genotypes of ancestors were confirmed based on the proportion of the major genotype of ancestors in the population being > 75% (e.g., among 18 Dorset sheep, when a genotype appeared in more than 14 individuals, it can be considered as a major genotype) and 2) to establish the ancestral source of an offspring allele, only the homozygous major genotypes in inconsistent sites were retained.

SNP and indel genotyping and screening procedures of Dorper sheep were as follows: 1) considering the most commonly missing sites in the intergenic region of the transcriptome data compared to the re-sequencing data, the sites with >50% missing data were filtered out; 2) the major genotype proportion of Dorper sheep must be >60% (e.g., if six genotypes were detected, when a genotype appeared in more than four individuals, it could be considered as a major genotype at this locus). Subsequently, heterozygous and missing genotypes were filtered out; 3) by comparing the major genotypes of the Dorper sheep with the main genotypes of the ancestors (Dorset and Persian sheep), the ancestor from which the site was mainly derived from can be determined; 4) through the variant annotation information which was obtained in the previous step, the information for each gene corresponding to the variation site was confirmed. This information included which allele and gene structure a variant is located on; and 5) Finally, we counted the number of SNPs and indels for each allele to determine which ancestor the allele is mainly derived from.

3 RESULTS

3.1 Genome Assembly and Annotation

3.1.1 Genome Assembly

In this study, ~163.08 Gb of filtered Illumina short-read sequencing data were obtained from the Dorper sheep (**Figure 1A**; **Supplementary Table S1**). The size of the Dorper genome was estimated to be around 2.65 Gb with 0.4% heterozygosity (**Supplementary Figure S1**).

After filtering out adaptor sequences, ~224.57 Gb of ONT subreads with an average length of 21.17 Kb was obtained. These clean reads were used to *de novo* assemble the genome which was refined and polished with Illumina short reads. About 91.58% of the conserved genes could be detected in the Dorper genome by the BUSCO software using mammalia_odb10 dataset, confirming the high completeness of the obtained genome (**Supplementary Table S2**).

Finally, we used the Hi-C technique to anchor assembly contigs in 27 chromosomes (2n = 54). We found that 831,180,895 uniquely mapped paired-end reads were generated and occupied ~70.41% of the total clean paired-end reads (1,180,468,259). The frequency of contig interactions was estimated on the basis of pairs mapped to the contigs. We found that 107 contigs, representing 77.54% of all contigs and 99.57% of the whole genome nucleotide bases were successfully anchored on 27 chromosomes (**Supplementary Table S3** and **Figure 1B**). The final assembly resulted in a 2.64 Gb genome with a contig N50 of 73.33 Mb. The genome consisted of 140 contigs, with the longest contig being 158.3 Mb (**Table 1**).

3.1.2 Genome Characteristics

We found that the GC content of the Dorper genome was 41.99% (**Figure 2**, **Supplementary Figure S2**), which was similar to that of other domestic sheep breeds (42.12%), snow sheep (42.12%), and goats (41.5%) (Upadhyay et al., 2020). TEs contributed 1,202,782,366 bp of the genome and accounted for 45.42% of the genome length (**Supplementary Table S4**). We found that class I TEs (RNA transposons or retrotransposons) occupied 42.54% of the genome. The most abundant retrotransposons found in the Dorper genome were long interspersed nuclear elements (LINE), which constituted 78.51% of all identified class I transposons. Moreover, the Dorper genome was not rich in class II TEs (DNA transposons), which occupied only 2.68% of the genome content. The assembly quality statistics comparison is listed in **Table 2**.

The obtained consensus gene set included 20,450 protein-coding genes (**Supplementary Table S5**). For the completeness of protein-coding genes, 94.18 and 2.22% of the "total complete BUSCOs" and "fragmented BUSCOs" were identified by BUSCO annotation, respectively. The average coding sequence length (CDS), average exon length, and average intron length were 1,571, 160, and 5,477 bp, respectively (**Supplementary Figure S5** and **Supplementary Table S6**). There were on average 9.8 exons per gene. We found functional annotation for 18,491 protein-coding genes, which represented about 90.42% of all the genes (**Supplementary Table S7**). Non-protein-coding genes included 251,525 tRNAs, 412 rRNAs, and 769 microRNAs (miRNAs) (**Figure 2**, **Supplementary Table S8**).

To evaluate quality of the genome annotation, the collinearity analysis of the Dorper sheep was conducted with Texel (Oar_v4.0) and Rambouillet (ARS-UI_Ramb_v2.0) sheep. The high collinearity observed among these three genomes illustrated that the accuracy of the Dorper genome assembly and annotation was high (**Figure 1C**).

3.2 Ancestral Genomic Components

3.2.1 Genomic Variants

A single-nucleotide polymorphism (SNP) and an indel database were developed between Dorset sheep (DSS), Persian sheep (PSS), and Dorper sheep (DPS). The re-sequencing data of DSS and PSS achieved an average depth of 8× and a mapping rate of 99.70% (**Supplementary Table S9**). The RNA-Seq data of DPS achieved an average depth of 4× and a mapping rate of 99.74% (**Supplementary Table S9**). A total of 21,289,550 SNPs were found in all data. The proportion of transitions (15,147,780, 71.15%) was much higher than that of transversions (6,141,770, 28.85%). The transition: transversion ratio was 2.47, which was similar to that found in other studies (Guan et al., 2016). A total of 2,388,815 indels were sought out among the three breeds. There were more deletions (1,306,121) than insertions (1,082,694).

3.2.2 Allele-Specific Expression Analysis

The ancestral alleles for all the SNPs and indels were inferred by comparing these variants to the Dorper sheep genome (**Table 3**, **Table 4**, **Figure 2**). Initially, we determined that 5,701 SNPs (1,000 bp centered around the gene site) were located inside 1,002 genes with at least one discriminating SNP, of which 260 SNP alleles from the Persian sheep, 723 SNP alleles from the Dorset sheep, and 19 SNP alleles were present in both breeds (**Supplementary Table S10**). At the above 5,701 SNP sites, 1,247 SNPs were from Persian sheep and 4,454 SNPs were from Dorset sheep. The same method for indels was used. The analysis detected 456 indels located inside 294 genes. In these indel mutant alleles, 66 alleles belonged to Persian sheep and 228 alleles to Dorset sheep (**Supplementary Table S11**).

3.2.3 Enrichment Analysis of ASE Genes

To explore the role of the genes carrying ASE, functional enrichment analyses were performed. Gene Ontology (GO) enrichment analysis of the 260 SNP alleles of Persian sheep showed that there were seven significant GO terms in molecular function (MF) and two significant GO terms in biological process (BP) ($p < 0.05$) (**Supplementary Table S12**, **Supplementary Figure S4**). GO enrichment analysis of the 723 SNP alleles of Dorset sheep showed that GO terms were significantly enriched in 125 ASE genes ($p < 0.05$), which were mainly involved in eight MF, five BP, and one Cellular Component (CC) (**Supplementary Table S13**, **Supplementary Figure S5**). For indel mutant alleles, GO enrichment analysis identified eight significantly ($p < 0.05$) enriched GO terms composed of four GO terms in MF and four GO terms in BP in Persian sheep (**Supplementary Table S14**, **Supplementary Figure S6**). The GO enrichment analysis showed that the indel mutant alleles of Dorset sheep were involved in nine GO terms ($p < 0.05$): four GO terms in MF, four GO terms in MF, and one GO terms in CC (**Supplementary Table S15**, **Supplementary Figure S7**).

KEGG analysis resulted in 61 significant ($p < 0.05$) SNP alleles of Persian sheep to be annotated to 15 KEGG pathways (**Supplementary Table S16**, **Supplementary Figure S8**). According to the annotation of KEGG, 218 SNP alleles of Dorset sheep were significantly ($p < 0.05$) annotated to 27 KEGG pathways (**Supplementary Table S17**, **Supplementary Figure S9**). After KEGG analysis, indel alleles of Persian sheep were mapped to seven significant KEGG pathways ($p < 0.05$) (**Supplementary Table S18**, **Supplementary Figure S10**). Indel alleles of Dorset sheep were mapped to 11 significant KEGG pathways ($p < 0.05$) (**Supplementary Table S19**, **Supplementary Figure S11**).

4 DISCUSSION

4.1 Genome Assembly and Annotation

Returning to the question posed at the beginning of this study, it is now possible to state that the first chromosome-scale reference genome of hair sheep is assembled. As of 2022, more than 15 domestic sheep breeds genome sequences have been recorded in the National Center for Biotechnology Information (NCBI). From the first version of the sheep reference genome (PRJNA33937) published by The International Sheep Genomics Consortium in 2010 to the 13 versions of the sheep genome involved in the pan-genome article published by Li et al., in 2021 (Li et al., 2021), sheep genome sequencing assembly has undergone a process from first-generation sequencing technology to third-generation sequencing technology. The ARS-UI_Ramb_v2.0 (Davenport et al., 2022), and previously Oar_rambouillet_v1.0, and Oar_v4.0 (Jiang et al., 2014) is the current reference genome for sheep. Comparing these versions, we found that the Contig N50 of Oar_v4.0 to ARS-UI_Ramb_v2.0 became longer gradually, from 145 kb to 43 M, indicating an obvious improvement in the assembly level. The advent of third-generation sequencing, long read-sequencing has meant scientists can now generate many sheep genomes from different breeds and populations from around the world. Recently a sheep pan-genome was published (Li et al., 2021) that included long-read genome assemblies 13 breeds. Many years of natural and artificial selection have produced abundant phenotypic variation in sheep populations. Different breeds contribute genetic diversity to global sheep genetic resources. Genome assembly of different breeds helps to reveal the origins and evolutionary forces of sheep population structure and constitutes a valuable resource for sheep breeding programs and genetic diversity studies.

Like most genome assembly strategies nowadays, third-generation sequencing technology was used in this study. The major advantage of the third-generation sequencing technology is the long read length. Specifically, we used Oxford Nanopore Technology (ONT) sequencing strategy which was also used in Rambouillet (Davenport et al., 2022) and Hu (Li et at., 2021) sheep. As a result, the final assembled genome size was 2.64 Gb with a scaffold N50 of 101.9 Mb and contig N50 of 73 M. The size of this genome is within the range of published sheep genome sizes, ranging from 2.61 Gb for the Texel sheep (Jiang et al., 2014) to 2.90 Gb for the East Friesian sheep (PRJNA721520). Compared to the other sheep breeds, the scaffold N50 of the Dorper sheep assembly is in the top quartile. In sheep genomes assembled using PacBio and ONT, the longest scaffold N50 is 107.7 M for Rambouillet sheep (Davenport et al., 2022) and the shortest one is 100 M for Texel sheep (Jiang et al., 2014). The contig N50 in our study is longer than those of most sheep genome assemblies. Comparison of Contig N50 with reference genomes is also detailed in **Table 2** of the manuscript. Our study implied that more and more high-quality sheep genomes of different breeds will be assembled with advances in sequencing

technologies and assembly methods and reduced sequencing costs. As the study in Li et al., 13 sheep breeds genomes were assembled at the same time, which include Dorper sheep (Li et at., 2021). The Dorper sheep genome published in the study of Li et al. was assembled using PacBio HiFi sequencing. There were certain unique points to their study, compared with ours. Especially, they assembled 2 haplotype-resolved genome assemblies based on HIFI data. However, their assembly level is still in scaffolds, we generated Hi-C data from the same individual to cluster, order, and orient contigs onto chromosomes. We provided a new assembly for Dorper and a detailed description in this manuscript including annotation and analysis of ASE, providing additional resources for the Dorper breed to those included in the pangenome created by Li et al.

4.2 Ancestral Genomic Components

Allele-specific expression (ASE) analysis identified multiple ASE SNPs and ASE indels in Dorper sheep which were derived from ancestors (**Figure 3**). These ASE genes are related to many essential traits, including growth (*IGF1, DAAM, PHF17, SYNE2, OST1, KIF20*), immune responses (*ABCC4, ARI1, CELF2, TMCO3*), and reproduction (*TAF4B, HTF4, STK10, LAYN*). Here, we had some interesting findings from the enrichment analyses of these alleles. Several GO terms were found in both ancestors. For instance, metal ion binding, protein phosphorylation, protein serine/threonine kinase activity, regulation of Rho protein signal transduction, and Rho guanyl-nucleotide exchange factor activity. Protein phosphorylation is an important factor in the transition from muscle to edible meat (Huang, Larsen, & Lametsch, 2012). It also has important effects on many physiological and biochemical reactions in muscles. Rho is active when bound to GTP and inactive when bound to GDP. It is also known to participate in many physiological activities including cell migration, adhesion, cytokinesis, proliferation, differentiation, and apoptosis, and to a greater extend cell transformation (Heasman & Ridley, 2008). Among KEGG pathways, ubiquitin-mediated proteolysis, autophagy, lysosome, the mTOR signaling pathway, and cellular senescence were detected in both ancestral breeds. The common GO terms and KEGG pathways in these ancestors indicate that growth and development related traits of Dorper sheep are a result of combinations of the maternal and paternal ancestral genomes.

For Persian sheep, there were specific GO terms involved in lipid-related processes, such as lipid binding and galactosyltransferase activity. Galactosyltransferase activity is a catalysis of the transfer of a galactosyl group to an acceptor molecule, typically another carbohydrate or a lipid (Jensen, Schultink, Keegstra, Wilkerson, & Pauly, 2012). These results are consistent with characteristics of Persian sheep, which is a fat-tailed breed (Lundie, 2011). Therefore, we suggest that Persian sheep made more contributions to traits related to fat deposition. In Persian sheep, the metabolic pathways mainly involve carbohydrate metabolisms, such as fructose and mannose metabolism, glycosphingolipid biosynthesis, and other types of O−glycan biosynthesis. The Persian sheep originated in the arid regions of east Africa in what is now Somalia. Their glycolysis pathway and catabolism of carbohydrates were enhanced under drought conditions (Bowne et al., 2012). These metabolism-related genes may explain the genetic basis of drought resistance in Dorper sheep. The GO terms from Dorset sheep were typically involved in muscle-associated events. Such as myosin complex and motor activity. Myosin is a superfamily of motor proteins associated with muscle contraction and a wide range of other motility processes in eukaryotes (Sellers, 2000). In Dorset sheep, the related organismal systems pathways mainly involve multiple signaling pathways. According to these results, we can infer that Dorset sheep have more impact than Persian sheep in the growth rate, carcass quality, and carcass yield of Dorper sheep.

This study set out to gain a better understanding of a hair sheep genome and the genetic basis of Dorper sheep. The present results are significant in at least two major respects. First, we provide the first high-quality reference genome of hair sheep, representing a valuable resource for sheep genetic studies. Second, the evidence from this study suggests that the pipeline we constructed for heterosis evaluation based on ASE genes detection is feasible. Through this approach, we found a number of ASE genes in the ancestral population that potentially contributed to the genetic mechanism of important economic traits of Dorper sheep. The method that we designed to reveal the heterosis might help others to evaluate composite breeds, which has important implications for crossbreeding and improvement through the breeding and selection of new high-quality cultivar sheep. Despite these promising results, questions remain. The adaptation and phenotypic differences of the Dorper sheep may be mediated by a complex network of genes that act in tandem, rather than by the action of a single candidate gene (Lv et al., 2014; Kim et al., 2015). It is therefore difficult to directly draw conclusions regarding the genetic mechanisms underlying the observed traits based only on ASE. Furthermore, with only six Dorper, three Persian, and 18 Dorset sheep data set, the sample size was probably too small to obtain reliable estimates. Further studies are required to better understand the mechanisms underlying the genome of Dorper sheep. Notwithstanding the relatively limited sample, this work offers valuable insights into genetic basis research for composite breeds.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA721526.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Management and Ethics Committee of Lanzhou Institute of Animal Husbandry and Veterinary Medicine, Chinese Academy of Agricultural Sciences.

AUTHOR CONTRIBUTIONS

YY, BY, and GQ conceived the project, TG, ZL and WS collected the samples, PX, YW, XL, QZ and XH performed the genome assembly and data analysis, GQ wrote the manuscript and YY, BY

and PX revised the manuscript. All authors reviewed the manuscript.

FUNDING

This research was supported by the major scientific research task of the Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences "The cultivation and industrialization of new breeds of high-quality and efficient mutton sheep" (CAAS-ZDRW202106), the fine-wool sheep resources and breeding innovation team of the Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (25-LZIHPS-07), and National Wool Sheep Industry Technology System Breeding Technology and Method Post (CARS-39-02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.846449/full#supplementary-material

REFERENCES

Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., et al. (2018). Convergent Genomic Signatures of Domestication in Sheep and Goats. *Nat. Commun.* 9 (1), 813. doi:10.1038/s41467-018-03206-y

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes. *Mobile DNA* 6, 11. doi:10.1186/s13100-015-0041-9

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a Self-Training Method for Prediction of Gene Starts in Microbial Genomes. Implications for Finding Sequence Motifs in Regulatory Regions. *Nucleic Acids Res.* 29 (12), 2607–2618. doi:10.1093/nar/29.12.2607

Blanco, E., Parra, G., and Guigó, R. (2007). Using Geneid to Identify Genes. *Curr. Protoc. Bioinformatics* 18 (1), 1–28. doi:10.1002/0471250953.bi0403s18

Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., et al. (2012). Drought Responses of Leaf Tissues from Wheat Cultivars of Differing Drought Tolerance at the Metabolite Level. *Mol. Plant* 5 (2), 418–429. doi:10.1093/mp/ssr114

Burge, C., and Karlin, S. (1997). Prediction of Complete Gene Structures in Human Genomic DNA. *J. Mol. Biol.* 268 (1), 78–94. doi:10.1006/jmbi.1997.0951

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale Scaffolding of De Novo Genome Assemblies Based on Chromatin Interactions. *Nat. Biotechnol.* 31 (12), 1119–1125. doi:10.1038/nbt.2727

Cao, Y., Xu, H., Li, R., Gao, S., Chen, N., Luo, J., et al. (2019). Genetic Basis of Phenotypic Differences between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-specific Expression in Their F1 Hybrids. *Front. Genet.* 10, 145. doi:10.3389/fgene.2019.00145

Chen, N. (2004). Using Repeat Masker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinformatics* 5. Chapter 4Unit 4.10. doi:10.1002/0471250953.bi0410s05

Cloete, S. W. P., Snyman, M. A., and Herselman, M. J. (2000). Productive Performance of Dorper Sheep. *Small Ruminant Res.* 36 (2), 119–135. doi:10.1016/s0921-4488(99)00156-x

Davenport, K. M., BickhartBickhart, D. M., Worley, K., Murali, S. C., Salavati, M., Clark, E. L., et al. (2022). An Improved Ovine Reference Genome Assembly to Facilitate In-Depth Functional Annotation of the Sheep Genome. *GigaScience* 11, giab096. doi:10.1093/gigascience/giab096

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 (6), 910–923. doi:10.1111/age.12991

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., Bateman, A., et al. (2004). Rfam: Annotating Non-coding RNAs in Complete Genomes. *Nucleic Acids Res.* 33, D121–D124. Database issue). doi:10.1093/nar/gki081

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, H., Na, R., et al. (2016). Scanning of Selection Signature Provides a Glimpse into Important Economic Traits in Goats (*Capra hircus*). *Sci. Rep.* 6, 36372. doi:10.1038/srep36372

Haas, B. J., Delcher, A. L., and Mount, S. M. (2003). Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies. *Nucleic Acids Res.* 31 (19), 5654–5666. doi:10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9 (1), R7. doi:10.1186/gb-2008-9-1-r7

Heasman, S. J., and Ridley, A. J. (2008). Mammalian Rho GTPases: New Insights into Their Functions from *In Vivo* Studies. *Nat. Rev. Mol. Cel Biol* 9 (9), 690–701. doi:10.1038/nrm2476

Hiendleder, S., Kaupe, B., Wassmuth, R., and Janke, A. (2002). Molecular Analysis of Wild and Domestic Sheep Questions Current Nomenclature and Provides Evidence for Domestication from Two Different Subspecies. *Proc. R. Soc. Lond. B* 269 (1494), 893–904. doi:10.1098/rspb.2002.1975

Huang, H., Larsen, M. R., and Lametsch, R. (2012). Changes in Phosphorylation of Myofibrillar Proteins during Postmortem Development of Porcine Muscle. *Food Chem.* 134 (4), 1999–2006. doi:10.1016/j.foodchem.2012.03.132

Jensen, J. K., Schultink, A., Keegstra, K., and Wilkerson, M. (2012). RNA-seq Analysis of Developing Nasturtium Seeds (Tropaeolum Majus): Identification and Characterization of an Additional Galactosyltransferase Involved in Xyloglucan Biosynthesis. *Mol. Plant* 5 (5), 984–992. doi:10.1093/mp/sss032

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The Sheep Genome Illuminates Biology of the Rumen and Lipid Metabolism. *Science* 344 (6188), 1168–1173. doi:10.1126/science.1252806

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using Intron Position Conservation for Homology-Based Gene Prediction. *Nucleic Acids Res.* 44 (9), e89. doi:10.1093/nar/gkw092

Korf, I. (2004). Gene Finding in Novel Genomes. *BMC Bioinformatics* 5, 59. doi:10.1186/1471-2105-5-59

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and Open Software for Comparing Large Genomes. *Genome Biol.* 5 (2), R12. doi:10.1186/gb-2004-5-2-r12

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, R., Gong, M., Zhang, X. Fei. W., Wang, F., Liu, Z., Zhang, L., et al. (2021). The First Sheep Graph-Based Pan-Genome Reveals the Spectrum of Structural Variations and Their Effects on Tail Phenotypes. *bioRxiv* 1222, 472709. doi:10.1101/2021.12.22.472709

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi:10.1093/nar/25.5.955

Lundie, R. S. (2011). The Genetics of Colour in Fat-Tailed Sheep: a Review. *Trop. Anim. Health Prod.* 43 (7), 1245–1265. doi:10.1007/s11250-011-9850-0

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two Open Source Ab Initio Eukaryotic Gene-Finders. *Bioinformatics* 20 (16), 2878–2879. doi:10.1093/bioinformatics/bth315

Marçais, G., and Kingsford, C. (2011). A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of K-Mers. *Bioinformatics* 27 (6), 764–770. doi:10.1093/bioinformatics/btr011

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

Pollott, G. E. (2011). A Suggested Mode of Inheritance for Wool Shedding in Sheep1. *J. Anim. Sci.* 89 (8), 2316–2325. doi:10.2527/jas.2010-3713

Porter, V., Alderson, L., and Hall, S. J. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding*, 2. Volume Pack: Cabi.

Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an Assembly Pipeline for Highly Heterozygous Genomes. *Nucleic Acids Res.* 44 (12), e113. doi:10.1093/nar/gkw294

Roth, S. C. (2019). What Is Genomic Medicine? *jmla* 107 (3), 442–448. doi:10.5195/jmla.2019.604

Sellers, J. R. (2000). Myosins: a Diverse Superfamily. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1496 (1), 3–22. doi:10.1016/s0167-4889(00)00005-7

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., et al. (2015). HiC-Pro: an Optimized and Flexible Pipeline for Hi-C Data Processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, H. (2015). BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* 31 (19), 3210–3212. doi:10.1093/bioinformatics/btv351

Song, G., Guo, Z., Liu, Z., Cheng, Q., Qu, X., Chen, R., et al. (2013). Global RNA Sequencing Reveals that Genotype-dependent Allele-specific Expression Contributes to Differential Expression in rice F1 Hybrids. *BMC Plant Biol.* 13, 221. doi:10.1186/1471-2229-13-221

Stanke, M., and Waack, S. (2003). Gene Prediction with a Hidden Markov Model and a New Intron Submodel. *Bioinformatics* 19 (Suppl. 2), ii215–ii225. doi:10.1093/bioinformatics/btg1080

Tesema, Z., Deribe, B., Kefale, A., Lakew, M., Tilahun, M., Shibesh, M., et al. (2020). Survival Analysis and Reproductive Performance of Dorper X Tumele Sheep. *Heliyon* 6 (4), e03840. doi:10.1016/j.heliyon.2020.e03840

Upadhyay, M., Hauser, A., Kunz, E., Krebs, S., Blum, H., Dotsev, A., et al. (2020). The First Draft Genome Assembly of Snow Sheep (Ovis Nivicola). *Genome Biol. Evol.* 12 (8), 1330–1336. doi:10.1093/gbe/evaa124

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, T. (2018). The Third Revolution in Sequencing Technology. *Trends Genet.* 34 (9), 666–681. doi:10.1016/j.tig.2018.05.008

Wang, F., Shao, J., He, S., Guo, Y., Pan, X., Wang, Y., et al. (2022). Allele-specific Expression and Splicing Provide Insight into the Phenotypic Differences between Thin- and Fat-Tailed Sheep Breeds. *J. Genet. Genomics*, S1673–S8527. doi:10.1016/j.jgg.2021.12.008

Conflict of Interest: XL and QZ were employed by the company Tianjin Aoqun Animal Husbandry Pty., Ltd.

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: 12 May 2022
doi: 10.3389/fgene.2022.818420

Whole Genome Sequencing Provides New Insights Into the Genetic Diversity and Coat Color of Asiatic Wild Ass and Its Hybrids

Hong Dong[1†], Zheng Dong[2†], Fuwen Wang[2], Gang Wang[2], Xiaoyu Luo[2], Chuzhao Lei[2] and Jingbo Chen[1]*

[1]College of Animal Science and Technology, SHIHEZI University, Shihezi, China, [2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China

The diversity of livestock coat color results from human positive selection and is an indispensable part of breed registration. As an important biodiversity resource, Asiatic wild ass has many special characteristics, including the most visualized feature, its yellowish-brown coat color, and excellent adaptation. To explore the genetic mechanisms of phenotypic characteristics in Asiatic wild ass and its hybrids, we resequenced the whole genome of one Mongolian Kulan (a subspecies of Asiatic wild ass) and 29 Kulan hybrids (Mongolian Kulan ♂×Xinjiang♀), and the ancestor composition indicated the true lineage of the hybrids. XP-EHH (Cross Population Extended Haplotype Homozygosity), θπ-ratio (Nucleotide Diversity Ratio), CLR (Composite Likelihood Ratio) and θπ (Nucleotide Diversity) methods were used to detect the candidate regions of positive selection in Asiatic wild ass and its hybrids. Several immune genes (*DEFA1*, *DEFA5*, *DEFA7*, *GIMAP4*, *GIMAP1*, *IGLC1*, *IGLL5*, *GZMB* and *HLA*) were observed by the CLR and θπ methods. XP-EHH and θπ-ratio revealed that these genes are potentially responsible for coat color (*KITLG*) and meat quality traits (*PDE1B* and *MYLK2*). Furthermore, the heatmap was able to show the clear difference in the haplotype of the *KITLG* gene between the Kulan hybrids and Asiatic wild ass group and the Guanzhong black donkey group, which is a powerful demonstration of the key role of *KITLG* in donkey color. Therefore, our study may provide new insights into the genetic basis of coat color, meat quality traits and immunity of Asiatic wild ass and its hybrids.

Keywords: asiatic wild ass, whole genome analysis, KITLG, genetic diversity, coat color

OPEN ACCESS

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

Reviewed by:
Beatriz Gutierrez Gil,
Universidad de León, Spain
Qianjun Zhao,
Institute of Animal Sciences, (CAAS), China

***Correspondence:**
Jingbo Chen
chenjb126@126.com

[†]These authors have contributed equally to this work and share first authorship

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

Received: *19 November 2021*
Accepted: *25 April 2022*
Published: *12 May 2022*

Citation:
Dong H, Dong Z, Wang F, Wang G, Luo X, Lei C and Chen J (2022) Whole Genome Sequencing Provides New Insights Into the Genetic Diversity and Coat Color of Asiatic Wild Ass and Its Hybrids. Front. Genet. 13:818420. doi: 10.3389/fgene.2022.818420

INTRODUCTION

Coat color is one of the most visualized breed features of livestock. Long-term natural selection allows animals to have a variety of coat colors, which has become the focus of animal breeding research. *KIT* gene plays a vital role in the occurrence of animal coat color, and it is the pigmentary switch in domestic animal species (Klungland and Vage, 2003). The diversiform coat color phenotypes segregating in Duroc hybrid pigs demonstrated that *KIT* is responsible for the complex variation (Wu et al., 2019). In addition, the coat color phenotypes of several species showed an indispensable association with the *KIT* gene (Zhang et al., 2015; Anello et al., 2019; Voß et al., 2020; Wen et al., 2021). The *KITLG* gene encodes the KIT ligand protein, which has a vital

function in the pigment formation process (Yang et al., 2018). The research of *KITLG* has made some achievements in terms of mammalian coat color. The hair color of humans living in Europe is impacted by the regulatory region and variant of *KITLG* (Sulem et al., 2007; Guenther et al., 2014). A selection signature study in indigenous Chongming white goats indicated that *KITLG* was strongly selected and is also crucial for pigment intensity in dogs (Gao et al., 2020; Weich et al., 2020). A previous study indicated that several ROHs (Runs of homozygosity, ROH) exist in the *KITLG* region in horses (Metzger et al., 2015). A genome-wide study of six donkey breeds in China revealed *KITLG* as a candidate gene for donkey color, and the formation mechanism of diluted gray pigmentation (Dun phenotype) in wild ass has been reported (Wang et al., 2020; Zhou et al., 2020); however, there are few studies on the coat color of Asiatic wild ass.

The domestication of donkeys promoted early human communication, trade and transportation. Domestic donkeys are a significant transport and economic animal in China, and they are concentrated between 32° and 42° north latitude. Domestic donkey lives in dry and warm climates in northwest, north and southwest China (Chang, 2011). The ancestors of modern domestic donkeys are considered to be Nubian wild ass (*Equus africanus africanus*) and Somali wild ass (*Equus africanus somaliensis*) (Rosenbom et al., 2015), which are subspecies of African wild ass. Mongolian kulan (*Equus hemionus hemionus*) is a kind of subspecies of Asiatic wild ass (Oakenfull et al., 2000), and it is distributed in central and western Asia (Richard et al., 2001). The Gobi regions in Inner Mongolia, Gansu and Xinjiang of northern China constitute the most important remaining stronghold of the Mongolian kulan (Gerritsmann et al., 2016). Kulan is a representative species in deserts (Zhang et al., 2020). They live in harsh natural conditions, can adapt to dry environments nicely, and have the habit of seasonal short-distance migration and cluster activities. Xinjiang donkey is an excellent miniature local breed in China and is mainly distributed in Hotan, Kashgar and Aksu of the Xinjiang Uygur Autonomous Region (Lei et al., 2005). Mongolian Kulan has an earthy yellow coat on its back and a white belly, which is typical of its appearance, but Xinjiang donkey has a mainly gray coat. Interestingly, the hybrid populations of Mongolian kulan male parents and Xinjiang donkey female parents showed highly similar coat color characteristics to kulan, and this color was highly consistent within Kulan hybrid population (**Supplementary Figure S1**). There is not precedent for introducing the wild donkey lineage to improve domestic donkeys. In addition, it would be of interesting to assess if Kulan hybrids could produce better meat and skin products than the domestic breeds currently being exploited in China. Therefore, an initial genetic characterization of the Kulan hybrid ass population, as the one here presented, can provide a reference for breeding strategies in this species.

Whole-genome sequencing is an important method to evaluate species population structure and population patterns and identify genome regions related to important economic and environmental adaptation traits. We resequenced the whole genome of Mongolian kulan and 29 Kulan hybrids from Qinghe, Altay region, Xinjiang. The results are expected to fully describe the genomic diversity and population structure of this hybrid population and reveal possible signs of natural or artificial selection. For a meaningful exploration of the formation mechanism of hybrid coat color and other traits, it is necessary to conduct in-depth research.

MATERIALS AND METHODS

Sample Collection and Sequencing

We sampled one Mongolian kulan and 29 Kulan hybrids from the farm of Qinghe Mengyuan Biological Technology Co., LTD. (Qinghe, Altay region, Xinjiang) and extracted genomic DNA from ear tissue by the standard phenol-chloroform protocol (Chen, 1995). Paired-end libraries with an average insert size of 50 bp were constructed for each individual, with an average read length of 150 bp. Whole genome re-sequencing was performed by Illumina HiSeq2000 at Novogene Bioinformatics Institute, Beijing, China.

To determine the ancestry proportion of the Kulan hybrids and their genetic differences to other donkeys (**Supplementary Table S1**), we chose an additional 28 individuals, including Guanzhong donkey (n = 10), Xinjiang donkey (n = 6), Equus africanus (n = 8), and Asiatic wild ass (n = 4).

Alignments and Variant Identification

After performing quality controls, the cleaned reads were aligned against the *Equus asinus* reference assembly EquAsi1.0 (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/016/077/325/GCF_016077325.2_ASM1607732v2/GCF_016077325.2_ASM1607732v2_genomic.fna.gz) using BWA-MEN (Li and Durbin, 2009) with default settings. Duplicate reads were filtered by Picard tools (http://broadinstitute.github.io/picard), and we used the Genome Analysis Toolkit (GATK, version 3.8) (McKenna et al., 2010) to detect single nucleotide polymorphisms (SNPs). After SNP calling, we used the "VariantFiltration" to discard sequencing and alignment artifacts from the SNPs with the parameters"QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0 and SOR >3.0" and mean sequencing depth of variants (all individuals) "<1/3× and >3×". Then, the annotation file of the reference genome (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/016/077/325/GCF_016077325.2_ASM1607732v2/GCF_016077325.2_ASM1607732v2_genomic.gff.gz) was used to obtain functional annotation of the polymorphic sites in all individuals using SNPeff. All the SNPs were counted by effects (downstream_gene_variant, intron_variant, etc) and genomic region (downstream, exon, intron, ect).

Phylogenetic and Population Structure Analyses

Neighbor-joining (NJ) tree, PCA, and ADMIXTURE methods were used to account for the genetic relationships between Mongolian Kulan ass and the Mongolian Kulan hybrid populations. An unrooted NJ tree based on the matrix of pairwise genetic distances from the autosomal SNP data of 58

donkeys was constructed by PLINK v1.9 (Purcell et al., 2007) with the parameter (-indep-pair-wise 50 5 0.2) and visualized by MEGA v7.0 (Kumar et al., 2018) and FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). The SNPs data was pruned by PLINK to perform Principal Component Analysis (PCA) and ADMIXTURE analysis. PCA was constructed using the smartPCA program of EIGENSOFT v5.0 (Patterson et al., 2006), and ADMIXTURE v1.3 (Alexander and Lange, 2011) was used for the construction of the population structure.

Genetic Diversity, Linkage Disequilibrium and Runs of Homozygosity Detection

We calculated the nucleotide diversity for the 58 individuals with 50 kb windows and 50 kb increments using VCFtools (Danecek et al., 2011). PopLDdecay (Zhang C et al., 2019) was used to calculate and visualize the linkage disequilibrium (LD) decay with default parameters. Runs of homozygosity (ROHs) can evaluate the inbreeding degree of the population, and ROHs were identified the --homozyg option implemented in PLINK with the parameter (-homozyg-window-snp 50).

Genome-wide Selective Sweep Analysis

To detect selective sweeps in the studied Kulan hybrids, we used Guanzhong donkeys as a reference population, as Guangzhong donkeys are typical Chinese domestic donkeys (Shen et al., 2021) and their black coats are highly constant in the Guanzhong population. The four analysis methods used are described as follows. Large differences in genetic diversity (θπ-ratio) were calculated with 50 kb sliding windows and 20 kb steps along the autosomes using VCFtools and in-house scripts. Cross Population Extended Haplotype Homozogysity (XP-EHH) based on the extended haplotype statistics for population pairs was conducted by selscan v1.1 with 50 kb windows (Szpiech and Hernandez, 2014). For the XP-EHH selection scan, our test statistic was the average normalized XP-EHH score in each 50 kb region. The composite likelihood ratio (CLR) uses information from allele frequencies to detect selective scans and relies on determining skews in the allele spectrum to bias rare and frequent alleles, and the CLR test was calculated for sites in non-overlapping 50 kb windows by using SweepFinder2 (Vy and Kim, 2015). The θπ test calculates the nucleotide diversity to explain the base diversity of the genome within the population, and we performed SweeD (Pavlidis et al., 2013) for θπ with 50 kb windows. All the SNPs identified after GATK and variant quality control (excluding those located on chromosomes X, Y and mitochondrial DNA) were used to perform a genome-wide selection sweep analysis with the four analysis methods previously described. The top 0.05% candidate windows identified by the different methods were considered as potential candidate selection sweep regions. The detection of selective clearance is to scan the gene sequence regularly through a fixed scale, and the window is the scale which the software scans the genome at a time. Bedtools (Quinlan and Hall, 2010) was used for the annotation of candidate windows. To gain a better understanding of the biological functions and pathways of the identified candidate genes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses was performed using KOBAS 3.0 (Xie et al., 2011). Only FDR (false discovery rate, FDR) corrected p-values < 0.05 will be considered and given in the supplementary table.

In the present study, 58 whole genomes of donkeys were analyzed. The potential signatures of positive selection between the Kulan hybrids and Guanzhong donkey were evaluated by θπ-ratio and XP-EHH, and CLR and θπ revealed the selection pressure within the Kulan hybrids.

RESULTS

Resequencing of the Kulan Hybrids and Variant Identification

A total of 30 resequenced data datasets, including 1 Mongolian kulan and 29 Kulan hybrids (**Supplementary Table S1**), were combined with 28 samples from published data (Africa wild ass = 8, Guanzhong donkey = 10, Xinjiang donkey = 6, Asiatic wild ass = 4) (**Supplementary Table S2**) to form a 58-individual dataset, in which the average depth of the reads was approximately 12.6× coverage. Overall, 6.2 billion clean reads were generated, and the average alignment rate was 99.6% in the 30 individuals (**Supplementary Table S1**). Using GATK, we detected 23520252 biallelic SNPs in all individuals, and then we annotated them (**Supplementary Table S3**). The annotated results indicated 209262 SNPs distributed in exons and 7266346 SNPs distributed in introns.

Phylogenetic Relationship, Principal Component Analysis and Population Structure

Based on the genomic SNPs, we built a neighbor-joining (NJ) tree, principal component analysis and ADMIXTURE (**Figure 1**). In the NJ tree, all the Kulan hybrids formed a separate branch, and the branch of African wild ass was close to Chinese domestic donkeys, which included Xinjiang and Guanzhong donkeys. In addition, the Asiatic wild ass was far from the others (**Figure 1A**). The PCA provided similar results (**Figure 1B**). The first principal component, explaining 10.22% of the total variation, separated Chinese domestic donkeys and African wild ass from Asiatic wild ass and the Kulan hybrids, and the third PC, explaining 4.65% of the total variation, could dissociate Chinese domestic donkeys from African wild ass. In the ADMIXTURE analysis (**Figure 1C**), when K = 2, the African wild ass lineage and Asiatic wild ass lineage could be distinguished. The Asiatic wild ass, African wild ass and Chinese domestic ass can be distinguished at K = 3 (the lowest CV error, **Supplementary Table S4**), while the Asiatic wild ass lineage and the Chinese domestic donkey lineage had similar proportions in the Kulan hybrids. There was no doubt that the Kulan hybrids were influenced by the African wild ass, because the African wild ass was the ancestor of the Chinese domestic donkey (Beja-Pereira et al., 2004; Rossel et al., 2008). When K = 4, Asiatic wild ass, African wild ass, Chinese domestic donkey and the Kulan hybrids were independent of each other.

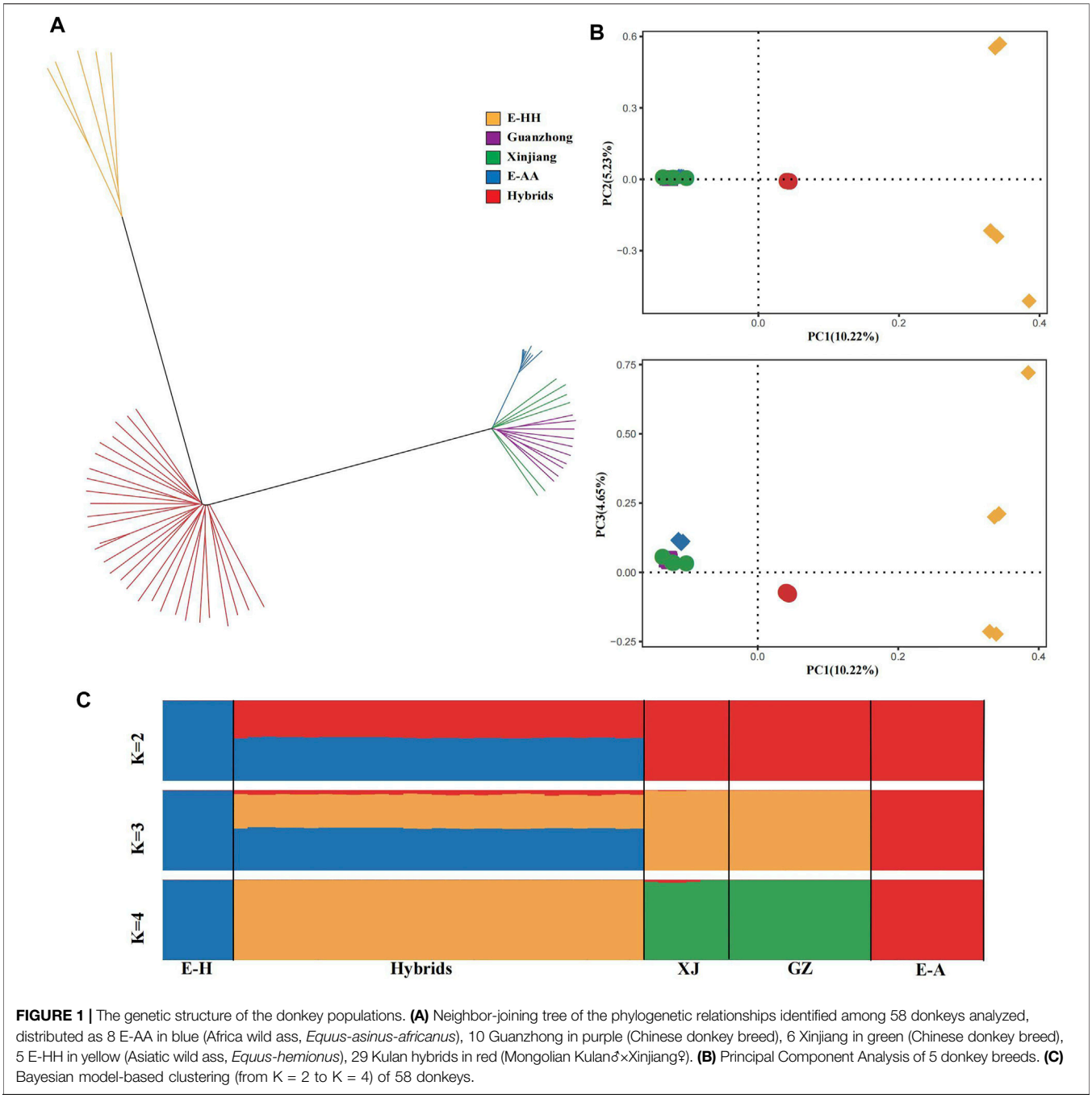

FIGURE 1 | The genetic structure of the donkey populations. **(A)** Neighbor-joining tree of the phylogenetic relationships identified among 58 donkeys analyzed, distributed as 8 E-AA in blue (Africa wild ass, *Equus-asinus-africanus*), 10 Guanzhong in purple (Chinese donkey breed), 6 Xinjiang in green (Chinese donkey breed), 5 E-HH in yellow (Asiatic wild ass, *Equus-hemionus*), 29 Kulan hybrids in red (Mongolian Kulan♂×Xinjiang♀). **(B)** Principal Component Analysis of 5 donkey breeds. **(C)** Bayesian model-based clustering (from K = 2 to K = 4) of 58 donkeys.

Patterns of Genomic Variation

Runs of homozygosity (ROHs) can evaluate the inbreeding degree of a population. We identified the length of ROHs into five types: <0.5 Mb, 0.5–1 Mb, 1–2 Mb, 2–4 Mb, and >4 Mb (**Figure 2A**). Long ROH fragments reflect inbreeding of recent generations, while short ROH fragments indicate inbreeding of distant generations, because the shorter the generation is, the less likely the ROH fragment will be interrupted by recombination. The majority of ROHs we identified in the five breeds were between 0.5–1 Mb. The length of ROHs in African wild ass and the hybrid population was longer than that in the other populations, which suggests that they have inbred in recent generations (**Figure 2B**). For nucleotide diversity, the Kulan hybrids had the highest value, while the African wild ass had the lowest value (**Figure 2C**). Linkage disequilibrium decay indicates that the genetic diversity of hybrid populations is high due to artificial hybridization, while African wild ass has the fastest rate of decay (**Figure 2D**).

Genome-wide Selective Sweep Test

We applied CLR and θπ methods to detect genomic regions related to selection in the Kulan hybrids. The top 0.05% of

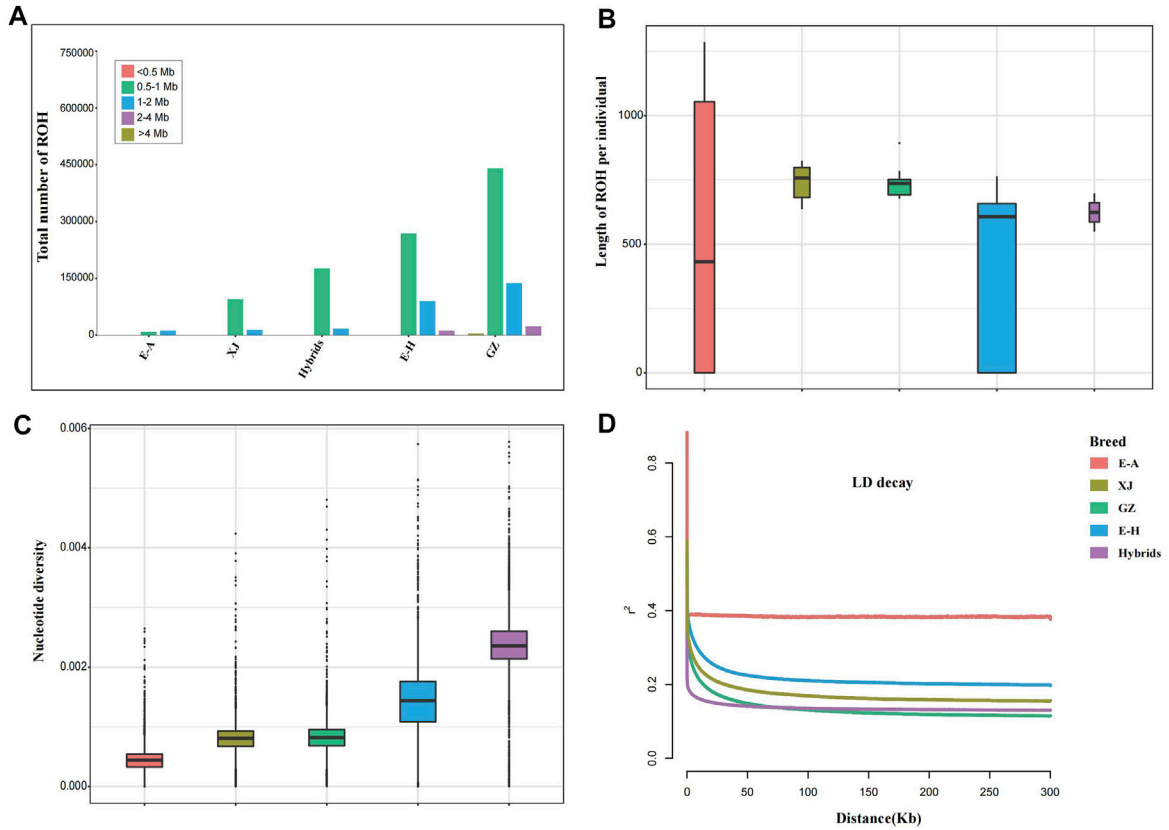

FIGURE 2 | Summary statistics for genomic variation. **(A)** The distribution of the total number of ROHs across chromosomes. **(B)** The distribution of lengths of ROHs in each breed. **(C)** Genome-wide distribution of nucleotide diversity of each breed in 50 kb windows with 50 kb steps. The horizontal line inside the box indicates the median of this distribution; box limits indicate the first and third quartiles; and points show outliers. Data points outside the whiskers can be considered outliers. **(D)** Genome-wide average LD decay estimated from each breed.

windows detected by the two methods were considered candidate regions in the Kulan hybrids. A total of 301 genes were detected by θπ (**Supplementary Table S5**), and 249 genes were detected by CLR (**Supplementary Table S6**). After overlapping, 77 genes shared by the two methods were obtained (**Supplementary Table S7**), which included several significant immune genes (*DEFA1*, *DEFA5*, *DEFA7*, *GIMAP4*, *GIMAP1*, *IGLC1*, and *IGLL5*). Numerous studies have shown that these genes play an integral role in the immune response. The hybrid population may have inherited the strong immunity of the Asiatic wild ass against natural pathogens. After enrichment analysis these 77 genes by KOBAS, we obtained 20 significant KEGG pathways when the corrected p-value < 0.05 (**Supplementary Table S8**).

XP-EHH and θπ-ratio are the methods that detect the selective sweep regions between two populations. With Guanzhong donkey as the reference population, a total of 790 (**Supplementary Table S9**) and 782 genes (**Supplementary Table S10**) were obtained using θπ-ratio and XP-EHH, respectively, while 346 genes were common (**Supplementary Table S11**). We performed functional enrichment to the 346 genes using KOBAS (http://kobas.cbi.pku.edu.cn/) to find vital KEGG pathways, and there were 11 pathways in the enrichment results (**Supplementary Table S12**). "Melanogenesis" pathway is essential for color phenotype in animals, and *KITLG* gene in the pathway has been proved to be associated with the phenotype of horse (Voß et al., 2020), donkey (Wang et al., 2020; Zhou et al., 2020), human (Yang et al., 2018) and dog (Weich et al., 2020).

DISCUSSION

In ancient China, donkey provided an excellent approach to servitude, but with the improvement of agricultural mechanization, the donkey industry has expanded the production of donkey skin and meat (Bennett and Pfuderer, 2020). This newer use of donkey poses a long-term challenge that has led to poor donkey breeding at present.

Understanding the population structure of Kulan hybrids can help to figure out the relatedness among the hybrids and other donkey breeds. Under the circumstance of knowing the origin of the Kulan hybrids, we built the NJ tree and performed PCA to explore the pinpoint of the Kulan hybrids at the population level. There was a clear distinction between the Kulan hybrids and the two wild breeds. Furthermore, the Chinese domestic donkey originated from the African wild ass (Beja-Pereira et al., 2004; Rossel et al., 2008), which caused those individuals to gather in

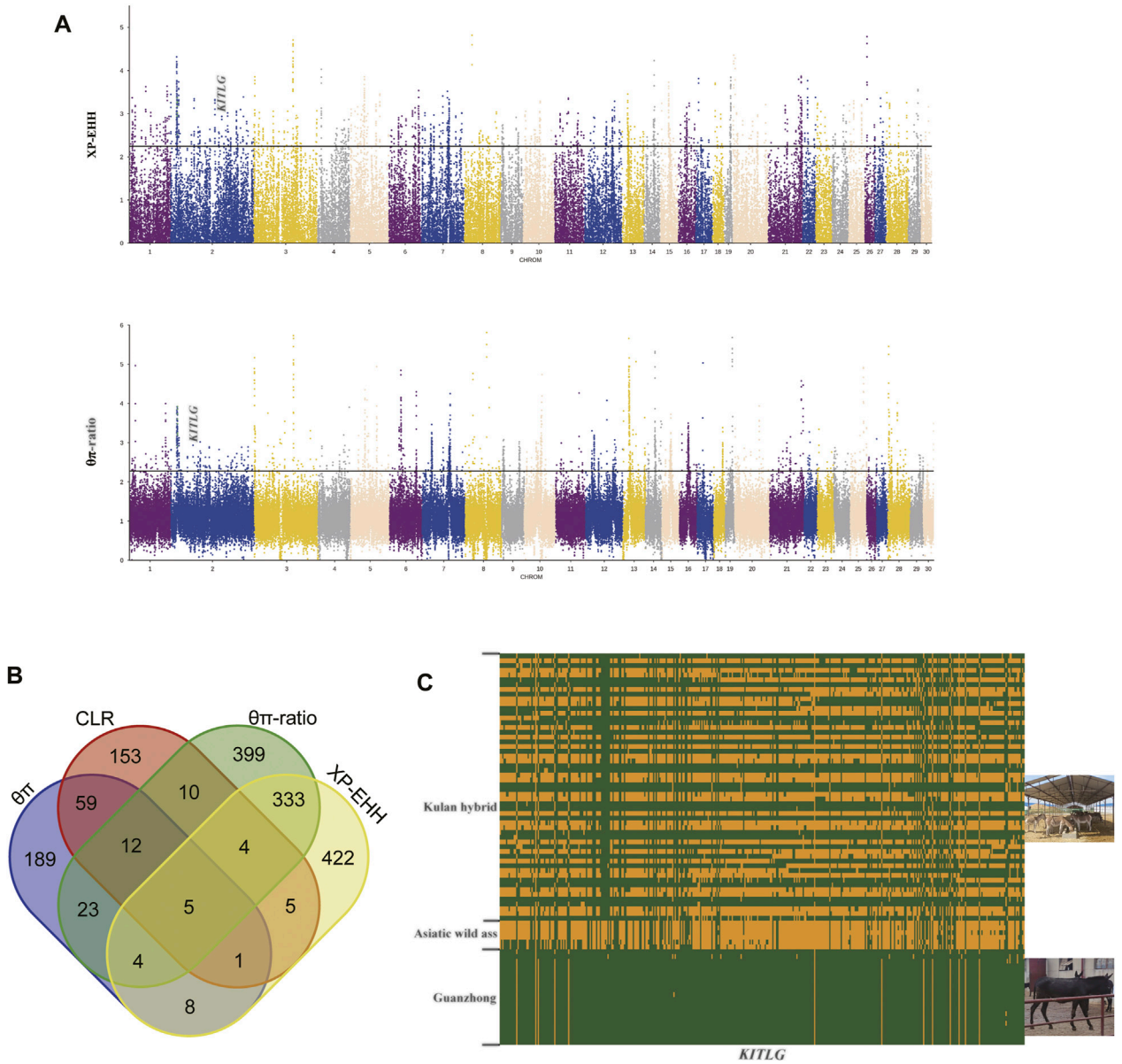

FIGURE 3 | Analysis of the signatures of positive selection in the genome. **(A)** Manhattan plot of selective sweeps. **(B)** Venn diagram showing the genes overlapping among θπ, CLR, θπ-ratio and XP-EHH. **(C)** Linkage disequilibrium analysis of the SNPs in *KITLG* regions.

one branch. The PCA was consistent with the NJ tree, and they both explained the specificality of the Kulan hybrids. The regular ADMIXTURE revealed the ancestral contributions of the Kulan hybrids, and the Asiatic wild ass contributed almost equally to the Kulan hybrids as the Chinese domestic donkey. Remarkably, a small portion of African wild ass SNPs existed in the Kulan hybrids at K = 3 (the lowest CV error), which agreed with the origin of the Chinese domestic donkey. When K = 4, the existence of the African wild ass lineage also revealed the uneven distribution of Xinjiang donkey in the NJ tree.

The length of ROH fragments can infer the history of inbreeding. Long ROH fragments reflect inbreeding in recent generations, while short ROH fragments indicate inbreeding in distant generations, because the shorter the generation is, the lower the possibility that the ROH fragments are interrupted by recombination (Ceballos et al., 2018). Because of the small number of wild donkeys in this study and the great differences among wild donkey subspecies, the results in **Figure 2B** are not satisfactory. Nucleotide diversity is used to describe the strength of polymorphisms in a population (Nei and Li, 1979). The Kulan hybrids had the highest nucleotide diversity and this was significantly higher than other populations with strong linkage to hybridization events. In addition, the patterns of LD were consistent with the nucleotide diversity results.

In the process of natural selection, positive selection usually results in a decrease in genetic polymorphism of the selected sites, while the accumulation of favorable mutations often leads to a hitchhiking effort or a selective sweep (Satta et al., 2018). Here, we used four methods to detect selective sweep regions, including θπ, CLR, XP-EHH and θπ-ratio.

To determine selective sweeps within the hybrid population, CLR and θπ methods were applied, and we observed 77 overlapping genes with both methods (**Figure 3B**). We focused on the function of the 77 genes, and some of them belong to the same family. Defensins are a family of antimicrobial and cytotoxic peptides thought to be involved in host defense (Holly et al., 2017), and we obtained three defensin genes (*DEFA1*, *DEFA5*, and *DEFA7*) that had strong signals of selection. In addition, previous studies have shown that *DEFA1* is crucial for the immunity of horses (Bruhn et al., 2009; Schlusselhuber et al., 2012). *GIMAP4* and *GIMAP1* belong to the GTP-binding superfamily and to the immunoassociated nucleotide (IAN) subfamily of nucleotide-binding proteins, which are most highly expressed in immune system cells and involved in T/B-cell development and survival (Heinonen et al., 2015; Saunders et al., 2010). *IGLC1* and *IGLL5* focused on the transcription of immunoglobulins, which are essential for antigen-specific binding in the process of immunity. For the results of KEGG enrichment analysis, most of these pathways are related to immunity, and we want to highlight *HLA-C*, *HLA-A* and *GZMB*, because these genes were obtained in several immune-related KEGG pathways ("Type I diabetes mellitus", "Allograft rejection", "Graft-versus-host disease", "Autoimmune thyroid disease", "Natural killer cell mediated cytotoxicity"). Human Leukocyte Antigen A (HLA) is also called Major Histocompatibility Complex (HMC), which has been proved to be involved in disease resistance in livestock (Minke et al., 2004; Mikko et al., 1999). *GZMB* encodes a member of the granzyme subfamily of proteins and part of the peptidase S1 family of serine proteases (Turner et al., 2019), and *GZMB* was identified by the four methods simultaneously. *GZMB* is associated with human autoimmune diseases (Xu et al., 2018; Jeong et al., 2021) and essential effector molecules for natural killer (NK)-cell cytotoxicity (Kim et al., 2011). The resistance of cattle to nematode infection and innate immunity requires the participation of *GZMB* (Van Meulder et al., 2013). In xenotransplantation, *GZMB* is often regarded as an important research object (Rodríguez-Gago et al., 2001; Matter-Reissmann et al., 2002). These studies can properly support the key role of *GZMB* in immunity, so the level of immunity of the hybrid population deserves further investigation. Asiatic wild ass has lived in southern Mongolia and northern China for a long time, and these areas have a harsh climate and natural environment. Without good artificial care, Asiatic wild asses struggle with various diseases and emergencies. There is a broad consensus that heterosis is an important method of animal genetic improvement, so Kulan hybrids may inherit strong immunity through hybridization. As a significant economic trait of livestock, disease resistance is considered the focus of breeding efforts. Therefore, wild ass may be a crucial genetic resource for improving the disease resistance of domestic donkeys.

Because the Kulan hybrids that come from Mongolian kulan and Chinese domestic donkeys, we chose Guanzhong donkey as the reference population, which is the representative breed of Chinese domestic donkeys (Shen et al., 2021). To understand the genetic differences between the Kulan hybrids and Guanzhong donkey, θπ-ratio and XP-EHH were performed to find the genomic traces left by selection. We performed enrichment analysis to the 346 overlapping genes (**Supplementary Table S11**), then we obtained 11 KEGG pathways when corrected p-values < 0.05(**Supplementary Table S12**). In the results, we would like to highlighted several vital terms identified by the analysis, which were "Calcium signaling pathway", "Regulation of actin cytoskeleton", "Melanogenesis". Among these terms, in relation to the phenotypes for which the two groups of animals compared in our pair-wise analyses differ, the Kulan hybrids and Guanzhong donkey, we would highlight the "Melanogenesis" pathway term (*WNT2B*, *GNAO1*, *KITLG*, *Camk2a*, *WNT6*, *DVL3*). Coat color is the most visualized feature of Asiatic wild ass and its Kulan hybrids. The coat color of the Kulan hybrids is earthy yellow while the Guanzhong donkey is black. Compared with Guanzhong donkey, *KITLG* was found in the top 0.5% of regions identified by both methods, then we built a heatmap of haplotypes to verify its true distribution. It was obvious that the Kulan hybrids and other populations had different patterns (**Figure 3C**). The *KITLG* gene is located on chromosome 2 of donkey, and *KITLG* have been proven to be involved in the formation of coat color in several species (Anello et al., 2019; Voß et al., 2020; Weich et al., 2020; Wen et al., 2021; Wu et al., 2019; Yang et al., 2018). Furthermore, there have been some studies on donkey and horse (Voß et al., 2020; Zhou et al., 2020). Therefore, our findings may provide new insights into the coat color formation of Asiatic wild ass and its Kulan hybrids. Besides, the "calcium signaling pathway" (*PDGFRB*, *Atp2b2*, *ATP2A3*, *PDE1B*, *Lap3*, *MYLK2*, *Camk2a*, *P2RX1*, *ATP2B2*, *P2RX5*) and "Regulation of actin cytoskeleton" (*PDGFRB*, *NCKAP1L*, *SOS1*, *MOS*, *Arhgef7*, *Lap3*, *MYLK2*, *ITGA5*, *ITGAE*) were also identified as the relevant terms based on the results of the KEGG enrichment analyses. Interestingly, the "calcium signalling pathway" and"Regulation of actin cytoskeleton" were related to meat quality (Cao et al., 2019; Malila et al., 2020; Xia et al., 2021) and from the mentioned genes we would highlight the association previously reported for *PDE1B* and *MYLK2*. Haplotypes based on the SNPs derived from a bacterial artificial chromosome containing the bovine gene *PDE1B* were associated with traits related to carcass fat (Stone et al., 2005) and the quantitative trait loci for fat deposition and carcass traits have been identified in the vicinity of the gene encoding *PDE1B* (Shin et al., 2012). *MYLK2* gene was associated with myosin components and may influence muscle development of early porcine embryos through miRNA regulatory network (Zhang X et al., 2019). Heterosis is common in animal breeding, and the Kulan hybrids may be able to provide higher quality meat than the Guanzhong donkey in point of genome-wide selective sweep test, but further research is needed to confirm it. Because of the large wide range of analyses presented here to assess the genetic diversity of the considered populations, and considering that this is not a deep study of selection sweep mapping, the results of the selection sweep mapping regions were assessed only through the enrichment analyses that have been discussed. However, it is important to clarify that some of the selection sweeps may be related to target selection genes that have not been highlighted by the global enrichment analyses. This is the case of the *LCORL* gene which was identified by both, the θπ-ratio and XP-EHH methods, as a

candidate gene for a selection sweep located on chromosome 3. This gene is known gene to regulate the body size of livestock species (Makvandi-Nejad et al., 2012; Takasuga, 2016), and previous studies have clearly shown the association between *LCORL* and donkey body size (Shen et al., 2021). Xinjiang donkey is a typical small donkey while Guanzhong donkey is a large breed, and the body size of Kulan hybrids is similar with Xinjiang donkey. Therefore, the identification of this selection sweeps confirms previous studies on the association of *LCORL* and body size. In any case, as previously said, further deeper analyses could be done on the selection sweeps reported here with the objective of identifying potential causal gene and mutations that could directly explain the selection signals here described.

Mongolian Kulan is an essential part of Asiatic wild ass, but hunting and deteriorating living conditions have caused their numbers to plummet leading them to nearly the level of a threatened species in the International Union for Conservation of Nature Red list. As donkey genetic resources are being exhausted, breeding planning would help improve and conserve Chinese native donkeys. For this purpose, our results provide a basis for exploring the genomic characteristics of Asiatic wild ass and its Kulan hybrids in relation to crucial economic traits.

CONCLUSION

Based on whole genome re-sequencing data, we provide an overview of Asiatic wild ass and its hybrids. Population genetic structure and genomic diversity analysis identified a new direction for the genetic improvement of Chinese domestic donkeys. Furthermore, we identified genes related to immunity, coat color and meat quality in the Kulan hybrids and the results may serve as valuable resources for research into disease-resistance breeding and the mechanism of animal coat color formation.

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 Institutional Animal Care and Use Committee of Northwest A & F University.

AUTHOR CONTRIBUTIONS

HD and JC conceived and designed the experiments. HD and ZD contributed equally to the construction and execution of this manuscript. FW and GW assisted in the data analysis of this study. XL performed the sample DNA extraction and data upload. CL provided laboratory instruments and equipment. JC and HD provided the funding for the research. All authors read and approved the final manuscript.

FUNDING

The present study is supported by Application of intelligent Information Collection technology in Breeding and Training of Thoroughbred Horses (Project Number: 2016E02035); Demonstration and promotion of efficient breeding technology of donkeys (Project Number: 2019E0264); Research and Demonstration of key Technologies for efficient Breeding of Donkeys in Key RESEARCH and Development projects of South Xinjiang Donkey Industry in Xinjiang Uygur Autonomous Region (Project Number: 2020B01002-2-1).

ACKNOWLEDGMENTS

We thank the High-Performance Computing of Northwest A&F University for providing computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.818420/full#supplementary-material

REFERENCES

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinforma.* 12, 246. doi:10.1186/1471-2105-12-246

Anello, M., Daverio, M. S., Silbestro, M. B., Vidal-Rioja, L., and Di Rocco, F. (2019). Characterization and Expression Analysis of KIT and MITF -M Genes in Llamas and Their Relation to White Coat Color. *Anim. Genet.* 50 (2), 143–149. doi:10.1111/age.12769

Beja-Pereira, A., England, P. R., Ferrand, N., Jordan, S., Bakhiet, A. O., Abdalla, M. A., et al. (2004). African Origins of the Domestic Donkey. *Science* 304 (5678), 1781. doi:10.1126/science.1096008

Bennett, R., and Pfuderer, S. (2020). The Potential for New Donkey Farming Systems to Supply the Growing Demand for Hides. *Animals* 10 (4), 718. doi:10.3390/ani10040718

Bruhn, O., Paul, S., Tetens, J., and Thaller, G. (2009). The Repertoire of Equine Intestinal α-defensins. *BMC Genomics* 10, 631. doi:10.1186/1471-2164-10-631

Cao, X.-K., Cheng, J., Huang, Y.-Z., Wang, X.-G., Ma, Y.-L., Peng, S.-J., et al. (2019). Growth Performance and Meat Quality Evaluations in Three-Way Cross Cattle Developed for the Tibetan Plateau and Their Molecular Understanding by Integrative Omics Analysis. *J. Agric. Food Chem.* 67 (1), 541–550. doi:10.1021/acs.jafc.8b05477

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, H. (2011). *Animal Genetic Resources in China-Horses, Donkeys, and Camels*. Beijing: China Agricultural Press. Reprinted.

Chen, H., and Leibenguth, F. (1995). Studies on Multilocus Fingerprints, RAPD Markers, and Mitochondrial DNA of a Gynogenetic Fish (Carassius Auratus Gibelio). *Biochem. Genet.* 33, 297–306. doi:10.1007/BF02399929

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

Gao, J., Lyu, Y., Zhang, D., Reddi, K. K., Sun, F., Yi, J., et al. (2020). Genomic Characteristics and Selection Signatures in Indigenous Chongming White Goat (*Capra hircus*). *Front. Genet.* 11, 901. doi:10.3389/fgene.2020.00901

Gerritsmann, H., Stalder, G. L., Kaczensky, P., Buuveibaatar, B., Payne, J., Boldbaatar, S., et al. (2016). Arterial pH and Blood Lactate Levels of Anesthetized Mongolian Khulan (*Equus hemionus* Hemionus) in the Mongolian Gobi Correlate with Induction Time. *J. Wildl. Dis.* 52, 642–646. Reprinted. doi:10.7589/2015-07-198

Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A., and Kingsley, D. M. (2014). A Molecular Basis for Classic Blond Hair Color in Europeans. *Nat. Genet.* 46 (7), 748–752. doi:10.1038/ng.2991

Heinonen, M. T., Kanduri, K., Lähdesmäki, H. J., Lahesmaa, R., and Henttinen, T. A. (2015). Tubulin- and Actin-associating GIMAP4 Is Required for IFN-γ Secretion during Th Cell Differentiation. *Immunol. Cell Biol.* 93 (2), 158–166. doi:10.1038/icb.2014.86

Holly, M. K., Diaz, K., and Smith, J. G. (2017). Defensins in Viral Infection and Pathogenesis. *Annu. Rev. Virol.* 4 (1), 369–391. doi:10.1146/annurev-virology-101416-041734

Jeong, K.-H., Kim, S. K., Seo, J.-K., Shin, M. K., and Lee, M.-H. (2021). Association of GZMB Polymorphisms and Susceptibility to Non-segmental Vitiligo in a Korean Population. *Sci. Rep.* 11 (1), 397. doi:10.1038/s41598-020-79705-0

Kim, T., Lee, S. U., Yun, S., Sun, H., Lee, S. H., and Kim, J. W. (2011). Human MicroRNA-27a* Targets Prf1 and GzmB Expression to Regulate NK-Cell Cytotoxicity. *Blood* 118, 5476–5486. doi:10.1182/blood-2011-04-347526

Klungland, H., and Våge, D. I. (2003). Pigmentary Switches in Domestic Animal Species. *Ann. Ny. Acad. Sci.* 994, 331–338. doi:10.1111/j.1749-6632.2003.tb03197.x

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. Reprinted. doi:10.1093/molbev/msy096

Lei, C. Z., Chen, H., Yang, G. S., Sun, W. B., Lei, X. Q., Ge, Q. L., et al. (2005). Study on Mitochondrial DNA D-Loop Polymorphism in Chinese Donkeys. *J Genet Genom* 32, 481–486. in chinese http://europepmc.org/abstract/MED/16018258.

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

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., et al. (2012). Four Loci Explain 83% of Size Variation in the Horse. *PLoS One* 7 (7), e39929. doi:10.1371/journal.pone.0039929

Malila, Y., Uengwetwanit, T., Arayamethakorn, S., Srimarut, Y., Thanatsang, K. V., Soglia, F., et al. (2020). Transcriptional Profiles of Skeletal Muscle Associated with Increasing Severity of White Striping in Commercial Broilers. *Front. Physiol.* 11, 580. doi:10.3389/fphys.2020.00580

Matter-Reissmann, U. B., Forte, P., Schneider, M. K. J., Filgueira, L., Groscurth, P., and Seebach, J. D. (2002). Xenogeneic Human NK Cytotoxicity against Porcine Endothelial Cells Is Perforin/granzyme B Dependent and Not Inhibited by Bcl-2 Overexpression. *Xenotransplantation* 9 (5), 325–337. doi:10.1034/j.1399-3089.2002.01074.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

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., et al. (2015). Runs of Homozygosity Reveal Signatures of Positive Selection for Reproduction Traits in Breed and Non-breed Horses. *BMC Genomics* 16, 764. doi:10.1186/s12864-015-1977-3

Mikko, S., Røed, K., Schmutz, S., and Andersson, L. (1999). Monomorphism and Polymorphism at Mhc DRB Loci in Domestic and Wild Ruminants. *Immunol. Rev.* 167, 169–178. doi:10.1111/j.1600-065x.1999.tb01390.x

Minke, J. M., Audonnet, J.-C., and Fischer, L. (2004). Equine Viral Vaccines: The Past, Present and Future. *Vet. Res.* 35 (4), 425–443. doi:10.1051/vetres:2004019

Nei, M., and Li, W. H. (1979). Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* 76 (10), 5269–5273. doi:10.1073/pnas.76.10.5269

Oakenfull, E. N., Lim, H. N., and Ryder, O. A. (2000). A Survey of Equid Mitochondrial DNA: Implications for the Evolution, Genetic Diversity and Conservation of Equus. *Conserv. Genet.* 1 (4), 341–355. doi:10.1023/A:1011559200897

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genet.* 2 (12), e190. doi:10.1371/journal.pgen.0020190

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol. Biol. Evol.* 30 (9), 2224–2234. doi:10.1093/molbev/mst112

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 (3), 559–575. doi:10.1086/519795

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

Richard, P. R., Henry, M. M., Badamjaviin, L., Claudia, F., David, P. K., Dulamtseren, S., et al. (2001). Status and Distribution of Khulan (*Equus hemionus*) in Mongolia. *J. Zool.* 254 (3), 381–389. doi:10.1017/S0952836901000887

Rodríguez-Gago, M., de Heredia, A., Ramírez, P., Parrilla, P., Aparicio, P., and Yélamos, J. (2001). Human Anti-porcine Gammadelta T-Cell Xenoreactivity Is Inhibited by Human FasL Expression on Porcine Endothelial Cells. *Transplantation* 72 (3), 503–509. doi:10.1097/00007890-200108150-00024

Rosenbom, S., Costa, V., Al-Araimi, N., Kefena, E., Abdel-Moneim, A. S., Abdalla, M. A., et al. (2015). Genetic Diversity of Donkey Populations from the Putative Centers of Domestication. *Anim. Genet.* 46 (1), 30–36. doi:10.1111/age.12256

Rossel, S., Marshall, F., Peters, J., Pilgram, T., Adams, M. D., and O'Connor, D. (2008). Domestication of the Donkey: Timing, Processes, and Indicators. *Proc. Natl. Acad. Sci. U.S.A.* 105 (10), 3715–3720. doi:10.1073/pnas.0709692105

Satta, Y., Fujito, N. T., and Takahata, N. (2018). Nonequilibrium Neutral Theory for Hitchhikers. *Mol. Biol. Evol.* 35 (6), 1362–1365. doi:10.1093/molbev/msy093

Saunders, A., Webb, L. M. C., Janas, M. L., Hutchings, A., Pascall, J., Carter, C., et al. (2010). Putative GTPase GIMAP1 Is Critical for the Development of Mature B and T Lymphocytes. *Blood* 115 (16), 3249–3257. doi:10.1182/blood-2009-08-237586

Schlusselhuber, M., Jung, S., Bruhn, O., Goux, D., Leippe, M., Leclercq, R., et al. (2012). *In Vitro* Potential of Equine DEFA1 and eCATH1 as Alternative Antimicrobial Drugs in Rhodococcosis Treatment. *Antimicrob. Agents Chemother.* 56 (4), 1749–1755. doi:10.1128/AAC.05797-11

Shen, J., Yu, J., Dai, X., Li, M., Wang, G., Chen, N., et al. (2021). Genomic Analyses Reveal Distinct Genetic Architectures and Selective Pressures in Chinese Donkeys. *J. Genet. Genomics* 48 (8), 737–745. doi:10.1016/j.jgg.2021.05.012

Shin, S., Heo, J., Yeo, J., Lee, C., and Chung, E. (2012). Genetic Association of Phosphodiesterase 1B (PDE1B) with Carcass Traits in Korean Cattle. *Mol. Biol. Rep.* 39 (4), 4869–4874. doi:10.1007/s11033-011-1280-6

Stone, R. T., Casas, E., Smith, T. P. L., Keele, J. W., Harhay, G., Bennett, G. L., et al. (2005). Identification of Genetic Markers for Fat Deposition and Meat Tenderness on Bovine Chromosome 5: Development of a Low-Density Single Nucleotide Polymorphism Map. *J. Anim. Sci.* 83, 2280–2288. Reprinted. doi:10.2527/2005.83102280x

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., et al. (2007). Genetic Determinants of Hair, Eye and Skin Pigmentation in Europeans. *Nat. Genet.* 39 (12), 1443–1452. doi:10.1038/ng.2007.13

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31, 2824–2827. Reprinted. doi:10.1093/molbev/msu211

Takasuga, A. (2016). PLAG1 and NCAPG-LCORL in Livestock. *Anim. Sci. J.* 87 (2), 159–167. doi:10.1111/asj.12417

Turner, C. T., Lim, D., and Granville, D. J. (2019). Granzyme B in Skin Inflammation and Disease. *Matrix Biol.* 75-76, 126–140. doi:10.1016/j.matbio.2017.12.005

Van Meulder, F., Van Coppernolle, S., Borloo, J., Rinaldi, M., Li, R. W., Chiers, K., et al. (2013). Granule Exocytosis of Granulysin and Granzyme B as a Potential Key Mechanism in Vaccine-Induced Immunity in Cattle against the Nematode Ostertagia Ostertagi. *Infect. Immun.* 81 (5), 1798–1809. doi:10.1128/IAI.01298-12

Voß, K., Tetens, J., Thaller, G., and Becker, D. (2020). Coat Color Roan Shows Association with KIT Variants and No Evidence of Lethality in Icelandic Horses. *Genes* 11 (6), 680. doi:10.3390/genes11060680

Vy, H. M. T., and Kim, Y. (2015). A Composite-Likelihood Method for Detecting Incomplete Selective Sweep from Population Genomic Data. *Genetics* 200 (2), 633–649. doi:10.1534/genetics.115.175380

Wang, C., Li, H., Guo, Y., Huang, J., Sun, Y., Min, J., et al. (2020). Donkey Genomes Provide New Insights into Domestication and Selection for Coat Color. *Nat. Commun.* 11 (1), 6014. doi:10.1038/s41467-020-19813-7

Weich, K., Affolter, V., York, D., Rebhun, R., Grahn, R., Kallenberg, A., et al. (2020). Pigment Intensity in Dogs Is Associated with a Copy Number Variant Upstream of KITLG. *Genes* 11 (1), 75. doi:10.3390/genes11010075

Wen, J., Shao, P., Chen, Y., Wang, L., Lv, X., Yang, W., et al. (2021). Genomic Scan Revealed KIT Gene Underlying White/gray Plumage Color in Chinese Domestic Geese. *Anim. Genet.* 52 (3), 356–360. doi:10.1111/age.13050

Wu, Z., Deng, Z., Huang, M., Hou, Y., Zhang, H., Chen, H., et al. (2019). Whole-Genome Resequencing Identifies KIT New Alleles that Affect Coat Color Phenotypes in Pigs. *Front. Genet.* 10, 218. doi:10.3389/fgene.2019.00218

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22 (1), 43. doi:10.1186/s12864-020-07340-0

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: A Web Server for Annotation and Identification of Enriched Pathways and Diseases. *Nucleic Acids Res.* 39 (Web Server issue), W316–W322. doi:10.1093/nar/gkr483

Xu, M., Liu, Y., Liu, Y., Li, X., Chen, G., Dong, W., et al. (2018). Genetic Polymorphisms of GZMB and Vitiligo: A Genetic Association Study Based on Chinese Han Population. *Sci. Rep.* 8 (1), 13001. doi:10.1038/s41598-018-31233-8

Yang, Z., Shi, H., Ma, P., Zhao, S., Kong, Q., Bian, T., et al. (2018). Darwinian Positive Selection on the Pleiotropic Effects of KITLG Explain Skin Pigmentation and Winter Temperature Adaptation in Eurasians. *Mol. Biol. Evol.* 35 (9), 2272–2283. doi:10.1093/molbev/msy136

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinforma. Oxf. Engl.* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, J., Liu, F., Cao, J., and Liu, X. (2015). Skin Transcriptome Profiles Associated with Skin Color in Chickens. *PLoS One* 10 (6), e0127301. doi:10.1371/journal.pone.0127301

Zhang, X. C., Shao, C. L., Ge, Y., Chen, C., Xu, W. X., and Yang, W. K. (2020). Suitable Summer Habitat of the Khulan in the Mt. Kalamaili Ungulate Nature Reserve and Estimation of its Population. *Ying Yong Sheng Tai Xue Bao* 31 (9), 2993–3004. doi:10.13287/j.1001-9332.202009.032

Zhang, X., Cai, S., Chen, L., Yuan, R., Nie, Y., Ding, S., et al. (2019). Integrated miRNA-mRNA Transcriptomic Analysis Reveals Epigenetic-Mediated Embryonic Muscle Growth Differences between Wuzhishan and Landrace Pigs1. *J. Anim. Sci.* 97 (5), 1967–1978. doi:10.1093/jas/skz091

Zhou, Z., Fan, Y., Wang, G., Lai, Z., Gao, Y., Wu, F., et al. (2020). Detection of Selection Signatures Underlying Production and Adaptive Traits Based on Whole-Genome Sequencing of Six Donkey Populations. *Animals* 10 (10), 1823. doi:10.3390/ani10101823

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: 27 May 2022
doi: 10.3389/fgene.2022.816379

Whole-Genome Resequencing of Xiangxi Cattle Identifies Genomic Diversity and Selection Signatures

*Xiaoyu Luo[1,2†], Jianbo Li[1,3†], Chentong Xiao[2], Luyang Sun[2], Weixuan Xiang[2,4], Ningbo Chen[2], Chuzhao Lei[2], Hong Lei[1,3], Yun Long[3,5], Ting Long[3,5], Quji Suolang[6] and Kangle Yi[1,3]**

[1]Hunan Institute of Animal and Veterinary Science, Changsha, China, [2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Xianyang, China, [3]Xiangxi Cattle Engineering Technology Center of Hunan Province, Huayuan, China, [4]School of Life Science, University of Bristol, Bristol, United Kingdom, [5]Hunan De Nong Animal Husbandry Group Co. Ltd., Huayuan, China, [6]Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Science, Lhasa, China

Understanding the genetic diversity in Xiangxi cattle may facilitate our efforts toward further breeding programs. Here we compared 23 Xiangxi cattle with 78 published genomes of 6 worldwide representative breeds to characterize the genomic variations of Xiangxi cattle. Based on clustering models in population structure analysis, we displayed that Xiangxi cattle had a mutual genome ancestor with Chinese indicine, Indian indicine, and East Asian taurine. Population genetic diversity was analyzed by four methods (nucleotide diversity, inbreeding coefficient, linkage disequilibrium decay and runs of homozygosity), and we found that Xiangxi cattle had higher genomic diversity and weaker artificial selection than commercial breed cattle. Using four testing methods ($\theta\pi$, CLR, F_{ST}, and XP-EHH), we explored positive selection regions harboring genes in Xiangxi cattle, which were related to reproduction, growth, meat quality, heat tolerance, and immune response. Our findings revealed the extent of sequence variation in Xiangxi cattle at the genome-wide level. All of our fruitful results can bring about a valuable genomic resource for genetic studies and breed protection in the future.

Keywords: Xiangxi cattle, genetic diversity, selection signatures, DNAJC8, whole genome analysis

OPEN ACCESS

Edited by:
El Hamidi Hay, United States Department of Agriculture, United States

Reviewed by:
Xiangdong Ding, China Agricultural University, China
Kwan-Suk Kim, Chungbuk National University, South Korea

****Correspondence:***
Kangle Yi
yikangle@yeah.net

†These authors have contributed equally to this work and share first authorship

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

Received: *16 November 2021*
Accepted: *13 April 2022*
Published: *27 May 2022*

Citation:
Luo X, Li J, Xiao C, Sun L, Xiang W, Chen N, Lei C, Lei H, Long Y, Long T, Suolang Q and Yi K (2022) Whole-Genome Resequencing of Xiangxi Cattle Identifies Genomic Diversity and Selection Signatures. Front. Genet. 13:816379. doi: 10.3389/fgene.2022.816379

INTRODUCTION

One of the most economically important breeds of livestock is domestic cattle, providing human beings with basic resources, such as milk, beef, draft energy, and so on. They can be artificially classified into two subspecies: humpless taurine (*Bos taurus*) and humped indicine (*Bos indicus*) (Loftus et al., 1994), which are hybrid with each other. Recently, the research shows that domestic cattle can be categorized into five groups worldwide: European taurine, Eurasian taurine, East Asian taurine, Chinese indicine, and Indian indicine (Chen et al., 2018). As a significant part of cattle variety resources, Chinese native cattle have abundant genetic diversity. Xiangxi cattle, a representative breed of cattle in southwestern China, is one of the breeds recognized by China and renowned abroad, which belongs to the beef-serving type (**Supplementary Note**). Xiangxi cattle is highly valued because of its strong immunity, excellent meat quality, high service endurance and heat resistance capacity (Wang et al., 2012; Long and Ma, 2009; Zhang et al., 2011). Previous research on genetic diversity, population polymorphism, and hybridization adopted methods, for instance, autosomal, Y

chromosome or mitochondrial data (Cai et al., 2019; Luoreng et al., 2014; Yao et al., 2017) to demonstrate that the Xiangxi cattle originated from *Bos taurus* X *Bos indicus*.

With the application of whole-genome sequencing, it has become possible to study the genetic diversity of various breeds on the group size (Daetwyler et al., 2014). Since the implementation of the "The 1,000 Bull Genomes" project, whole-genome sequencing has been used to investigate the world's commercial breeds (Daetwyler et al., 2014; Stothard et al., 2011) and indigenous breeds (Kim et al., 2017; Lee et al., 2013), providing precious genome resources for future molecular breeding and genetic improvement. As a hot research methodology, the whole-genome sequencing has been progressively applied to Chinese native cattle such as Qinchuan cattle, Jiaxian Red cattle, and Yanbian cattle (Choi et al., 2015; Mei et al., 2018; Xia et al., 2021).

Considering that the entire genome variation of Xiangxi cattle was largely unexplored, we performed whole-genome sequencing on 23 Xiangxi cattle. SNPs were identified in Xiangxi cattle by basing on *Bos taurus* reference genome assembly (ARS-UCD1.2). Then SNPs of Xiangxi cattle made a comparison with those of commercial and native breeds previously collected from around the world. Finally, by scanning the whole genome of Xiangxi cattle, the selective sweeping results were determined. Our analyses fully described the genomic diversity and population structure, revealed the possible signs of natural and artificial selection, as well as provided new insights for breeding Xiangxi cattle.

METHODS

Samples, DNA Extraction, and Sequencing

Ear tissue samples of Xiangxi cattle ($n = 23$) were collected from the Xiangxi cattle engineering technology center of Hunan Province, Huayuan, China. Genomic DNA of the ear tissue samples was extracted using a standard phenol/chloroform-based protocol. The DNA library was constructed for each sample (500 bp insert size). Sequencing via Illumina NovaSeq 6000 with 2 × 150 bp model at Novogene Bioinformatics Institute, Beijing, China, and 150 bp paired-end sequence data were generated.

Furthermore, we obtained sequence data of 78 cattle, including 25 European cattle (Hereford-14, Jersey-11), 15 Hanwoo, 31 Chinese native cattle (Wannan-5, Guangfeng-4, Ji'an-4, Leiqiong-3, Jinjiang-3, Qinchuan-12), and 7 India-Pakistan cattle (Tharparkar-1, Sahiwal-1, Hariana-1, Nelore-1, Gir-2 and unknown-1) publicly. In total, 101 whole genomes of cattle were used for the subsequent analysis.

Reads Mapping and SNP Calling

The clean reads were mapped onto the *Bos taurus* reference genome assembly ARS-UCD1.2 using BWA-MEM (0.7.13-r1126) with default parameters (Li and Durbin, 2009). After mapping, the single nucleotide polymorphisms (SNPs) were detected by using Samtools (Li et al., 2009), Picard tools (http://broadinstitute.github.io/picard), and Genome Analysis Toolkit (GATK, version 3.6–0-g89b7209). All SNPs were filtered using the module "Variant Filtration" of GATK to obtain high-quality SNPs. By using the ANNOVAR(Wang et al., 2010), SNPs were annotated based on the latest reference assembly (ARS-UCD1.2).

Population Genetics Analysis

After pruning in PLINK with the parameter (--indep-pair-wise 50 5 0.2), a set of SNPs were generated for the following analyses. An unrooted neighbor-joining (NJ) was constructed based on the matrix of pairwise genetic distances using MEGA v7.0 (Kumar et al., 2016) and iTOL v5 (Letunic and Bork, 2021). Principal component analysis (PCA) was performed using the smartPCA of the EIGENSOFT v5.0 package (Patterson et al., 2006). Population structure analysis was assessed with genetic clusters K ranging from 2 to 7 using the ADMIXTURE v1.3 (Alexander and Large, 2011).

Linkage disequilibrium (LD) decay was calculated using PopLDdecay (Zhang et al., 2019) with default parameters. By using the VCFtools (Danecek et al., 2011), we respectively calculated the inbreeding coefficient (--het) and the Nucleotide diversity (π). Additionally, we identified the runs of homozygosity (ROHs) using the--homozyg option implemented in the PLINK. The number and length of ROH for each breed were calculated and divided into four categories (0.5–1 Mb, 1–2 Mb, 2–4 Mb, >4 Mb). The plot as mentioned above was depicted using the R script (http://www.r-project.org).

Selective Sweep Test

To identify the selective sweep regions in Xiangxi cattle, two different statistics were used, including the nucleotide diversity ($\theta\pi$) with 50 kb sliding window and 20 kb step in VCFtools (Danecek et al., 2011) and the composite likelihood ratio (CLR) with 50 kb windows in SweepFinder2 (DeGiorgio et al., 2016). We took the top 1% of the overlapping genes obtained by the two above methods.

For comparison in Xiangxi cattle and Qinchuan cattle, we separately calculated fixation index (F_{ST}) and cross-population extended haplotype homozygosity (XP-EHH). We estimated the F_{ST} values using VCFtools (50 kb windows with 20 kb step) and the XP-EHH using selscan v1.1 (Szpiech et al., 2014) with default settings. It is noteworthy that the normalized XP-EHH score was tested about each 50-kb region in the XP-EHH statistic. Significant genomic regions were identified ($p < 0.005$). In order to make the results more reliable, we used two or more methods to determine that the overlapping regions had been used as candidate regions. We calculated the Tajima's D statistic to consolidate our results by using VCFtools.

Moreover, KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/) was used to understand the function and complex pathways of candidate genes, including the Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Gene Ontology (GO) in the study (corrected p-value < 0.05).

RESULTS

Sequencing and Identification of Single Nucleotide Polymorphisms

Using BWA-MEM software, the clean reads were aligned with the *B. taurus* reference genome (ARS-UCD1.2), resulting in a total of

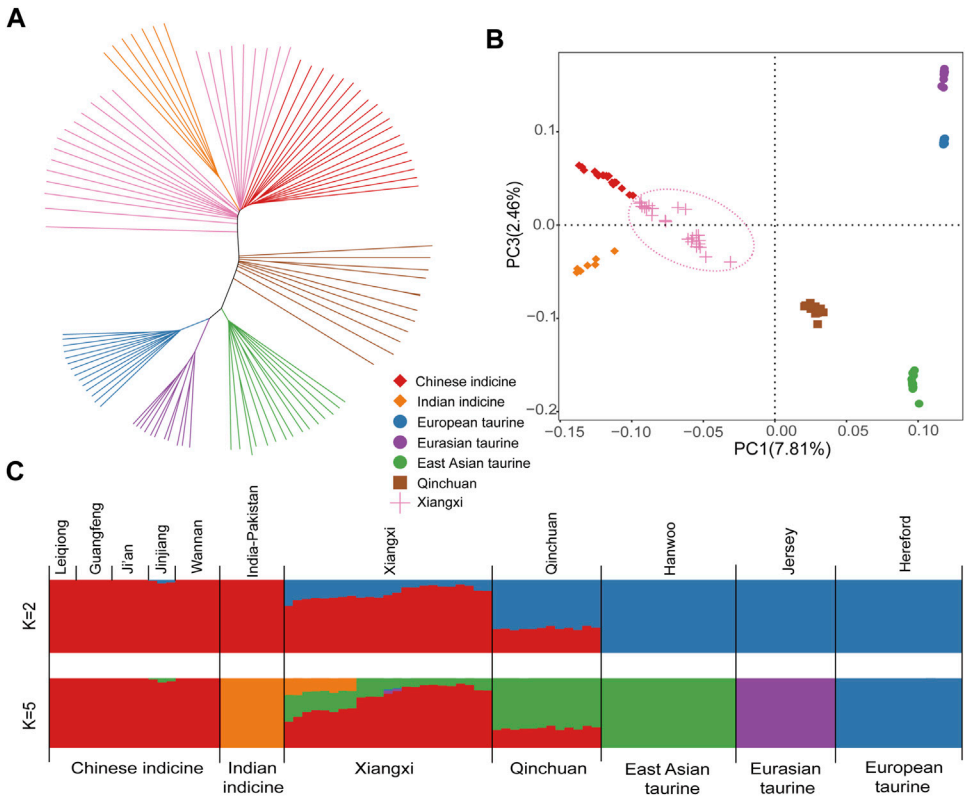

FIGURE 1 | Population genetic analysis of Xiangxi cattle. **(A)** The Neighbor-joining tree of relationships among 7 populations **(B)** The principal component analysis of cattle with PC1 against PC3 **(C)** genetic structure of cattle using ADMIXTURE with K = 2 and K = 5.

~4.7 billion reads and ~10.9 × coverage each (**Supplementary Table S1**). Xiangxi cattle were jointly analyzed with the public genomic data of five "core" populations (Chen et al., 2018; Xia et al., 2021)and Qinchuan cattle (native cattle in North-central China). Five "core" populations: European taurine (Hereford), Eurasian taurine (Jersey), East Asian taurine (Hanwoo), Chinese indicine (Leiqiong, Guangfeng, Ji'an, Jinjiang and Wannan), and Indian indicine (Tharparkar, Nelore, Sahiwal, Hariana and Gir) (**Supplementary Table S2**).

Xiangxi cattle have been discovered with 40573969 biallelic SNPs. The functional annotation of the polymorphic loci displayed that the vast majority of SNPs existed in the intergenic region (60.1%) or intron region (38.2%). Exon accounted for 0.7% of the total SNPs, including 112572 nonsynonymous and 179112 synonymous SNPs (**Supplementary Figure S1**). In addition, 0.9% of the SNPs were detected in an untranslated region (UTR), and the remaining 0.1% in splice. Xiangxi cattle was found to have the largest number of SNPs which was consistent with Chinese indicine on the whole, whereas Qinchuan was only second to them (**Supplementary Table S3**).

Population Structure and Characterization

Based on the study of autosomal SNPs, we discussed the phylogenetic relationships among 101 samples from 7 populations (**Figure 1**). The NJ (neighbor-joining) tree showed that all "core" cattle populations form separate clusters. Qinchuan cattle are located between taurine cattle and indicine cattle (Mei et al., 2018), while Xiangxi cattle are aggregated near Chinese indicine and Indian indicine (**Figure 1A**). Principal component analysis (PCA) provides similar results to the above conclusions (**Figure 1B**).

We estimated the ancestral populations of all cattle samples by using clustering models. When K = 2, the lineage of cattle can be fundamentally distinguished from that of taurine and indicine cattle. When K = 5, five "core" cattle herds were separated, while Xiangxi cattle exhibited mixed phenomena (**Figure 1C**). The result displayed that it had a mutual genome ancestor with Chinese indicine, Indian indicine, and East Asian taurine. Moreover, a more dramatic genetic influence of Chinese indicine than the others was shown in the results.

Population Genetic Diversity

As shown in **Figure 2A**, the nucleotide diversity of Xiangxi cattle (0.00339) was second only to Chinese indicine (0.00357) and close to Qinchuan cattle (0.00247) and Indian indicine (0.00238), while that of Eurasian taurine (0.00095) was the lowest. The results of the linkage disequilibrium analysis agreed with these findings. The lowest LD level was found at short distances in Xiangxi cattle, followed by Chinese indicine, and the Eurasian taurine (Jersey) showed a higher LD level (**Figure 2B**).

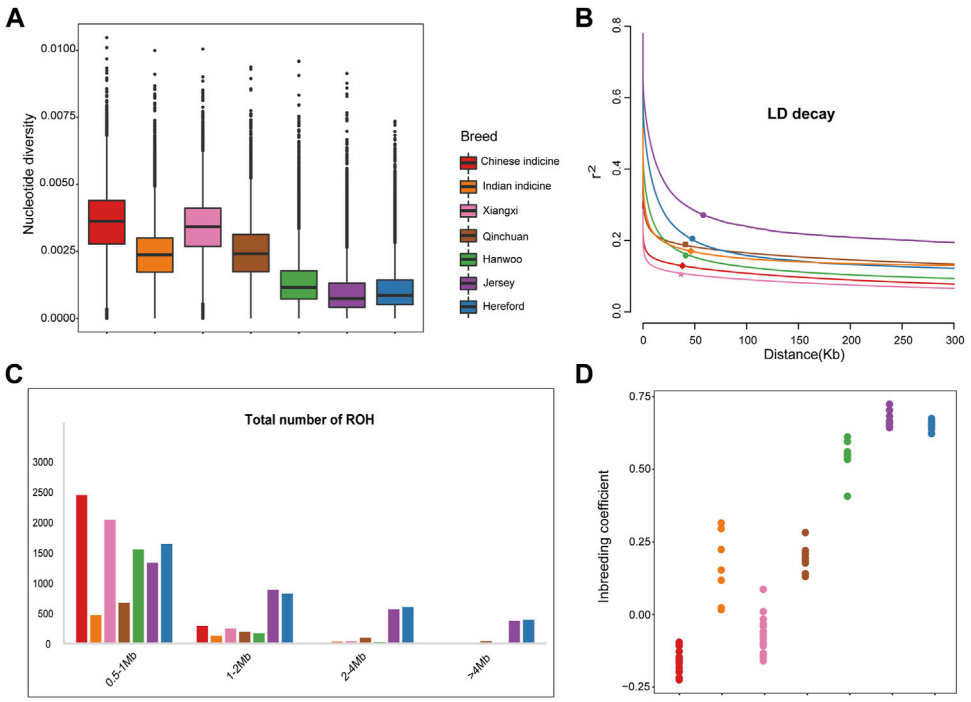

FIGURE 2 | Genetic diversity among 101 samples from 7 populations. **(A)** Box plots of the nucleotide diversity for each group. The points which were on the outside the of whiskers showed outliers **(B)** Decay of linkage disequilibrium on cattle autosomes estimated from each breed. **(C)** The estimation of total number of ROH for each group **(D)** Inbreeding coefficient for each individual.

Meanwhile, the ROH of each breed was divided by length (0.5–1 Mb, 1–2 Mb, 2–4 Mb and >4 Mb). In order of the number of ROH, it demonstrated that the European taurine took the largest number. By contrast, the number of the Indian indicine is lowest, that of the Chinese Crossbreed and Chinese indicine are in the middle respectively (**Figure 2C**). The Xiangxi cattle we were concerned with showed more extensive amounts of short/medium ROH (0.5–2 Mb) but there were no long ROH (2–4 Mb). The number of ROH in European taurine cattle is much larger than in Xiangxi cattle due to high breeding. It is noteworthy that highly selective breeding may be at risk of inbreeding depression. As expected, the inbreeding coefficient also proved these findings (**Figure 2D**).

Genome-Wide Selective Sweep Test

For the sake of screening out the associated genetic variation, the nucleotide diversity analysis (θπ) and the composite likelihood ratio (CLR) were used to explore the selected genomic regions in Xiangxi cattle (**Supplementary Tables S4, S5**) (**Figure 3**). We found that partial genome regions in Xiangxi cattle might have been selected during domestication. A total of 2225 (θπ) and 182(CLR) genes were identified. Among them, 127 genes were overlapped which were considered to be candidate genes (**Supplementary Table S6**). The annotations of candidate genes revealed the functions that may be associated with economic traits, including reproduction (*KHDRBS2*, *GOLGA4*, *BBX*, *BANF2*, *INSL6*) (Cai et al., 2017; Guo et al., 2018; Ortega et al., 2016; Pini et al., 2020; Wang et al., 2016), growth (*RALGAPA1*, *NCK1*, *VGLL2*, *FGFR3*) (Aryal et al., 2015; Honda et al., 2017; Sun et al., 2020; Wen et al., 2016), meat quality (*PLIN4*) (Pena et al., 2019), and somatotype (*SYN3*) (An et al., 2019). We also obtained genes (*ELANE, AZU1*) (Cui et al., 2021; Zhang et al., 2021) related to immune response, which may result from long-term natural selection.

Furthermore, we implemented two methods, F_{ST} ($p < 0.005$, $F_{ST} \geq 0.29979$) and XP-EHH ($p < 0.005$, XP-EHH≥2.27), to elucidate further the positive selection characteristics between Xiangxi cattle and Qinchuan cattle (**Figure 4**), which contained 1,264 and 553 hypothetical selection genes, respectively (**Supplementary Tables S7, S8**). 102 overlapping genes were also tested in both methods (**Supplementary Table S9**). Among them, regions containing known candidate genes related to heat tolerance (*DNAJC8*) (Li G. et al., 2020) showed intense differentiation signals. As for functional enrichment analysis (**Figure 5**), the KEGG pathway had only one significant pathway called "Regulation of actin cytoskeleton" (corrected *p*-value < 0.05) and 5 genes were performed (*PIP4K2A*, *FGF22*, *DIAPH1*, *WASF2*, *SLC9A1*), which were related to meat quality, growth efficiency, and the balance of ionic concentrations (**Supplementary Table S10**). GO terms were enriched, in particular including the terms of skin barrier and epidermal cell differentiation ("establishment of skin barrier, GO: 0061436" "positive regulation of epidermal cell differentiation, GO:0045606" "negative regulation of keratinocyte proliferation, GO:0010839"). These several genes (*SFN*, *PALLD*, *KDF1*, *FA2H*) were found in Xiangxi cattle which were associated with the

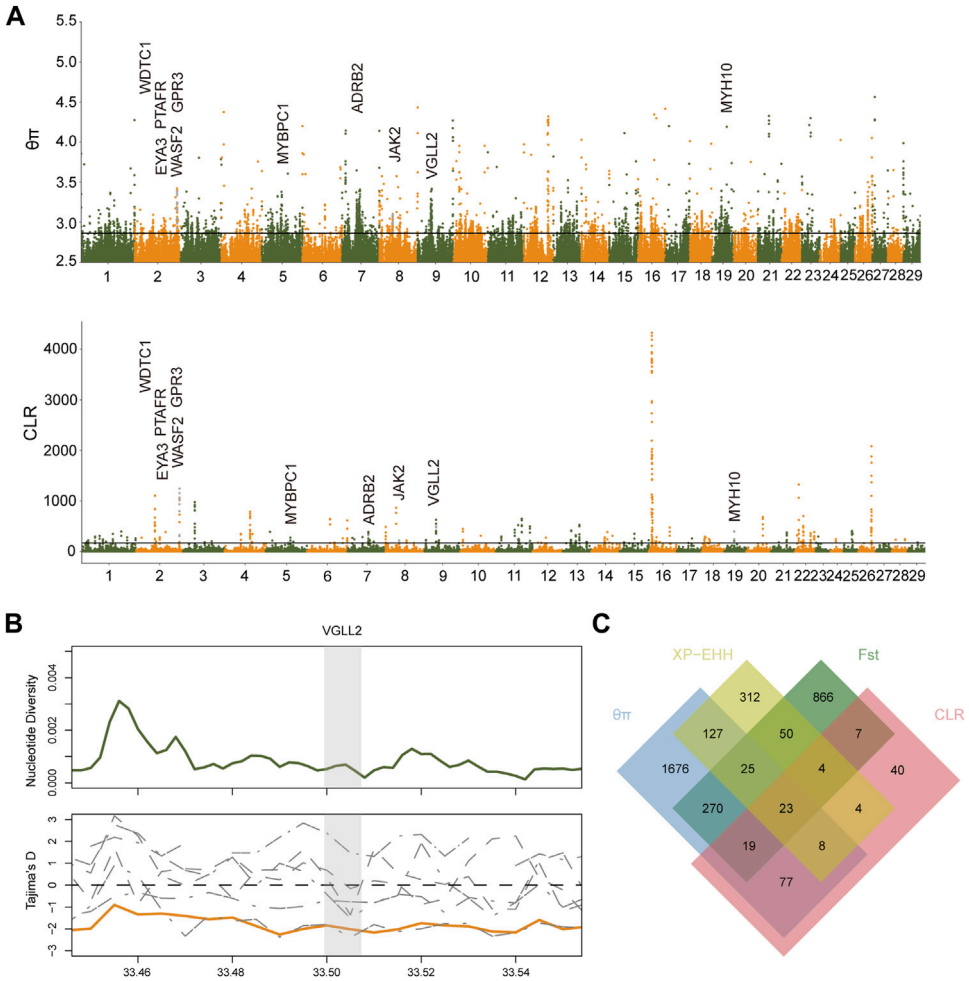

FIGURE 3 | The signatures of positive selection in Xiangxi cattle. **(A)** Manhattan plot of selective sweeps by θπ and CLR methods **(B)** Nucleotide diversity of *VGLL2* gene region in Xiangxi cattle. and Tajima's D value in each group (orange line in Xiangxi cattle). **(C)** Number of candidate genes identified in Xiangxi cattle by the four methods (θπ, CLR, F_{ST} and XP-EHH) listed in each of the Venn diagram components.

development of sebum, sebaceous glands, and epidermal cells. In addition, we also detected significant GO terms responsible for meat quality ("positive regulation of lipophagy, GO:1904504" "calmodulin binding, GO:0005516") (**Supplementary Table S11**) involving relevant genes (*ADRB2*, *SESN2*, *MYH10*, *ATP5IF1*, *SMTNL1*, *SLC9A1*).

Above all, among the above four applied methods, 23 candidate genes were involved in production and immunization traits (**Supplementary Table S12**). e.g., The *ADRB2* gene was associated with muscle pH (Murani et al., 2013). The *MYH10* gene was discovered to potentially affect the beef quality of the longissimus lumborum muscle following castration (Li et al., 2020b). *WASF2* gene affects cytoskeleton arrangement. The *EYA3* gene was revealed to have a significant correlation with muscle development (Xia et al., 2021). *MYBPC1* genes are well-known marbling-related genes, which were first identified in Japanese Black beef cattle (Li et al., 2020c). The *GPR3* gene was relative to fat development in mice (Godlewski et al., 2015). The *WDTC1* gene was involved in the regulation of fat-related gene transcription (Ducos et al., 2017). PTAFR was also reported to be involved in inflammatory responses in cattle (Trovato et al., 2015; Vanvanhossou et al., 2020; Xerxa et al., 2016). *JAK2* mediates essential signaling events in innate and adaptive immunity (Lukashova et al., 2003).

DISCUSSION

In order to figure out the population structure and genetic diversity of Xiangxi cattle, we utilized the representative worldwide cattle dataset, making a comparison in a number of contexts and analyses as follows. Ancestor Component Analysis can reflect the communicating extent of genetic information. Xiangxi cattle was complex in the lineage composition which included Chinese indicine (~73%), East Asian taurine (~20%), and Indian indicine (~7%) (**Supplementary Figure S2**). The heterogeneity of Chinese cattle was induced by different taurine and indicine lineage proportions. Since Xiangxi cattle

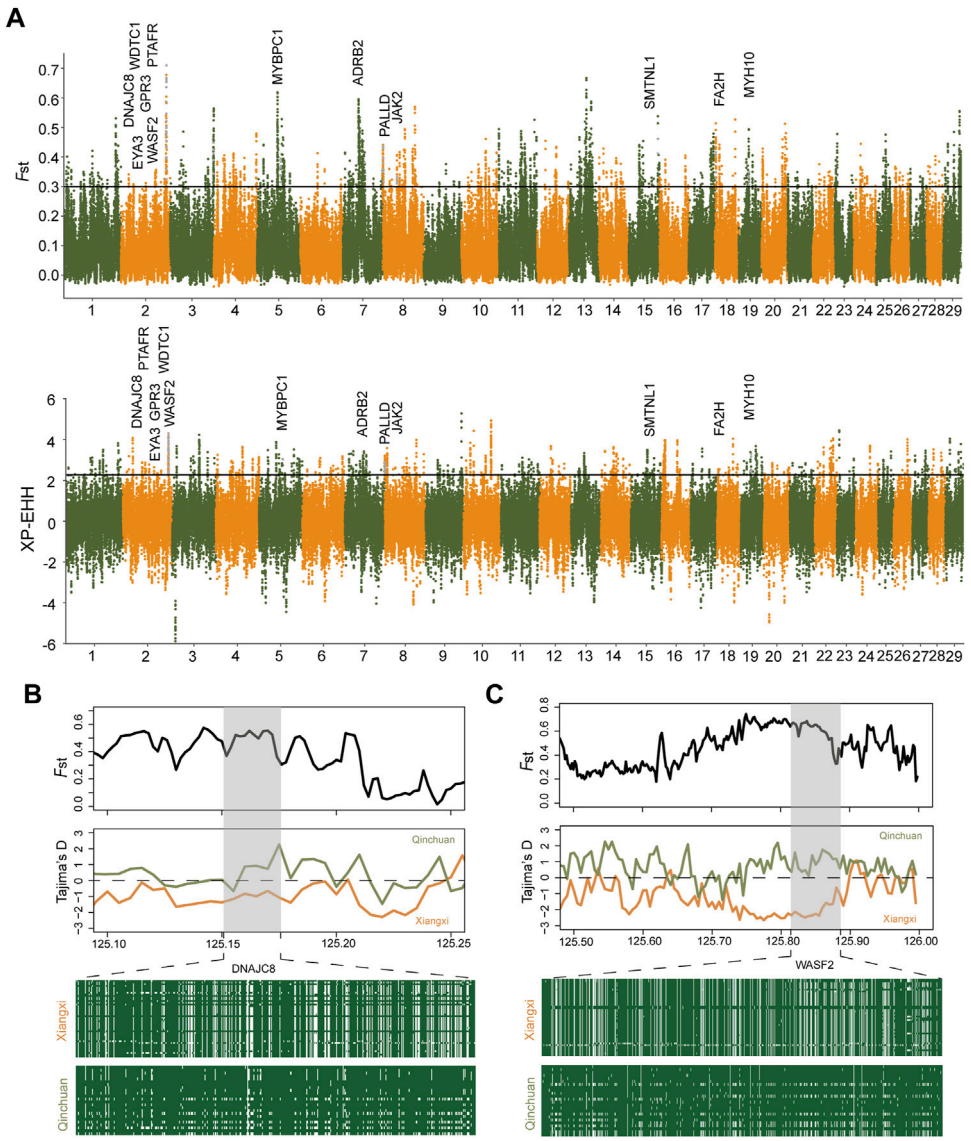

FIGURE 4 | Selective signals between Xiangxi cattle and Qinchuan cattle. **(A)** Manhattan plot of selective sweeps by F_{ST} and XP-EHH methods. F_{ST} and Tajima's D plots of *DNAJC8* gene **(B)** and *WASF2* gene **(C)**. Haplotype diversity at the example genes **(B,C)**.

was situated in the southern regions of China, its ancestry was affected by indicine more than taurine cattle. On the contrary, Qinchuan cattle was situated in North-central China. Using the principal component and neighbor-joining tree analytical methods, we observed an accurate division of seven populations. In particular, Xiangxi cattle were markedly separated into two branches (**Figure 1A**), which may be caused by the short-term hybridization of foreign commercial cattle, or the self-domesticated Indian zebu which have been migrated from the Indus Valley in the East and entered China from Yunnan Province about three thousand years ago. It may subsequently have an effect on southern Chinese cattle.

In the present study, the genomic variation parameters of the populations all had a similar trend. We ranked all breeds according to their nucleotide diversity (Chinese indicine > Xiangxi cattle > Qinchuan cattle > Indian indicine > taurine cattle), which was consistent with the previous study (Liu et al., 2020). By contrasting the two Chinese breeds, it was found that the genetic diversity of Xiangxi cattle exceeded the Qinchuan cattle, possibly because of the fact that Xiangxi cattle contained more Chinese indicine cattle lineage. The highest genetic diversity observed in Chinese indicine was induced by introgression of other bovines (*Bos javanicus*) (Chen et al., 2018). Similarly, the analysis of SNPs, inbreeding coefficient, and LD decay was highly concurrent with the abundant genetic diversity of Xiangxi cattle. The length and distribution density of ROH revealed that the degree of Xiangxi cattle was obviously less than taurine (**Figure 2**). Our results showed that Xiangxi cattle had more

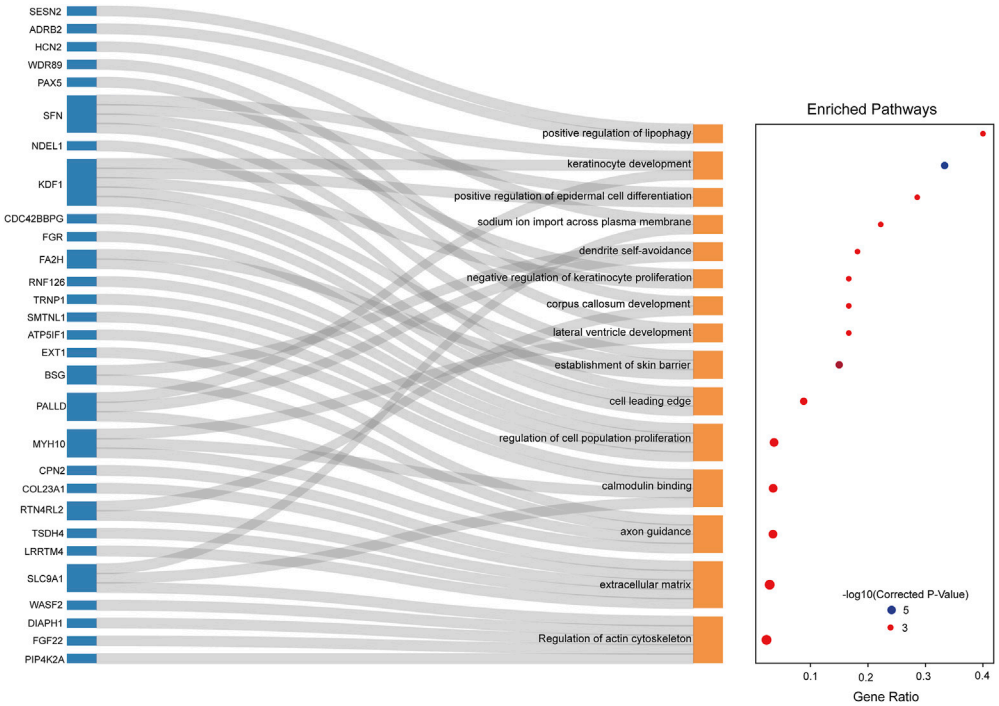

FIGURE 5 | KEGG and GO enrichment analysis of Xiangxi cattle candidate genes overlapped by F_{ST} and XP-EHH methods (corrected *p*-value < 0.05).

genetic diversity than taurine cattle due to insufficient breeding. After a long period of intensive breeding, taurine cattle have formed a mature commercial system, whereas the breeding system of Xiangxi cattle was imperfect and has tremendous potential.

Xiangxi cattle is an advantageous resource for the development of animal husbandry because of its excellent characteristics, especially for its disease-resistance and environmental adaptability. To increase the assay efficiency and decrease false positives, four selective sweeping methods were performed on Xiangxi cattle. If the gene was conspicuously detected by at least two methods, then it will be served as an actual candidate gene. Xiangxi cattle is muscular, and have long been used as draft animals. According to θπ and CLR (**Figure 3A**), the overlapped candidate gene (*VGLL2*) played a vital role in the process of locomotion. The *VGLL2* gene can affect the development of skeletal muscles (Gabriel et al., 2016; Honda et al., 2017,2019). Subsequently the positive selection region including the *VGLL2* gene was intensively confirmed by the lower values of Tajima'D and nucleotide diversity analysis in Xiangxi cattle (**Figures 3B,C**).

The surroundings of Xiangxi cattle are hotter than Qinchuan cattle. In the long-term natural and artificial selection, Xiangxi cattle has adapted to the thermal climate in southern China. Using Qinchuan cattle as the reference population, we calculated F_{ST} and XP-EHH (**Figure 4A**). The *DNAJC8* gene, a member of DnaJ Heat Shock Protein Family, was observed to act as an important part of heat stress. Previous studies suggested that *DNAJC8* protected bees from heat stress by regulating heat-inducible and antioxidant genes (Li G. et al., 2020). Heat stress triggers the production of reactive oxygen species (ROS) (Slimen et al., 2014). Previous studies have shown that DnaJ protein plays an important role in protecting antioxidant enzyme activity and removing excess ROS(Kong et al., 2014; Wang et al., 2014). In order to avoid false positives, the verification analysis was carried out on the gene region. *DNAJC8* was located on chromosome 2 of cattle, about 0.025 Mbp. This region showed extremely apparent differentiation and distinct haplotype patterns in two populations. Also, Tajima'D analysis exhibited that it was significantly lower in Xiangxi cattle (**Figure 4B**). These results provided forceful evidence that the region containing *DNAJC8* may have a crucial effect on heat resistance in Xiangxi cattle.

In addition, GO terms (e.g., keratinocyte development, establishment of skin barrier, positive regulation of epidermal cell differentiation) were enriched, involving epidermal cell related genes (e.g., *SFN*, *PALLD*, *KDF1*, *FA2H*) (**Figure 5**). The *SFN* gene can affect the expression of key matrix metalloproteinase (MMPs) in skin fibroblasts, thereby controlling the process of cutaneous wound healing (Kilani et al., 2008; Medina et al., 2007). *FA2H* is essential for the synthesis of HFA-sphingolipids in various organs which is highly expressed in the skin and crucial for the formation of the epidermal barrier (Maier et al., 2011). Xiangxi cattle strengthen the resistance to external toxins and prevent excessive loss of body fluids through the action of keratinocytes, thus adapting to the local heat and humid environment better. A KEGG pathway called "Regulation of

actin cytoskeleton" enriched genes (*PIP4K2A*, *FGF22*, *DIAPH1*, *WASF2*, *SLC9A1*) associated with the muscle contraction. *SLC9A1* improves the adjustment capacity of muscle pH value and maintains body endurance (Iaia et al., 2008; Skovgaard et al., 2014). It was worth noting that the *WASF2* gene, which was overlapped in the four selection methods, was a downstream effector molecule that was involved in the signal transduction from tyrosine kinase receptors and small GTPases to actin cytoskeleton and promoted actin filament formation (Miki et al., 1998). The gene can respond to changes of the external environment by reregulating its expression or distribution to further influence the cytoskeleton arrangement (Qian et al., 2009). Besides, *WASF2* was also related to innate immunity, which may be attributed to the pleiotropy of the gene (Liu et al., 2021; Nolz et al., 2006). Compared to Qinchuan cattle, Xiangxi cattle conserved more *WASF2*. The haplotype patterns of the *WASF2* gene were different between the two populations (**Figure 4C**). Therefore, it was speculated that the *WASF2* gene was selected in Xiangxi cattle. In the past, due to the local geographical environment and harsh production conditions, Xiangxi cattle were mainly selected for draft, resulting in the characteristics of high endurance and strong assault draft. The above genes might be related to these characteristics.

In four selected methods (θπ, CLR, F_{ST}, and XP-EHH), six genes (*ADRB2*, *WDTC1*, *MYBPC1*, *EYA3*, *MYH10*, *GPR3*) also showed positive selection, which may be related to superior meat quality traits in Xiangxi cattle. Two genes (*PTAFR*, *JAK2*) were related to the disease-resistance of Xiangxi cattle. *PTAFR* gene, platelet activating factor receptor, can regulate inflammatory, smooth-muscle contractile and hypotensive activity (Honda et al., 2002; Marathe et al., 1999). Additionally, *JAK2* is one of the tyrosine kinase family genes related to cytokine receptors, which has a certain correlation with *PTAFR* gene in immune response (Lukashova et al., 2003). Therefore, they were vital for the disease resistance of Xiangxi cattle.

In summary, Xiangxi Cattle is one of the famous local excellent breeds in China with superior traits (disease-resistance, environmental adaptability, meat quality). Our selective sweep identification of Xiangxi cattle may promote the cognition of the hereditary mechanism of potential population characteristics, and hopefully serve as a theoretical reference for the breeding in the future.

CONCLUSION

In this study, we used WGS data to study the population structure of Xiangxi cattle, so as to provide the first in-depth study of its gene diversity, phylogenetic relationship, and Selective Sweep Test. Some genes were identified that will not only be helpful to us for a better understanding of the features of Xiangxi cattle, but also further study the characteristics of other native cattle of China. In conclusion, the revelation of the genetic diversity of Xiangxi cattle will establish a sound foundation for conservation breeding in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, accession number: PRJNA779877.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of Northwest A&F University.

AUTHOR CONTRIBUTIONS

XL and JL execution of this manuscript. CX, LS, WX, NC, and CL revised the manuscript. HL, YL, TL, and QS contributed to the sample collections. KY provided the laboratories for statistical analysis and the funding for the research.

FUNDING

This work was supported by the key research and development project of Hunan Province, China (No. 2020NK 2066), Innovation platform and talent plan of Hunan Province, China (No. 2019TP 2010) and the China Agriculture Research System of MOF and MARA (CARS-37).

ACKNOWLEDGMENTS

We thank the High-Performance Computing of Northwest A&F University for providing computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.816379/full#supplementary-material

REFERENCES

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinforma.* 12, 246. [Journal Article; Research Support, N.I.H., Extramural]. doi:10.1186/1471-2105-12-246

An, B., Xia, J., Chang, T., Wang, X., Xu, L., Zhang, L., et al. (2019). Genome-wide Association Study Reveals Candidate Genes Associated with Body Measurement Traits in Chinese Wagyu Beef Cattle. *Anim. Genet.* 50 (4), 386–390. doi:10.1111/age.12805

Aryal, A. C, S., Miyai, K., Izu, Y., Hayata, T., Notomi, T., Noda, M., et al. (2015). Nck Influences Preosteoblastic/osteoblastic Migration and Bone Mass. *P. Natl. Acad. Sci. Usa.* 112 (50), 15432–15437. doi:10.1073/pnas.1518253112

Cai, C., Ning, Q., Chen, Q., Zhu, Y., Chen, N., Chen, H., et al. (2019). Genetic Diversity of Y-SNPs in Xiangxi Cattle (In Chinese). *China Cattle Sci.* 45 (05), 34–36.

Cai, X., Yu, S., Mipam, T., Yang, F., Zhao, W., Liu, W., et al. (2017). Comparative Analysis of Testis Transcriptomes Associated with Male Infertility in Cattleyak. *Theriogenology.* 88, 28–42. doi:10.1016/j.theriogenology.2016.09.047

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome Resequencing Reveals World-wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nat. Commun.* 9 (1), 2337. doi:10.1038/s41467-018-04737-0

Choi, J. W., Choi, B. H., Lee, S. H., Lee, S. S., Kim, H. C., Yu, D., et al. (2015). Whole-Genome Resequencing Analysis of Hanwoo and Yanbian Cattle to Identify Genome-wide SNPs and Signatures of Selection. *Mol. Cells.* 38 (5), 466–473. [Journal Article.; Research Support, Non-U.S. Gov't]. doi:10.14348/molcells.2015.0019

Cui, C., Chakraborty, K., Tang, X. A., Zhou, G., Schoenfelt, K. Q., Becker, K. M., et al. (2021). Neutrophil Elastase Selectively Kills Cancer Cells and Attenuates Tumorigenesis. *Cell.* 184 (12), 3163–3177. doi:10.1016/j.cell.2021.04.016

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., et al. (2014). Whole-genome Sequencing of 234 Bulls Facilitates Mapping of Monogenic and Complex Traits in Cattle. *Nat. Genet.* 46 (8). doi:10.1038/ng.3034

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. [Journal Article; Research Support, N.I.H., Extramural; Research Support, N.I.H., Intramural; Research Support, Non-U.S. Gov't]. doi:10.1093/bioinformatics/btr330

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., and Nielsen, R. (2016). SweepFinder2: Increased Sensitivity, Robustness and Flexibility. [Journal Article]. *Bioinformatics.* 32 (12), 1895–1897. doi:10.1093/bioinformatics/btw051

Ducos, E., Vergès, V., Dugé de Bernonville, T., Blanc, N., Giglioli-Guivarc'h, N., and Dutilleul, C. (2017). Remarkable Evolutionary Conservation of Antiobesity ADIPOSE/WDTC1 Homologs in Animals and Plants. *Genetics.* 207 (1), 153–162. doi:10.1534/genetics.116.198382

Gabriel, B. M., Hamilton, D. L., Tremblay, A. M., and Wackerhage, H. (2016). The Hippo Signal Transduction Network for Exercise Physiologists. *J. Appl. physiology.* 1985120 (10), 1105–1117. doi:10.1152/japplphysiol.01076.2015

Godlewski, G., Jourdan, T., Szanda, G., Tam, J., Cinar, R., Harvey-White, J., et al. (2015). Mice Lacking GPR3 Receptors Display Late-Onset Obese Phenotype Due to Impaired Thermogenic Function in Brown Adipose Tissue. *Sci. Rep.-UK.* 5, 14953. doi:10.1038/srep14953

Guo, J., Tao, H., Li, P., Li, L., Zhong, T., Wang, L., et al. (2018). Whole-genome Sequencing Reveals Selection Signatures Associated with Important Traits in Six Goat Breeds. *Sci. Rep.-UK.* 8 (1), 10405. doi:10.1038/s41598-018-28719-w

Honda, M., Hidaka, K., Fukada, S., Sugawa, R., Shirai, M., Ikawa, M., et al. (2017). Vestigial-like 2 Contributes to Normal Muscle Fiber Type Distribution in Mice. *Sci. Rep.-UK.* 7 (1), 7168. doi:10.1038/s41598-017-07149-0

Honda, M., Tsuchimochi, H., Hitachi, K., and Ohno, S. (2019). Transcriptional Cofactor Vgll2 Is Required for Functional Adaptations of Skeletal Muscle Induced by Chronic Overload. *J. Cell. Physiol.* 234 (9), 15809–15824. doi:10.1002/jcp.28239

Honda, Z., Ishii, S., and Shimizu, T. (2002). Platelet-activating Factor Receptor. *J. Biochem.* 131 (6), 773–779. doi:10.1093/oxfordjournals.jbchem.a003164

Iaia, F. M., Thomassen, M., Kolding, H., Gunnarsson, T., Wendell, J., Rostgaard, T., et al. (2008). Reduced Volume but Increased Training Intensity Elevates Muscle Na+-K+ Pump Alpha1-Subunit and NHE1 Expression as Well as Short-Term Work Capacity in Humans. [Journal Article; Research Support, Non-U.S. Gov't]. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 294 (3), R966–R974. doi:10.1152/ajpregu.00666.2007

Kilani, R. T., Medina, A., Aitken, A., Jalili, R. B., Carr, M., and Ghahary, A. (2008). Identification of Different Isoforms of 14-3-3 Protein Family in Human Dermal and Epidermal Layers. *Mol. Cell. Biochem.* 314 (1-2), 161–169. doi:10.1007/s11010-008-9777-6

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The Genome Landscape of Indigenous African Cattle. [Journal Article; Research Support, Non-U.S. Gov't]. *Genome Biol.* 18 (1), 34. doi:10.1186/s13059-017-1153-y

Kong, F., Deng, Y., Wang, G., Wang, J., Liang, X., and Meng, Q. (2014). LeCDJ1, a Chloroplast DnaJ Protein, Facilitates Heat Tolerance in Transgenic Tomatoes. *J. Integr. Plant Biol.* 56 (1), 63–74. doi:10.1111/jipb.12119

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1093/molbev/msw054

Lee, K. T., Chung, W. H., Lee, S. Y., Choi, J. W., Kim, J., Lim, D., et al. (2013). Whole-genome Resequencing of Hanwoo (Korean Cattle) and Insight into Regions of Homozygosity. *BMC Genomics.* 14, 519. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1186/1471-2164-14-519

Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) V5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* 49 (W1), W293–W296. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1093/nar/gkab301

Li, G., Zhao, H., Guo, H., Wang, Y., Cui, X., Li, H., et al. (2020a). Analyses of the Function of DnaJ Family Proteins Reveal an Underlying Regulatory Mechanism of Heat Tolerance in Honeybee. [Journal Article]. *Sci. Total Environ.* 716, 137036. doi:10.1016/j.scitotenv.2020.137036

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics.* 25 (14), 1754–1760. [Journal Article.; Research Support, Non-U.S. Gov't]. 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. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1093/bioinformatics/btp352

Li, Y., Cheng, G., Yamada, T., Liu, J., Zan, L., and Tong, B. (2020b). Effect of Expressions and SNPs of Candidate Genes on Intramuscular Fat Content in Qinchuan Cattle. *Animals.* 10 (8), 1370. doi:10.3390/ani10081370

Li, Y., Wang, M., Li, Q., Gao, Y., Li, Q., Li, J., et al. (2020c). Transcriptome Profiling of Longissimus Lumborum in Holstein Bulls and Steers with Different Beef Qualities. *PLoS One.* 15 (6), e235218. doi:10.1371/journal.pone.0235218

Liu, M., Zhang, J., Pinder, B. D., Liu, Q., Wang, D., Yao, H., et al. (2021). WAVE2 Suppresses mTOR Activation to Maintain T Cell Homeostasis and Prevent Autoimmunity. *Sci. (New York, N.Y.).* 371 (6536). doi:10.1126/science.aaz4544

Liu, Y., Xu, L., Yang, L., Zhao, G., Li, J., Liu, D., et al. (2020). Discovery of Genomic Characteristics and Selection Signatures in Southern Chinese Local Cattle. [Journal Article]. *Front. Genet.* 11, 533052. doi:10.3389/fgene.2020.533052

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for Two Independent Domestications of Cattle. *Comp. Study.* 91, 2757–2761. (Reprinted. doi:10.1073/pnas.91.7.2757

Long, Z., and Ma, W. (2009). Characteristics and Utilization of Xiangxi Cattle Breeds (In Chinese). *Mod. Agric. Sci. Technol.* (16), 274–275.

Lukashova, V., Chen, Z., Duhé, R. J., Rola-Pleszczynski, M., and Stanková, J. (2003). Janus Kinase 2 Activation by the Platelet-Activating Factor Receptor (PAFR): Roles of Tyk2 and PAFR C Terminus. *J. Immunol.* 1950 171 (7), 3794–3800. doi:10.4049/jimmunol.171.7.3794

Luoreng, Z. M., Wang, X. P., Ma, Y., Li, F., Guo, D. S., Li, N., et al. (2014). Three Novel SNPs in the Coding Region of the Bovine MC3R Gene and Their Associations with Growth Traits. [Journal Article; Research Support, Non-U.S. Gov't]. *Biochem. Genet.* 52 (1-2), 116–124. doi:10.1007/s10528-013-9632-7

Maier, H., Meixner, M., Hartmann, D., Sandhoff, R., Wang-Eckhardt, L., Zöller, I., et al. (2011). Normal Fur Development and Sebum Production Depends on Fatty Acid 2-hydroxylase Expression in Sebaceous Glands. *J. Biol. Chem.* 286 (29), 25922–25934. doi:10.1074/jbc.M111.231977

Marathe, G. K., Davies, S. S., Harrison, K. A., Silva, A. R., Murphy, R. C., Castro-Faria-Neto, H., et al. (1999). Inflammatory Platelet-Activating Factor-like Phospholipids in Oxidized Low Density Lipoproteins Are Fragmented Alkyl Phosphatidylcholines. *J. Biol. Chem.* 274 (40), 28395–28404. [Journal Article; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, P.H.S.]. doi:10.1074/jbc.274.40.28395

Medina, A., Ghaffari, A., Kilani, R. T., and Ghahary, A. (2007). The Role of Stratifin in Fibroblast-Keratinocyte Interaction. *Mol. Cell. Biochem.* 305 (1-2), 255–264. doi:10.1007/s11010-007-9538-y

Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., et al. (2018). Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome

Resequencing. [Journal Article]. *Mol. Biol. Evol.* 35 (3), 688–699. doi:10.1093/molbev/msx322

Miki, H., Suetsugu, S., and Takenawa, T. (1998). WAVE, a Novel WASP-Family Protein Involved in Actin Reorganization Induced by Rac. *EMBO J.* 17 (23), 6932–6941. doi:10.1093/emboj/17.23.6932

Murani, E., Ponsuksili, S., Reyer, H., Wittenburg, D., and Wimmers, K. (2013). Expression Variation of the Porcine ADRB2 Has a Complex Genetic Background. *Mol. Genet. Genomics.* 288 (11), 615–625. doi:10.1007/s00438-013-0776-2

Nolz, J. C., Gomez, T. S., Zhu, P., Li, S., Medeiros, R. B., Shimizu, Y., et al. (2006). The WAVE2 Complex Regulates Actin Cytoskeletal Reorganization and CRAC-Mediated Calcium Entry during T Cell Activation. *Curr. Biol. Cb.* 16 (1), 24–34. doi:10.1016/j.cub.2005.11.036

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

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis, [Journal Article; Research Support. N.I.H., Extramural; Research Support, Non-U.S. Gov't]. *PLoS Genet.* 2 (12), e190. doi:10.1371/journal.pgen.0020190

Pena, R. N., Noguera, J. L., García-Santana, M. J., González, E., Tejeda, J. F., Ros-Freixedes, R., et al. (2019). Five Genomic Regions Have a Major Impact on Fat Composition in Iberian Pigs. *Sci. Rep.-UK.* 9 (1), 2031. doi:10.1038/s41598-019-38622-7

Pini, T., Parks, J., Russ, J., Dzieciatkowska, M., Hansen, K. C., Schoolcraft, W. B., et al. (2020). Obesity Significantly Alters the Human Sperm Proteome, with Potential Implications for Fertility. *J. Assist. Reprod. Gen.* 37 (4), 777–787. doi:10.1007/s10815-020-01707-8

Qian, A., Di, S., Gao, X., Zhang, W., Tian, Z., Li, J., et al. (2009). CDNA Microarray Reveals the Alterations of Cytoskeleton-Related Genes in Osteoblast under High Magneto-Gravitational Environment. *Acta bioch. bioph. Sin.* 41 (7), 561–577. doi:10.1093/abbs/gmp041

Skovgaard, C., Christensen, P. M., Larsen, S., Andersen, T. R., Thomassen, M., and Bangsbo, J. (2014). Concurrent Speed Endurance and Resistance Training Improves Performance, Running Economy, and Muscle NHE1 in Moderately Trained Runners. *J. Appl. Physiol.* 117 (10), 1097–1109. [Journal Article; Randomized Controlled Trial; Research Support, Non-U.S. Gov't]. doi:10.1152/japplphysiol.01226.2013

Slimen, I. B., Najar, T., Ghram, A., Dabbebi, H., Ben Mrad, M., and Abdrabbah, M. (2014). Reactive Oxygen Species, Heat Stress and Oxidative-Induced Mitochondrial Damage. A Review. *Int. J. Hyperth. official J. Eur. Soc. Hyperthermic Oncol. North Am. Hyperth. Group.* 30 (7), 513–523. doi:10.3109/02656736.2014.971446

Stothard, P., Choi, J. W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., et al. (2011). Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery. *BMC Genomics.* 12, 559. [Comparative Study; Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1186/1471-2164-12-559

Sun, T., Huang, G., Wang, Z., Teng, S., Cao, Y., Sun, J., et al. (2020). Selection Signatures of Fuzhong Buffalo Based on Whole-Genome Sequences. *BMC Genomics.* 21 (1), 674. doi:10.1186/s12864-020-07095-8

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31 (10), 2824–2827. [Letter; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1093/molbev/msu211

Trovato, A., Panelli, S., Strozzi, F., Cambulli, C., Barbieri, I., Martinelli, N., et al. (2015). Expression of Genes Involved in the T Cell Signalling Pathway in Circulating Immune Cells of Cattle 24 Months Following Oral Challenge with Bovine Amyloidotic Spongiform Encephalopathy (BASE). *BMC Vet. Res.* 11, 105. doi:10.1186/s12917-015-0412-y

Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., and König, S. (2020). A Multi-Breed GWAS for Morphometric Traits in Four Beninese Indigenous Cattle Breeds Reveals Loci Associated with Conformation, Carcass and Adaptive Traits. *BMC Genomics.* 21 (1), 783. doi:10.1186/s12864-020-07170-0

Wang, C., Tang, M., Chang, W., Furushima, K., Jang, C., Behringer, R. R., et al. (2016). PiggyBac Transposon-Mediated Mutagenesis in Rats Reveals a Crucial Role of Bbx in Growth and Male Fertility. *Biol. Reprod.* 95 (3), 51. doi:10.1095/biolreprod.116.141739

Wang, G., Cai, G., Kong, F., Deng, Y., Ma, N., and Meng, Q. (2014). Overexpression of Tomato Chloroplast-Targeted DnaJ Protein Enhances Tolerance to Drought Stress and Resistance to Pseudomonas Solanacearum in Transgenic Tobacco. *Plant physiology Biochem. PPB.* 82, 95–104. doi:10.1016/j.plaphy.2014.05.011

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. [Evaluation Study; Journal Article; Research Support, N.I.H., Extramural]. doi:10.1093/nar/gkq603

Wang, X. P., Luoreng, Z. M., Li, F., Wang, J. R., Li, N., and Li, S. H. (2012). Genetic Polymorphisms of Lipoprotein Lipase Gene and Their Associations with Growth Traits in Xiangxi Cattle. [Journal Article; Research Support, Non-U.S. Gov't]. *Mol. Biol. Rep.* 39 (12), 10331–10338. doi:10.1007/s11033-012-1910-7

Wen, X., Li, X., Tang, Y., Tang, J., Zhou, S., Xie, Y., et al. (2016). Chondrocyte FGFR3 Regulates Bone Mass by Inhibiting Osteogenesis. *J. Biol. Chem.* 291 (48), 24912–24921. doi:10.1074/jbc.m116.730093

Xerxa, E., Barbisin, M., Chieppa, M. N., Krmac, H., Vallino Costassa, E., Vatta, P., et al. (2016). Whole Blood Gene Expression Profiling in Preclinical and Clinical Cattle Infected with Atypical Bovine Spongiform Encephalopathy. *PLoS One.* 11 (4), e153425. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1371/journal.pone.0153425

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics.* 22 (1), 43. [Journal Article; Review]. doi:10.1186/s12864-020-07340-0

Yao, Y., Jia, Y., Xia, X., Zhu, Y., Lan, X., Huang, Y., et al. (2017). Genetic Diversity of Mitochondrial DNA D-Loop Region in Xiangxi Cattle (In Chinese). *China Cattle Sci.* 43 (05), 11–13.

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. [Journal Article; Research Support, Non-U.S. Gov't]. *Bioinformatics.* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, R., Li, Y., and Xing, X. (2021). Comparative Antler Proteome of Sika Deer from Different Developmental Stages. *Sci. Rep.-UK.* 11 (1), 10484. doi:10.1038/s41598-021-89829-6

Zhang, Y., Xu, S., and Wang, Z. (2011). *Animal Genetic Resources in China-bovines (In Chinese)*. Beijing: China Agriculture Press.

Conflict of Interest: Authors YL and TL is employed by Hunan De Nong Animal Husbandry Group Co. Ltd.

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: 30 May 2022
doi: 10.3389/fgene.2022.844653

Genome-Wide Selection Signatures and Human-Mediated Introgression Events in *Bos taurus indicus*-influenced Composite Beef Cattle

Seyed Milad Vahedi[1], *Siavash Salek Ardestani*[2], *Kian Pahlevan Afshari*[3], *Seyed Mohammad Ghoreishifar*[4], *Sima Moghaddaszadeh-Ahrabi*[5]*, *Mohammad Hossein Banabazi*[6†] *and Luiz Fernando Brito*[7]

[1]Department of Animal Science and Aquaculture, Dalhousie University, Truro, NS, Canada, [2]Department of Animal Science, Science and Research Branch, Islamic Azad University, Tehran, Iran, [3]Department of Animal Sciences, Islamic Azad University, Varamin-Pishva Branch, Varamin, Iran, [4]Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, [5]Department of Animal Science, Faculty of Agriculture and Natural Resources, Islamic Azad University, Tabriz Branch, Tabriz, Iran, [6]Department of Animal Breeding and Genetics (HGEN), Centre for Veterinary Medicine and Animal Science (VHC), Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden, [7]Department of Animal Sciences, Purdue University, West Lafayette, IN, United States

OPEN ACCESS

Edited by:
Li Ma,
University of Maryland, United States

Reviewed by:
Wai Yee Low,
University of Adelaide, Australia
Hussain Mahdi Bahbahani,
Kuwait University, Kuwait

***Correspondence:**
Sima Moghaddaszadeh-Ahrabi
s.moghaddaszadeh@iaut.ac.ir

†Present address:
Mohammad Hossein Banabazi,
Department of Biotechnology, Animal Science Research Institute of Iran, Agricultural Research, Education & Extension Organization, Karaj, Iran

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

Received: *28 December 2021*
Accepted: *09 February 2022*
Published: *30 May 2022*

Citation:
Vahedi SM, Salek Ardestani S, Pahlevan Afshari K, Ghoreishifar SM, Moghaddaszadeh-Ahrabi S, Banabazi MH and Brito LF (2022) Genome-Wide Selection Signatures and Human-Mediated Introgression Events in Bos taurus indicus-influenced Composite Beef Cattle. Front. Genet. 13:844653. doi: 10.3389/fgene.2022.844653

Genetic introgression from interbreeding hybridization of European *Bos taurus taurus* (EBT) and Indian *Bos taurus indicus* (IBI) cattle breeds have been widely used to combine the climatic resilience of the IBI cattle and the higher productivity of EBT when forming new composite beef cattle (CB) populations. The subsequent breeding strategies have shifted their initial genomic compositions. To uncover population structure, signatures of selection, and potential introgression events in CB populations, high-density genotypes [containing 492,954 single nucleotide polymorphisms (SNPs) after the quality control] of 486 individuals from 15 cattle breeds, including EBT, IBI, and CB populations, along with two *Bos grunniens* genotypes as outgroup were used in this study. Then, in-depth population genetics analyses were performed for three CB breeds of Beefmaster, Brangus, and Santa Gertrudis. Neighbor-joining, principal components, and admixture analyses confirmed the historical introgression of EBT and IBI haplotypes into CB breeds. The f_{dM} statistics revealed that only 12.9% of CB populations' genetic components are of IBI origin. The results of signatures of selection analysis indicated different patterns of selection signals in the three CB breeds with primary pressure on pathways involved in protein processing and stress response in Beefmaster, cell proliferation regulation and immune response in Brangus, and amino acids and glucose metabolisms in Santa Gertrudis. An average of >90% of genomic regions underlying selection signatures were of EBT origin in the studied CB populations. Investigating the CB breeds' genome allows the estimation of EBT and IBI ancestral proportions and the locations

Abbreviations: BMA, beefmaster; BRM, brahman; BRG, brangus; CHL, charolais; CB, composite beef cattle; DCMS, de-correlated composite of multiple signals; EBT, european *Bos taurus taurus*; EHH, extended haplotype homozygosity; FDR, false discovery rate; GOBP, gene ontology biological process; GYR, gyr; HFD, hereford; IBI, indian *Bos taurus indicus;* iHH12, integrated haplotype homozygosity pooled; iHS, integrated haplotype score; KEGG, kyoto encyclopedia of genes and genomes; NEL, nellore; OUT, outgroup (yak); PMT, piedmontese; PPI, protein-protein interaction; SGT, santa gertrudis; SNP, single nucleotide polymorphism; Z(D), Z-transformed D; $Z(f_{dM})$, Z-transformed f_{dM}.

within the genome where either taurine or indicine origin alleles are under selective pressure. Such findings highlight various opportunities to control the selection process more efficiently and explore complementarity at the genomic level in CB populations.

Keywords: indicine, taurine, composite beef cattle, population genetics, introgression, recent selection signatures

INTRODUCTION

Various cattle breeds with particular phenotypes and genetic backgrounds have been artificially selected worldwide mainly for agricultural purposes, such as milk yield, meat production, and climatic resilience (Bélanger and Pilling, 2019; Freitas et al., 2021). Two main types of modern cattle breeds of *Bos taurus taurus* (humpless taurine) and *Bos taurus indicus* (humped indicine) are the result of two domestication processes in the Fertile Crescent and the Indus Valley, respectively (Hiendleder et al., 2008; Bickhart et al., 2016). Population genetic studies based on genomic single-nucleotide polymorphism (SNP) data identified three major groups of cattle, comprising of Asian indicine, Eurasian taurine, and African taurine (Bovine HapMap Consortium, 2009; McTavish et al., 2013). However, a prominent whole-genome sequencing analysis of domestic cattle populations demonstrated that the worldwide cattle population could be classified into five continental groups based on Y-chromosome haplotypes and autosomal variants: European taurine, Eurasian taurine, East Asian taurine, Chinese indicine, and Indian indicine (Chen et al., 2018). In contrast, composite cattle breeds have been developed by human-mediated crossing of two or more breeds, in specific proportions, to combine their desirable and complementary traits into one breed (Gregory and Cundiff, 1980; Gregory et al., 1994). Once the composite breed is formed, intensive selection programs, mainly for traits valued by breeders, are applied to improve the rates of genetic gain and animal productive efficiency (Heaton et al., 2002; Grigoletto et al., 2020).

Over the last half-century, some *Bos taurus indicus*-influenced composite beef cattle (CB) breeds, such as Beefmaster, Brangus, and Santa Gertrudis, have been developed in the US to combine the climatic resilience of the Indian *Bos taurus indicus* cattle (IBI) and the higher productivity of European *Bos taurus taurus* (EBT) (Buzanskas et al., 2017; Gregory and Cundiff, 1980). Beefmaster is the first American composite breed which was developed through crossing between Brahman, Shorthorn, and Hereford on the Lasater Ranch in Falfurrias, Texas in 1908. The pedigree-estimated breed composition in Beefmaster is expected to be ½ Brahman, ¼ Hereford, and ¼ Shorthorn (Warwick, 1958). Brangus cattle was derived from crosses between Angus and Brahman cattle in Oklahoma, Mississippi, Texas, and Louisiana in the 1930s and the breed's genome content of ⅜ Brahman and ⅝ Angus is expected (International Brangus Breeders Association; https://gobrangus.com/). Santa Gertrudis cattle was developed on the King Ranch in Kingsville, Texas, by experimental crossbreeding between Shorthorn and Brahman cattle between 1910 and 1920. This breed is expected to have a genetic composition of ⅜ Brahman and ⅝ Shorthorn (Rhoad, 1949; Warwick, 1958).

One of the primary sources of genetic variability, particularly in composite breeds, is adaptive variation transmitted to the breed by introgression, a phenomenon referred to as "adaptive introgression" (Hedrick, 2013). Therefore, almost all the variants present in the populations of CB breeds will have been of introgressed origin from IBI and EBT breeds. The availability of whole-genome DNA information on a large number of individuals enables assigning ancestry along a chromosome and identify ancestry genomic regions introgressed into hybrid breeds using high-throughput genomic data (Malinsky et al., 2021). Introgression analysis of CB cattle breeds represents an appealing model for understanding the genomic consequences of crossbreeding with the goal of improving traits of interest.

The processes of domestication, breed formation, and subsequent selection have left detectable footprints within the genome of CB cattle breeds (Paim et al., 2020; Singh et al., 2020). Some of these footprints of selection (also known as selection signatures) reflect the historical selection during cattle domestication, whereas some represent selection within the past few generations for economically important traits, including meat production or environmental adaptation (Decker et al., 2014). Regarding the recent crossbreeding events in CB breeds, attention must be given to the more recent systematic, organized selection following the breed formation than old selective sweeps. Meanwhile, several extended haplotype homozygosity (EHH)-related statistics have been developed to detect more recent selection events (Sabeti et al., 2002). Moreover, the results of multiple EHH-based tests can be integrated into different composite measures of selection (Lotterhos et al., 2017). One of the well-known composite measures is the de-correlated composite of multiple signals (DCMS), proposing higher resolution and power of detecting selection signals compared to most single statistics (Ma et al., 2015) and composite measures of selection footprints (Lotterhos et al., 2017).

This study uses genomic data generated from the Illumina BovineHD SNP Beadchip (San Diego, CA, United States) to trace the ancestry components and understand the introgression processes that formed the currently observed diversity of CB breeds. Moreover, we applied the DCMS method to explore recent selection signatures in the autosomal genome of CB breeds to investigate the possible determinants of adaptation and production.

MATERIALS AND METHODS

Data Collection and Quality Control

Genotype data of 15 cattle breeds (N = 486), including nine EBT, three IBI, and three CB breeds, along with two *Bos*

TABLE 1 | Descriptive statistics for the studied cattle breeds.

Type	Breed	Abbreviation	N	Total
European *Bos taurus* (EBT)	Angus	ANG	42	315
	Brown Swiss	BSW	22	
	Charolais	CHL	37	
	Guernsey	GNS	21	
	Herford	HFD	28	
	Holstein	HOL	60	
	Jersey	JER	34	
	Limousin	LMS	50	
	Piedmontese	PMT	21	
Indian *Bos taurus indicus* (IBI)	Brahman	BRM	46	104
	Gyr	GYR	27	
	Nellore	NEL	31	
Composite beef cattle (CB)	Beefmaster	BMA	23	67
	Brangus	BRG	12	
	Santa Gertrudis	SGT	32	
Bos grunniens	Yak	OUT	2	2

grunniens samples as outgroup was obtained from the WIDDE database (Sempéré et al., 2015). Only breeds genotyped with the Illumina Bovine HD Genotyping BeadChip (www.illumina.com; San Diego, CA, United States) and that had at least 10 individuals were included in this study. **Table 1** presents the cattle breeds, their abbreviations, and number of individuals included in this study. Quality control of 732,993 SNP was performed using the *VCFtools* 0.1.16 (Danecek et al., 2011) for all genotypes. Markers with minor allelic frequency <0.01, SNP calling rate <0.90, extreme departure from Hardy-Weinberg equilibrium p-value < 10^{-7}, and SNPs located on non-autosomal chromosomes were removed. Furthermore, samples with a genotype call rate of less than 90% were discarded from downstream analyses. After the quality control, 492,954 SNPs from 488 animals remained for further analyses.

Population Structure and Phylogenetic Tree

Principal component analysis (PCA) was used to estimate genetic relationships and population structure. PCA was conducted using the PLINK 2.0 software (Purcell et al., 2007). Subsequently, a 3D plot of principal components was constructed using a custom-made Python 3.8 (http://www.python.org) script (https://github.com/Siavash-cloud/3D-PCA-plot). The phylogenetic analysis was performed using the neighbor-joining approach in VCF-kit 0.2.9 (Cook & Andersen, 2017). Visualization of the phylogenetic analysis was based on rooting of the outgroup in *FigTree* 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree). Maximum likelihood analysis of population structure of the studied cattle breeds was conducted using *Admixture* 1.3 (Alexander et al., 2009) for K values ranging from 2 to 20 with 10 iterations per K value. The *Admixture* software (Alexander et al., 2009) uses a cross-validation procedure to estimate the most likely number of ancestral populations (K). The cross-validation error estimates were plotted using the *ggplot2* package (Villanueva and Chen, 2019) in the R 4.0.5 software (R Core Team, 2020) to compare the K values.

Genomic Introgression Analyses

Patterson's D (Green et al., 2010) and the related estimate of admixture fraction f, referred to as f_4-ratio (Patterson et al., 2012) were estimated using the *Dtrios* program implemented in the *Dsuite* package (Malinsky et al., 2021) to assess the evidence for gene flow from EBT and IBI breeds into CB populations. These two statistics are applied to assess correlations of allele frequencies across populations (Patterson et al., 2012). These methods can be successfully used for learning about hybridization and introgression events within groups of closely related populations (Patterson et al., 2012; Pease and Hahn, 2015). In the Patterson's D and f_4-ratio statistics, a simple explicit phylogenetic tree model is fitted to a quartet of populations, and a formal test for a history of admixture is performed in that context (Patterson et al., 2012). The D and f_4-ratio statistics were applied to biallelic SNPs across four populations: P1, P2, P3, and OUT, related by the rooted tree [(P1,P2), P3, OUT], where yak (*Bos grunniens*) was used as the outgroup (OUT) to test if the P1 (EBT cattle) and P2 (IBI cattle) shared more alleles at the SNP level with a candidate introgressor—P3, including the three CB breeds. The site patterns were ordered as follows: BBAA represents P1 and P2 sharing the derived allele, ABBA referred to P2 and P3 sharing the derived allele, and BABA to P1 and P3 sharing the derived allele. Under the null hypothesis, which assumes no gene flow, the ABBA and BABA patterns are expected to occur due to incomplete lineage sorting with equal frequencies, and a significant deviation from that expectation is consistent with introgression between P3 and either P1 or P2. For more details, see Patterson et al. (2012) and Durand et al. (2011). Therefore, for all n biallelic sites, *nABBA*, *nBABA*, and *nBBAA* were calculated as follows (Patterson et al., 2012):

$$nABBA = \sum_{i=1}^{n} (1-\hat{p}_{i1})\hat{p}_{i2}\hat{p}_{i3}(1-\hat{p}_{iO}) + \hat{p}_{i1}(1-\hat{p}_{i2})(1-\hat{p}_{i3})\hat{p}_{iO}$$

$$nBABA = \sum_{i=1}^{n} \hat{p}_{i1}(1-\hat{p}_{i2})\hat{p}_{i3}(1-\hat{p}_{iO}) + (1-\hat{p}_{i1})\hat{p}_{i2}(1-\hat{p}_{i3})\hat{p}_{iO}$$

$$nBBAA = \sum_{i=1}^{n} \hat{p}_{i1}\hat{p}_{i2}(1-\hat{p}_{i3})(1-\hat{p}_{iO}) + (1-\hat{p}_{i1})(1-\hat{p}_{i2})\hat{p}_{i3}\hat{p}_{iO}$$

where $\hat{p}_{i1}$, $\hat{p}_{i2}$, $\hat{p}_{i3}$, and $\hat{p}_{iO}$ are the derived allele frequency estimate at site i in P1, P2, P3, and OUT, respectively. Based on Patterson et al. (2012)'s definition, the Dsuit package applies OUT as the fourth population, not necessarily an outgroup; for more details, see Malinsky et al. (2021). Therefore, the implicit assumption that the outgroup population is fixed for the ancestral allele and is necessarily used in that case is no longer applicable (Malinsky et al., 2021). Patterson's D statistics was calculated by $D = \frac{\sum_{i=1}^{n}(\hat{p}_{i2}-\hat{p}_{i1})(\hat{p}_{i3}-\hat{p}_{iO})}{\sum_{i=0}^{n}(\hat{p}_{i2}+\hat{p}_{i1}-2\hat{p}_{i2}\hat{p}_{i1})(\hat{p}_{i3}+\hat{p}_{iO}-2\hat{p}_{i3}\hat{p}_{iO})}$ (Patterson et al., 2012). To calculate the f_4-ratio, P3 was split into two subsets, P3a and P3b, and Dsuite randomly sampled from P3 alleles at each SNP. Then, f_4-ratio was calculated using $f_4\text{-ratio} = \frac{\sum_{i=1}^{n}(\hat{p}_{i3a}-\hat{p}_{iO})(\hat{p}_{i2}-\hat{p}_{i1})}{\sum_{i=1}^{n}(\hat{p}_{i3a}-\hat{p}_{iO})(\hat{p}_{i3b}-\hat{p}_{i1})}$ (Patterson et al., 2012).

To localize the introgressed loci from the breeds with the highest Z-transformed D (Z(D)) and f_4-ratio values, the *Dinvestigate* software of the *Dsuite* package (Malinsky et al., 2021) and f_{dM} statistics were applied. f_{dM} is a conservative version of the f statistic that is particularly appropriate for analysis of small genomic windows (Malinsky et al., 2018). The f_{dM} statistics (Malinsky et al., 2015) was estimated using a sliding window size of 10 SNPs and a step size of two SNPs (Wang et al., 2020) as follows:

$$f_{dM} = \frac{S(P1, P2, P3, OUT)}{-S(Pd, P2, Pd, OUT)}$$

Where $S(P1, P2, P3, OUT)$ stands for the numerator of Patterson's D, and $-S(Pd, P2, Pd, OUT)$ is equivalent to the f_{dM} numerator when $Pd = P1$ and $Pd = P3$, depending on which of P_1 or P_2 populations has the higher frequency of the derived allele. Under the null hypothesis, i.e., no introgression, the f_{dM} value is symmetrically distributed around zero, and it can equally quantify shared variation between P3 and P2 (positive values) or between P3 and P1 (negative values) (Malinsky et al., 2021).

Analysis of Recent Signatures of Selection

To detect recent signatures of selection in each CB breed, three EHH-related within-population signatures of selection tests, including integrated haplotype score (iHS; Voight et al., 2006), integrated haplotype homozygosity pooled (iHH12; Torres et al., 2018), and nSL (Ferrer-Admetlla et al., 2014) were applied using *selscan* 2.0.0 software (Szpiech and Hernandez, 2014). Prior to analyses, missing genotypes were removed using *VCFtools* 0.1.16 (genotype call rate = 100%) since selscan cannot handle missing genotypes. Genotype imputation was not performed because there were not enough genotyped individuals to accurately perform imputation, especially when considering that the CB populations have high genetic diversity. The nSL procedure calculates the SL statistic that measures the length of a segment of haplotype homozygosity in terms of segregating sites (Ferrer-Admetlla et al., 2014). Then, the three statistics of iHS, iHH12, and nSL were combined into a single DCMS framework. The approach of combining multiple selection signals in the DCMS framework has been previously performed in several studies (e.g., Yurchenko et al., 2018; Ghoreishifar et al., 2020). DCMS combines *p*-values produced by multiple statistics for each locus into a single measure considering the correlation between the statistics (Lotterhos et al., 2017). The DCMS statistic is calculated at the position *l* as follows (Ma et al., 2015):

$$DCMS_l = \sum_{t=1}^{n} \frac{\log\left[\frac{1-p_{lt}}{p_{lt}}\right]}{\sum_{i=1}^{n}|r_{it}|}$$

where p_{lt} shows the *p*-value at position *l* for statistic *t*; r_{it} refers to the correlation between the test statistic of the *i* th and *t* th methods, and *n* is the total number of test statistics combined in the DCMS. The expression $\frac{1}{\sum_{i=1}^{n}|r_{it}|}$ is called weight factor, which ranges from $\frac{1}{n}$ to 1.

Therefore, within a given CB breed, for each statistic of iHS, iHH12, and nSL, genome-wide *p*-values were calculated based on fractional ranks using the *stat_to_pvalue* function of the R *MINOTAUR* package (Verity et al., 2017). Fractional rank *p*-values were estimated using a right-tailed test for all statistics (two.tailed = FALSE, right. tailed = TRUE); then, DCMS statistics were calculated using the *DCMS* function of the R *MINOTAUR* package (Verity et al., 2017). DCMS values were then fitted to a normal distribution using the robust linear model (*rlm* function) of the *MASS* R package (Venables & Ripley, 2013) and model = rlm (dcms ~ 1), where the dcms object is a vector containing the raw DCMS values. The outputs of the fitted model, including mean and standard deviation, were used by the *pnorm* R function to calculate the *p*-values of the DCMS statistics (lower.tail = FALSE, log. p = FALSE). Finally, to control for multiple testing false discovery rate (FDR) among rejected null hypotheses, the DCMS *p*-values were transformed to the corresponding *q*-values using the *qvalue* R function and the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

We extracted the f_{dM} value of genomic regions harboring SNPs with significant selection signals (q-value < 0.05). The average f_{dM} values of regions harboring each significant SNP under selection were used to identify the ancestry. To show the overlaps between signatures of selection and introgression analyses, graphical visualization of bovine chromosomes was applied using the *chromoMap* 0.3.1 package (Anand and Lopez, 2022) in the R software (R Core Team, 2020).

Gene Annotation

After z-transforming the f_{dM} values ($Z(f_{dM})$), genomic regions ranked the highest and lowest 95 percentile of $Z(f_{dM})$ values were considered as putative introgressed genomic regions into CB breeds. This threshold was set since significant Patterson's D values indicated that on average 6.39% of bovine genome was introgressed into the CB breeds. A similar threshold level and approach have been previously applied by Barbato et al. (2020) and Morales-Cruz et al. (2021). To locate the putative regions under selection, chromosome intervals harboring SNPs with a q-value < 0.05 were considered as statistically significant intervals, and boundaries of each interval were defined by the locations of the first flanking SNPs exhibiting a q-value > 0.10. The advantage of this approach is that fewer candidate genes are obtained for the selection peaks (Yurchenko et al., 2018). Then, protein-coding genes were extracted from the significant regions based on the UMD 3.1 bovine reference genome assembly (Zimin et al., 2009). Manhattan plots of the results were created using the R package *CMplot* 3.6.2 (Yin et al., 2021).

Protein-Protein Interaction Network Construction and Functional Enrichment Analysis

The genes identified for each CB breeds were separately applied for functional enrichment and protein-protein interaction (PPI) network analyses. Prior to the analyses, duplicated genes were removed from the genes list of each breed. Functional profiling (g:GOSt) of *gProfiler* software (Raudvere et al., 2019) was used to determine terms of Gene Ontology (GO) biological processes

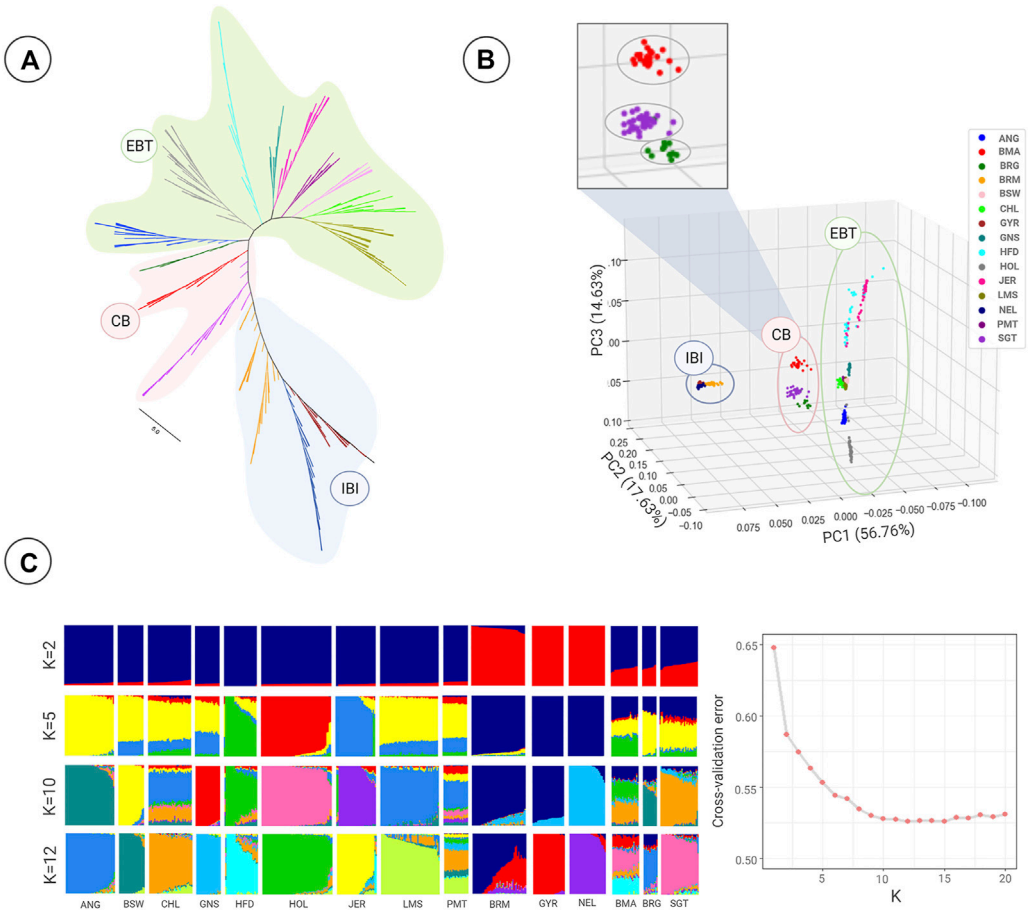

FIGURE 1 | Population structure and relationship of European *Bos taurus taurus* (EBT), Indian *Bos taurus indicus* (IBI), and composite beef cattle (CB) breeds tested in this study using genome-wide SNPs. **(A)** A neighbor-joining phylogenetic tree was constructed using whole-genome SNP data. The scale bar represents pairwise distances between different individuals. Colors reflect the different geographic regions of samples. **(B)** Principal component analysis showing PC1 against PC2 and PC3. **(C)** Model-based clustering of 15 cattle breeds using admixture analysis with the assumed number of ancestries of 2, 5, 10, and 12. The linear plot shows the cross-validation error as a function of K for admixture analysis. Abbreviations of all breeds are given in **Table 1**.

(GO:BP) or biological pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways, and Reactome databases in which the candidate genes identified in the introgression study were statistically overrepresented. The functional enrichment analysis was conducted based on the reference gene list of *Bos taurus*. Regarding candidate genes found in selection signature analyses, PPI network and functional enrichment analyses were conducted using the *Cytoscape* 3.8.2 software (Shannon et al., 2003) and the reference gene list of *Bos taurus*. The *stringApp* (Doncheva et al., 2018) of *Cytoscape* was used to import and augment networks from the STRING protein database (https://string-db.org/). Physical interactions with the confidence cutoff ratio of 0.4 (default setting) and a maximum of 20 additional interactions were tested. The *stringApp* predicts interactions based on co-expression analysis, evolutionary signals across the cattle genome, automatic text-mining, and orthology-based evidence transfer across organisms. To identify biological pathways enriched in the candidate gene lists, gene enrichment analysis was performed for GO:BP, KEGG, WikiPathways, and Reactome terms. The FDR value < 0.05 was considered as the threshold for identifying the overrepresented terms in all functional enrichment analyses.

RESULTS

A total of 488 cattle from a publicly-available database were included in this study (**Table 1**). After quality control, 492,954 SNPs from all animals remained for further analyses. The average ± standard deviation of inter-marker distance of the different breeds was 4.9 ± 8.0 Kb, and the minimum and maximum distance between SNPs were 0.1 and 36.6 Kb, respectively.

Population Genetic Structure

The phylogenetic analysis illustrated EBT and IBI breeds in two separate main branches (**Figure 1A**). We noted that CB breeds

were located at intermediate positions between these two major clades, as expected. The CB breeds gathered in a paraphyletic pattern, most likely due to the gene flow between IBI and CB. The Brangus was clustered in one clade with the Angus breed, suggesting a closer genetic relationship with EBT cattle, particularly Angus, compared to other populations (**Figure 1A**). In contrast, Santa Gertrudis demonstrated more intimate genetic relationships with the Brahman breed from the IBI group. The Beefmaster breed was also located between Santa Gertrudis and Brangus clades, with a slight affinity to the EBT clades.

Similar population affinities were obtained based on the PCA results, in which a clear genetic structure with samples from each geographical region clustering together was observed (**Figure 1B**). The first component (PC1 = 56.76%) was driven by the difference between three large clusters of EBT, IBI, and CB. However, within the CB cluster, a separation was found between Beefmaster, Brangus, and Santa Gertrudis breeds along the second (PC2 = 17.63%) and third (PC3 = 14.63%) principal components.

The ancestral lineage compositions of 15 cattle breeds from global populations are shown in **Figure 1C**. The K value, representing the number of ancestral populations, indicated that the lowest cross-validation error is K = 12. However, with regards to the history of the different populations and the practice of crossbreeding, we chose to plot the admixture results from two assumed ancestries (K = 2) to optimal K of 12 (**Figure 1C**). In K = 2, the samples were split into two groups: 1) pure EBT or IBI animals and 2) crossbred CB breeds (combination of the two ancestral populations). Results revealed the presence of EBT and IBI breeds admixture in all three CB breeds populations, confirming the principal component analysis plot. On average, Beefmaster showed 0.71 ± 0.04 and 0.29 ± 0.04 of EBT and IBI ancestry. In contrast, the average EBT and IBI ancestral contribution in Brangus was estimated at 0.69 ± 5 and 0.31 ± 0.05, respectively. In Santa Gertrudis, we estimated an average of 0.64 ± 0.03 and 0.36 ± 0.03 of EBT and IBI ancestry, respectively. Among the IBI breeds, Gyr had the highest percentage of IBI ancestry in the Beefmaster genome (0.21 ± 0.04, K = 12), whereas Brahman was the most introgressed IBI breed to Brangus genome (0.18 ± 0.05, K = 12). Furthermore, Nellore achieved the highest level of admixture in Santa Gertrudis among IBI breeds (0.07 ± 0.04, K = 12). Among EBT breeds, Hereford (0.28 ± 0.06, K = 8) and Angus (0.59 ± 0.05, K = 12) had the highest genomic contribution in Beefmaster and Brangus, respectively. Moreover, Hereford (0.16 ± 0.06, K = 12) gained the most significant introgression among EBT breeds in the Santa Gertrudis genome.

Detection of Introgression

Among 455 possible breed trios, we found 306 trios with significant Z(D) values ($p < 0.05$). Among them, 38, 44, and 56 trios revealed the significant admixture in Beefmaster, Brangus, and Santa Gertrudis breeds. We also estimated 230 trios with f_4-ratio > 0.1%, among which 37, 36, and 51 trios demonstrated the gene flow from EBT and IBI populations to Beefmaster, Brangus, and Santa Gertrudis breeds, respectively.

Beefmaster

Among 29 significant Z(D) values indicating introgression events with EBT breeds, the highest value belonged to the Hereford breed while the Piedmontese was considered as P1 ((((PMT,BMA),HFD),OUT); D = 0.100; Z(D) = 30.020). This result was confirmed by f_4-ratio as the highest f_4-ratio value belonged to Hereford (((((((ANG,HFD),BMA),OUT); f_4-ratio = 8.450). In contrast, the most significant Z(D) value representing introgression events with IBI breeds was obtained from the trio of (((HFD,BMA),GYR),OUT); D = 0.015; Z(D) = 9.433). In parallel, the Gyr population possessed the highest f_4-ratio value ((((HFD,BMA),GYR),OUT); f_4-ratio = 0.154) among those showing IBI introgression. Our results indicated clear evidence of genome-wide level introgression from Hereford and Gyr breeds to the Beefmaster population, which is consistent with the results of PCA and admixture analyses.

The $Z(f_{dM})$ values were estimated using the trio of (((HFD,GYR),BMA),OUT). The top 5% genomic regions introgressed from Hereford (n = 6,318) and Gyr (n = 5,288) into Beefmaster were identified, among which the highest and lowest $Z(f_{dM})$ values were obtained by genomic windows on BTA16 (16:40.21–40.26 Mb; $Z(f_{dM})$ = 4.415) and BTA24 (24: 0.10–0.16 Mb; $Z(f_{dM})$ = −2.218), respectively (**Figure 2**). Regions with $f_{dM} > 0$ and $f_{dM} < 0$ comprised 11.7% and 88.3% of the Beefmaster genome. The gene annotation analyses detected 14 and eight candidate genes potentially transferred from Hereford and Gyr ancestry into Beefmaster by the top 5% introgressed genomic regions. The gene enrichment analyses detected three biological pathway terms in which the candidate genes potentially introgressed from Gyr into Beefmaster were significantly over-represented (**Supplementary Table S1**). In contrast, no significant pathway was detected for candidate genes with Hereford ancestry.

Brangus

Among the significant Z(D) values, the highest value was obtained by Angus when Charolais was considered as P1 (((CHL,ANG),BRG),OUT); D = 0.087; Z(D) = 29.092). The highest f_4-ratio among EBT breeds was obtained by the same trio ((((CHL,ANG),BRG),OUT); f_4-ratio = 2.839). In contrast, the highest Z-score among Z(D) values representing introgression events from IBI to CB breeds was achieved by Brahman ((((HFD,BRG),BRM),OUT); D = 0.015; Z(D) = 6.417). Four trios with f_4-ratio > 0.1% showed introgression with IBI breeds, out of which the highest value (f_4-ratio = 0.220) belonged to Brahman. Our results of Patterson's D and f_4-ratio was in parallel with the results of PCA and admixture analyses.

The top 5% genomic regions introgressed from Angus (n = 1,198) and Brahman (n = 982) into Brangus were identified (**Figure 2**). The highest and lowest $Z(f_{dM})$ values were obtained by genomic windows on BTA6 (6:77.64–77.66 Mb; $Z(f_{dM})$ = 4.444) and BTA16 (16:34.27–34.28 Mb; $Z(f_{dM})$ = −2.135). Estimation of the f_{dM} of Brangus cattle using the tree topology

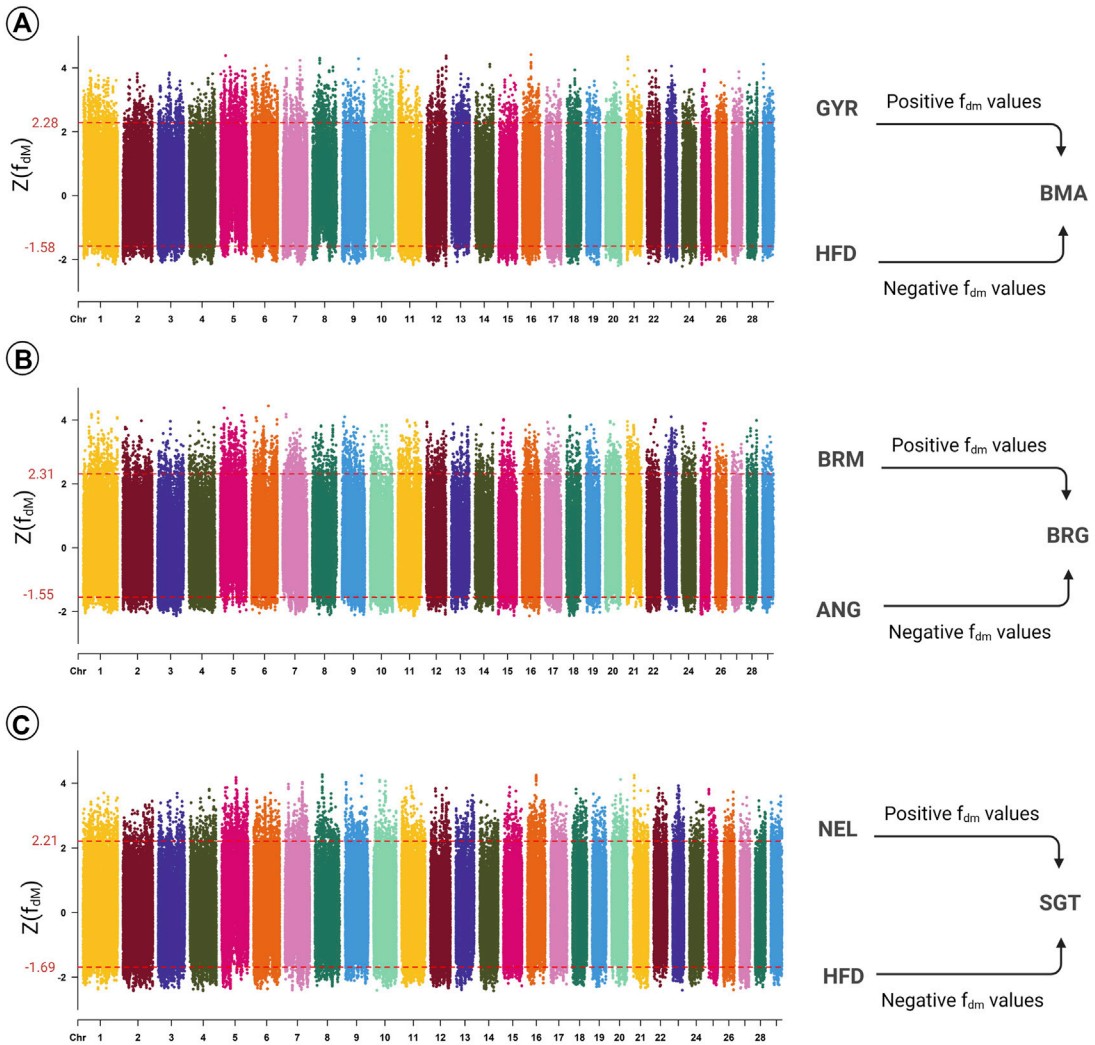

FIGURE 2 | Manhattan plot of the Z-transformed f_{dM} ($Z(f_{dM})$) at the significance threshold of 5% in Beefmaster **(A)**, Brangus **(B)**, and Santa Gertrudis **(C)**. The $f_{dM} > 0$ represents regions with Indian *Bos taurus indicus* (IBI) ancestry, while $f_{dM} < 0$ represents regions with European *Bos taurus* (EBT) origin. Abbreviations of all breeds are given in **Table 1**.

(((ANG,BRM),BRG),OUT) revealed that the introgressed regions with $f_{dM} > 0$ and $f_{dM} < 0$ formed 11.8% and 88.2% of the Brangus genome, respectively. The gene annotation analyses revealed that 581 and 10 candidate genes were potentially transferred by the top 5% introgressed genomic regions from Angus and Brahman into Brangus. Based on the gene enrichment analysis, we found 13 and two biological processes and pathway terms in which the candidate genes introgressed from Angus and Brahman were overrepresented (**Supplementary Table S1**).

Santa Gertrudis

Among the significant Z(D) values, the highest value belonged to the Hereford breed (((BRG,HFD), GT),OUT); D = 0.104; Z(D) = 26.542). This result was further confirmed by f_4-ratio test (((CHL,HFD),SGT),OUT); f_4-ratio = 10.280). On the contrary, 10 significant Z(D) values were estimated for IBI breeds, among which the Nellore breed obtained the highest Z-score ((((HFD,SGT),NEL),OUT); D = 0.016; Z(D) = 9.830). In agreement with this result, Nellore had the highest value of f_4-ratio among those representing IBI introgression (f_4-ratio = 0.146). Therefore, in agreement to the admixture results, the introgression assessment of Patterson's D and f_4-ratio clearly depicted the higher levels of genome-wide introgression from Hereford and Nellore into the Santa Gertrudis breed.

The top 5% genomic regions introgressed from Hereford (n = 1,237) and Nellore (n = 1,085) were determined (**Figure 2**). The highest and lowest $Z(f_{dM})$ values were obtained by genomic windows on BTA8 (8:8:33.30–33.32 Mb; $Z(f_{dM})$ = 4.264) and BTA14 (14:14:64.37–64.42 Mb; $Z(f_{dM})$ = −2.421). The estimation of the f_{dM} values using the trio of (((HFD,NEL),SGT),OUT) showed that the introgressed regions with $f_{dM} > 0$ and $f_{dM} < 0$ contributed to 15.3% and 84.7% of the Santa Gertrudis genome, respectively. The gene annotation analyses revealed that 42 and 15 candidate genes were potentially transferred by the top 5%

TABLE 2 | Genomic regions detected by the DCMS analyses as being under putative selection in Beefmaster (BMA), Brangus (BRG), and Santa Gertrudis (SGT) breeds.

Chr[a]	Position (Mb)	Q-value	Breed	Gene Id	Associated trait(s) in cattle
BTA2	90.41–90.49	0.001	BRG	*STRADB*	Immune response to intestinal parasite
				C2CD6	Marbling score
BTA3	80.75–80.92	0.001	SGT	*JAK1*	Residual feed intake, immune system response to *Mycobacterium bovis*
BTA4	25.26–25.26	0.029	BRG	*AGR2*	Immune response to intestinal parasites, heat tolerance
BTA4	103.31–103.32	0.018	BRG	*KIAA1549*	Conformation
BTA5	36.04–36.05	0.023	SGT	*NELL2*	Puberty
BTA6	81.55–81.56	0.000	BRG	*TECRL*	Puberty, feed conversion ratio
BTA6	103.80–103.81	0.011	SGT	*AFF1*	Puberty
BTA6	104.09–104.13	0.014	SGT	*NUDT9*	Uniformity of yearling weight
BTA7	35.69–35.70	0.017	SGT	*HSD17B4*	Backfat thickness
BTA7	35.83–35.85	0.001	SGT	*TNFAIP9*	—
BTA7	36.01–36.05	0.001	SGT	*DMXL1*	—
BTA7	63.98–64.02	0.030	SGT	*SYNPO*	Meat tenderness
BTA7	64.41–64.44	0.014	SGT	*CCDC69*	—
BTA7	64.72–64.73	0.043	SGT	*FAT2*	Placentation
BTA7	65.07–65.09	0.008	SGT	*GLRA1*	Myoclonus
BTA8	8.96–8.97	0.028	BMA	*MSRA*	Antioxidant defense and lifespan
BTA10	69.77–69.96	0.042	BRG	*EXOC5*	Infectious hoof lesions
				AP5M1	Infectious hoof lesions
BTA10	70.85–71.06	0.002	BRG	*ARID4A*	Bull fertility
				TIMM9	—
				KIAA0586	—
BTA11	10.56–10.58	0.000	BRG	*TET3*	—
BTA11	11.66–11.75	0.000	BRG	*EXOC6B*	Ketosis susceptibility
BTA11	12.92–12.94	0.000	BRG	*DYSF*	Development of the hind quarter, fertility
BTA11	13.55–13.59	0.002	BRG	*CD207*	Ketosis susceptibility
				CLEC4F	Ketosis susceptibility
BTA11	14.35–14.38	0.024	BRG	*SRD5A2*	Ketosis susceptibility, sperm motility
BTA11	26.48–26.73	0.038	SGT	*SLC3A*	—
				PREPL	Marbling score
				CAMKMT	—
BTA11	31.25–31.26	0.049	SGT	*FSHR*	Multiple birth
BTA11	32.16–32.17	0.046	SGT	*NRXN1*	Temperament
BTA11	36.80–36.81	0.008	SGT	*ACYP2*	Carcass and bone weight, feed efficiency, arthrogryposis, macroglossia
BTA11	38.70–38.71	0.011	SGT	*CCDC85A*	—
BTA11	101.28–101.29	0.000	BRG	*LAMC3*	Tick resistance, backfat thickness, conception rate
BTA13	25.46–25.47	0.029	SGT	*KIAA1217*	Gestation length
BTA13	25.80–25.82	0.030	SGT	*ARHGAP21*	Calving-to-first service interval, non-return after 56 days, puberty, post-partum anoestrus
BTA13	26.88–26.96	0.041	SGT	*MY O 3A*	Fertility
				GAD2	Fertility, average daily feed intake, meet percent
BTA13	43.59–43.61	0.048	SGT	*UCN3*	Heat tolerance and oxidative stress
BTA13	44.93–44.95	0.003	SGT	*KLF6*	Body conformation, intramuscular fat percentage
BTA13	72.79–72.81	0.030	SGT	*SRSF8*	—
BTA15	62.05–62.06	0.027	BRG	*MPPED2*	Muscle growth and development
BTA15	64.11–64.12	0.003	BRG	*EIF3M*	Fertility
				CCDC73	—
BTA15	65.01–65.04	0.002	BRG	*KIAA1549L*	—
BTA15	65.56–65.58	0.027	BRG	*ABTB2*	Slaughter weight
BTA15	67.69–67.70	0.042	BRG	*PRR5L*	Bovine respiratory disease susceptibility, conceptus development
BTA17	10.22–10.23	0.028	BMA	*ARHGAP10*	Intramuscular fat formation
BTA17	29.93–30.01	0.028	BMA	*LARP1B*	—
BTA17	30.09–30.18	0.028	BMA	*ABHD18*	—
				MFSD8	-—
BTA17	30.18–30.21	0.028	BMA	*PLK4*	—
BTA17	30.25–30.33	0.043	BMA	*HSPA4L*	Thermal stress
				SLC25A31	Semen quality
				INTU	—
BTA23	7.29–7.31	0.041	SGT	*COL11A2*	Marbling score, heifer pregnancy
BTA24	21.20–21.21	0.021	SGT	*MOCOS*	Yearling weight
BTA26	47.22–47.23	0.033	BRG	*DOCK1*	Flight speed
BTA29	36.70–36.71	0.029	SGT	*PRDM10*	Post-partum anoestrus

[a]Bos taurus *chromosome (BTA).*

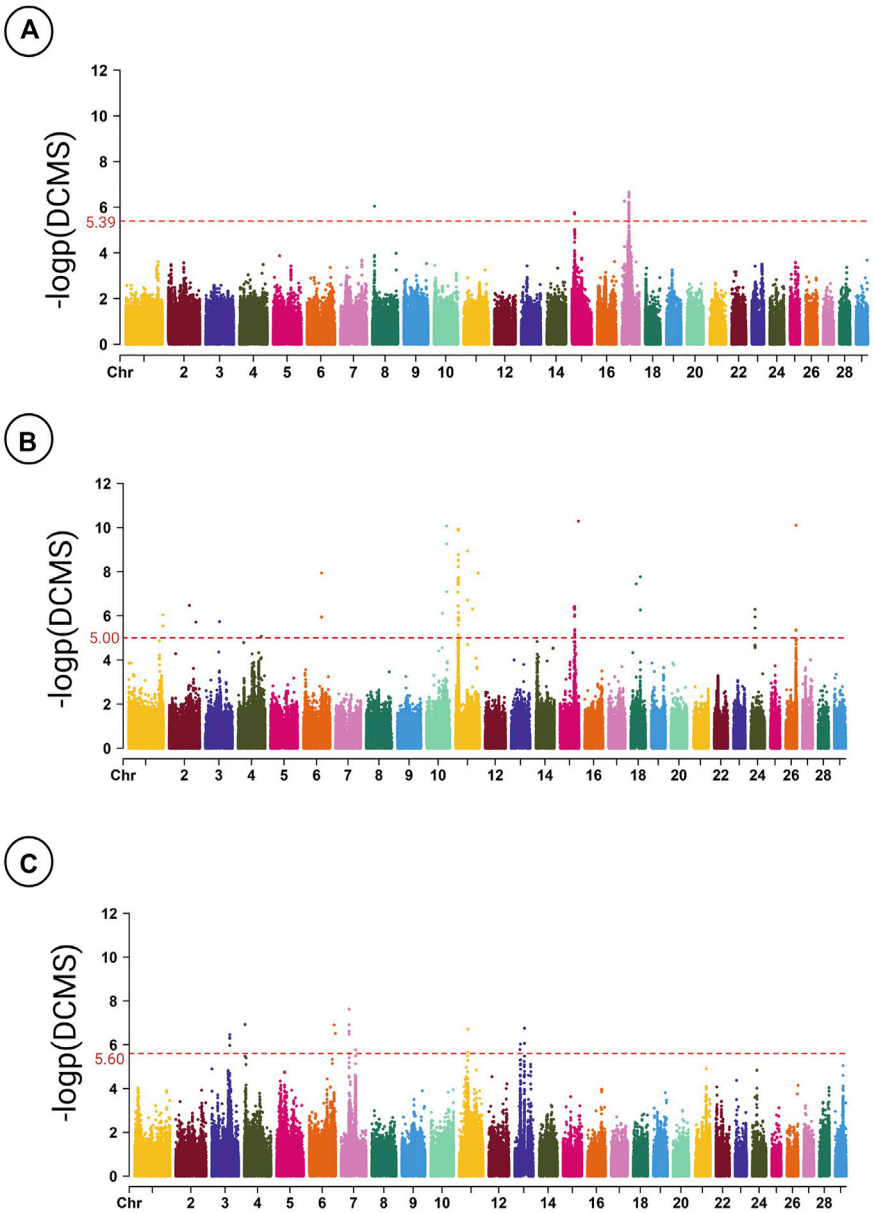

FIGURE 3 | Manhattan plot of the genomic regions under putative selection detected by DCMS in Beefmaster **(A)**, Brangus **(B)**, and Santa Gertrudis **(C)**. The dashed lines represent the significant threshold level at the FDR adjusted *p*-value < 5%.

introgressed genomic regions from Hereford and Nellore into Santa Gertrudis. In total, we identified 12 terms in which the introgressed genes from Nellore to Santa Gertrudis were significantly overrepresented (**Supplementary Table S1**). One term was also found significant overrepresented by introgressed genes from Hereford to Santa Gertrudis (**Supplementary Table S1**).

Recent Signatures of Selection Signals

After obtaining DCMS statistics for 492,954 SNPs, *p*-values were fitted to a normal distribution and corrected for multiple testing. Fourteen, 73, and 85 genomic regions with significant selection signals were found for Beefmaster, Brangus, and Santa Gertrudis, respectively (q-value < 0.05). The average ± standard deviation length of the regions under selection were 24.5 ± 14.1, 17.0 ± 14.9, and 28.5 ± 15.4 Kb in the Beefmaster, Brangus, and Santa Gertrudis breeds, respectively. The total length of regions under significant selection pressure was 299.3, 783.7, and 823.1 Kb in Beefmaster, Brangus, and Santa Gertrudis, respectively. Overall, nine, 24, and 28 candidate genes were identified in the genomic regions underlying selection in the Beefmaster,

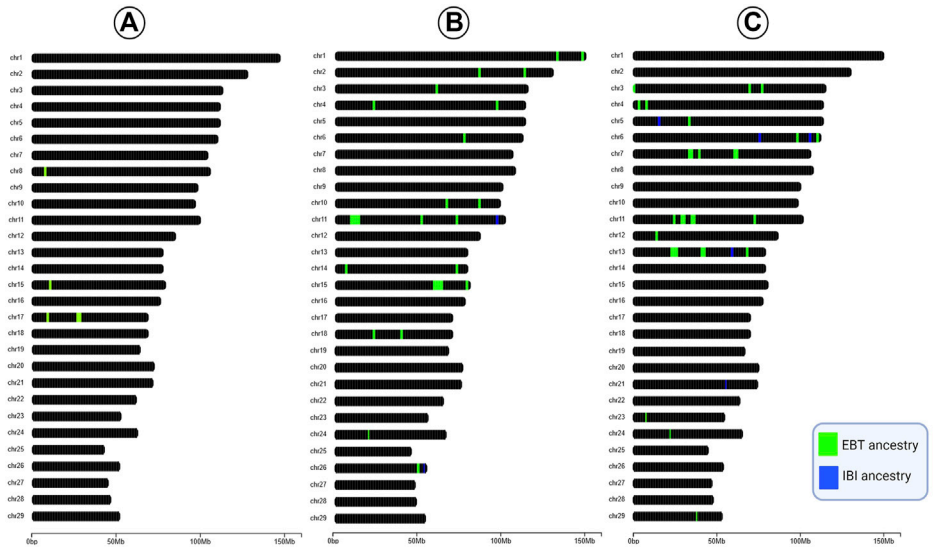

FIGURE 4 | Graphical visualization of the cattle chromosomes with regions underlying significant selection and their origin in Beefmaster **(A)**, Brangus **(B)**, and Santa Gertrudis **(C)**.

Brangus, and Santa Gertrudis breeds (**Table 2**). None of the candidate genes were detected in more than one breed.

The distribution of regions under selection across the genome of CB breeds is depicted in **Figure 3**. The most significant genomic region in Brangus was detected on BTA11 (11:12.91–12.93 Mb; q-value = 0.00002), whereas in Beefmaster, one region on BTA8 and four regions on BTA17 were found with the lowest q-value of 0.028. In Santa Gertrudis, the most significant region is located on BTA7 (7:35.82–35.84 Mb) with q-value = 0.00055.

Gene annotation analysis of the regions underlying selection revealed various previously-reported and novel candidate genes that are associated with a diverse range of traits, including reproduction, susceptibility to infectious diseases, meat quality, thermotolerance, susceptibility to metabolic diseases, genetic diseases, puberty, feed efficiency, carcass quality, sex determination, and temperament (**Table 2**). In the PPI network analysis, nine, 24, and 26 identifiers were uploaded from the STRING database for Beefmaster, Brangus, and Santa Gertrudis. PPI networks including 28, 33, and 40 genes with significant interactions were constructed for Beefmaster, Brangus, and Santa Gertrudis (**Supplementary Figure S1**). Gene enrichment analysis revealed that in the Beefmaster breed, the discovered candidate genes were significantly overrepresented in chaperon-mediated protein folding term (GO:0061077), along with three terms related to heat stress response (BTA-2262752, BTA-3371453, bta04141). Regarding the Brangus breed, we found multiple significant pathways associated with immune system function (bta04666, bta04664, bta04662, bta04650), metabolism (BTA-194840, bta04024, GO:0007265), intracellular signaling (bta04014, bta04310, bta04015), and muscle development (bta04810). In the Santa Gertrudis breed, we detected significant metabolic pathways, including lysin (bta00310), beta-alanine (bta00410), pyruvate (bta00620), and arginine and proline (bta00330).

Figure 4 shows the overlaps between the results of signatures of selection and introgression analyses. In the Beefmaster breed, all 112 SNPs with significant selection signal had $f_{dM} < 0$, whereas in Brangus, among 113 significant SNPs identified in DCMS, 104 (92.04%) and nine (7.96%) had $f_{dM} < 0$ and $f_{dM} > 0$, respectively. Considering Santa Gertrudis, we found 111 significant markers in DCMS of which 91 (81.98%) and 20 (18.02%) had $f_{dM} < 0$ and $f_{dM} > 0$, respectively.

DISCUSSION

We investigated patterns of ancestry in three *Bos taurus indicus*-influenced CB cattle breeds using high-density genome-wide polymorphism data. Our results confirmed the hybrid nature of three CB cattle breeds of Beefmaster, Brangus, and Santa Gertrudis. We identified signals of indicine and taurine ancestry at genes that are primarily involved in the domestication and adaptation process. Our analyses of the recent signature of selection indicated that genome regions underlying selection pressure among CB breeds are different and are mainly of EBT origin.

Introgression Events

In neighbor-joining tree and principal component analyses, CB breeds were placed between IBI and EBT populations. The admixture results in K = 2 indicated more indicine genetics (36%) for SGT than BMA (28%) and BRG (30%). In parallel, neighbor-joining tree showed that SGT is closer to indicus ancestry compared with BMA and BRG. Both admixture and phylogenetic analyses were consistent with introgression analysis, which indicated less taurine origin (84.7%) for SGT than that for Brangus (88.2%) and Beefmaster (88.3%). The analyses of admixture, f_4-ratio, and D-statistic clearly showed gene flow

from different IBI and EBT breeds to CB populations. Our results were consistent with historical pedigree information of the introgression history of the hybrid cattle breeds (Ritchie, 2009). Using D, f_4-ratio, and f_{dM} statistics, we observed that, on average, only 12.9% of the studied CB breeds' genome was introgressed from indicine origin, which is in agreement with the results reported by McTavish et al. (2013), who indicated that only 11% of recently formed hybrid breeds' genome is of indicine ancestry. Moreover, Barbato et al. (2020) reported that approximately 11–13% of indicine ancestry were recorded in three Central Italian breeds of Marchigiana, Romagnola, and Chianina. Our admixture results showed a higher level of IBI introgression (average of 32.0%) in the studied CB population. The explanation for this discordance might be that a historical model is not fitted in admixture analysis, and it is unrealistically assumed that all populations have derived from a single ancestral group (Alexander et al., 2009). On the contrary, D, f_4-ratio, and f_{dM} statistics include fitting a simple phylogenetic tree model to a group of populations, and they provide a formal test for a history of admixture in that context (Patterson et al., 2012). Therefore, a more accurate estimation of introgression would be feasible (Malinsky et al., 2021).

Using the f_{dM} statistics we inferred genomic local ancestry in CB breeds. In all chromosomes, some segments ranked the highest 95 percentile or the lowest 95 percentile. Barbato et al. (2020) identified multiple introgressed genomic regions in three Italian breeds in all chromosomes except in BTA17 and BTA 28. However, they only searched for indicine-driven introgressed regions. Five genes were common among the candidate genes with EBT ancestry (**Supplementary Figure S2**). Among them, Neurexin 3 (*NRXN3*) is involved in the genetic architecture of cattle temperament (Paredes-Sánchez et al., 2020). *NRXN3* is a member of the Neurexins family that are cell adhesion molecules in the nervous system to specify and stabilize excitatory and inhibitory synapses (Aoto et al., 2013). Another gene from this family, Neurexin 1 (*NRXN1*), was shared between candidate genes with EBT ancestry in Brangus and Santa Gertrudis. This gene has been proposed as a candidate gene for a moderate temperament of cattle (Qanbari et al., 2014). The *Bos taurus indicus* crosses, particularly Brahman crosses, have been reported to be more temperamental than EBT breeds but less aggressive/fearful than IBI breeds (Hoppe et al., 2010). Therefore, the inheritance of alleles with EBT ancestry may contribute to less fearful or excitable animal's response to handling or forced movement by humans. Another candidate gene, TSC22 Domain Family Member 1 (*TSC22D1*), encodes a member of the TSC22 domain family of leucine zipper transcription factors and belongs to the large family of early response genes, which are activated rapidly in response to a wide range of cellular stimuli (Kester et al., 1999). This gene has been found to be associated with maternal lipomatous myopathy (Peletto et al., 2017) and calving difficulty (Purfield et al., 2020) in cattle. It has been generally assumed that IBI cattle have less calving difficulty than EBT breeds due to their pelvic structure. However, the greater emphasis on growth rate has increased the risk of calving difficulties in CB breeds (Morrison et al., 1989). Two candidate genes of CASP2 and RIPK1 Domain Containing Adaptor with Death Domain (*CRADD*) and Sodium/Potassium Transporting ATPase Interacting 2 (*NAKIN2*) were shared among candidate genes with IBI ancestry. *CRADD* encodes a death domain-containing protein that can induce cell apoptosis (Ahmad et al., 1997). This gene is involved in cattle embryonic development, a process in which apoptosis plays a critical role in adapting myometrial cells during pregnancy (Rehman et al., 2003). In contrast, the function of *NAKIN2* in cattle is not comprehensively understood yet.

In all CB populations, gene enrichment study of genes with IBI ancestry showed the detected candidate genes were significantly involved in biological processes associated with apoptosis regulation by tumor suppressor TP53 (**Supplementary Table S1**). TP53 is upregulated in response to several stimuli, including activation of oncogenes, DNA damage, or nutrient deprivation (Aubrey et al., 2018). Subsequently, apoptosis will be induced through intrinsic or extrinsic pathways (Aubrey et al., 2018). The *TP53* gene and its pathways are associated with some traits in cattle, such as heifer fertility (Neupane et al., 2017) and carcass weight (Carvalho et al., 2019). Carcass weight and quality are usually higher in EBT breeds than IBI's (Highfill et al., 2012). Indeed, cattle with IBI influence tend to have lower meat tenderness and growth rate than EBT (Wheeler et al., 2010; Elzo et al., 2012). Therefore, increases in the proportion of alleles with EBT origin in the crossbred progeny could result in higher meat quality and yield. Santa Gertrudis and Beefmaster genes with IBI ancestry are significantly involved in the Class C/3 (Metabotropic glutamate/pheromone receptors) pathway. Cattle pheromones are crucial in reproduction and social behaviors, e.g., sexual attraction, mother-young interactions, estrus indication, puberty acceleration, reducing the post-partum anestrus, hormonal stimulation, and erection (Mucignat-Caretta, 2014). Therefore, they could be critical factors in animals' fitness and reproductive success.

Santa Gertrudis candidate genes with EBT ancestry were significantly overrepresented in the neural crest differentiation process (**Supplementary Table S1**). Neural crest cells have the potential to differentiate into several different cell types of the vertebrate body. They are the origin of most parts of the peripheral nervous system, endocrine cells in the adrenal medulla and thyroid, the frontal part of the head, and parts of the cardiovascular system (for a review, please see Dupin et al., 2006). It has been hypothesized that a reduction in neural crest cell proliferation and migration is an essential genetic mechanism of early domestication (Wilkins et al., 2014). Moreover, further changes to the neural crest may potentiate the later evolution of other domestication traits (Wright et al., 2020). Oxidation by cytochrome P450 was the most significant pathway in which Brangus candidate genes with EBT ancestry were overrepresented. Overall, five genes were overrepresented in the Oxidation by cytochrome P450 WikiPathway term which includes 39 genes in cow. Drugs, toxins, and many foreign hydrophobic compounds are removed from the body by cytochrome P450 enzyme system, particularly in the liver (Moubarak & Rosenkrans, 2000). Moreover, the critical role of P450 genes in oxidative stress, inflammatory-based diseases, and synthesis and metabolism of sterols, steroid hormones, and lipid

biofactors such as eicosanoids, vitamin D3, and retinoids in mammals is well-documented (Omura, 1999; Kuhn et al., 2020). Therefore, understanding the inheritance of cytochrome P450 oxidation functional elements from EBT breeds might be essential in veterinary practice, nutrition, and metabolic priorities of CB breeds.

Recent Signatures of Selection

To the best of our knowledge, this study is the first report of signatures of selection in the Beefmaster breed. In this breed, DCMS found three locations on the bovine genome with the most significant peaks (q-value = 0.02809), including two locations on BTA17 (17:10.22–10.23 Mb and 17:29.93–30.18 Mb) and one on BTA8 (8:8.97–8.97 Mb) (**Figure 3**). We found six candidate genes in these regions, including Methionine Sulfoxide Reductase A (*MSRA*), Rho GTPase Activating Protein 10 (*ARHGAP10*), Polo Like Kinase 4 (*PLK4*), La-related protein 1 (*LARP1B*), Abhydrolase Domain Containing 18 (*ABHD18*), and Major Facilitator Superfamily Domain Containing 8 (*MFSD8*), of which the function of the last three genes in CB breeds are not comprehensively known. Oxidation of proteins by reactive oxygen species can occur in oxidative stress, such as heat stress and several diseases. Free and protein-bound methionine residues are susceptible to oxidation to methionine sulfoxide derivatives (Moskovitz et al., 2001). However, *MSRA* can repair this modification by catalyzing the thioredoxin-dependent reduction of free and protein-bound Met(O) to methionine (Moskovitz et al., 2001). Therefore, *MSRA* plays a vital role in stress response in mammals. *ARHGAP10* is also a member of the Rho-GTPase activating protein (Rho-GAP) family, regulating Rho-GTPase signaling pathways, and these pathways are involved in actin cytoskeleton dynamics, cell proliferation, and differentiation (Bassères et al., 2002). A recent study on Japanese black cattle indicated that this gene might be involved in intramuscular fat formation (Ueda et al., 2021). The PPI network analysis of Beefmaster candidate genes indicated the critical role of Heat Shock Protein Family A Member 4 Like (*HSPA4L*) with contribution into three biological pathways (**Supplementary Figure S1**). *HSPA4L* is a member of the HSP70 gene family, the largest and the most conserved protein family throughout evolution (Daugaard et al., 2007). HSP70 genes function in heat tolerance, protection of cells against apoptosis, and reactive oxygen species (Feder & Hofmann, 1999). A recent transcriptome study in cattle revealed that the expression of *HSPA4L* is upregulated in the milk somatic cells during heat challenge (Garner et al., 2020). Gene enrichment analysis revealed that Beefmaster candidate genes were significantly overrepresented in pathways related to post-translation protein modifications and stress response (**Supplementary Figure S1**). It is noteworthy that Beefmaster has been widely used not only for their heat tolerance and adaptability but excellent growth and carcass quality (Cartwright, 1970; Wheeler et al., 2010).

Regarding the Brangus breed, we identified 73 genomic windows underlying selection, of which the most significant signal was found on BTA11 (11:12.92–12.94 Mb; q-value = 1.02E-06) (**Figure 3**). The well-known candidate gene of Dysferlin (*DYSF*) is located in this region. Dysferlin, known as dystrophy-associated fer-1-like protein, is encoded by the *DYSF* gene (Blandin et al., 2012). This calcium-dependent transmembrane protein is mainly expressed in skeletal and cardiac muscles to enhance calcium-mediated membrane fusion and sarcolemmal repair (Blandin et al., 2012). Mutations in this gene cause multiple different phenotypes of muscular dystrophies in humans (Vilchez et al., 2005). In contrast, studies in cattle demonstrated its strong association with the development of cattle muscularity (Doyle et al., 2020), sire conception rate (Rezende et al., 2018), and female reproductive performance (Fonseca et al., 2020). A recent study conducted by Zhang et al. (2020) indicated that this gene might be under strong positive selection for high altitudes adaptation in Chinese indigenous cattle. We could not find any overlapping genomic regions between our results and a recently-conducted signatures of selection study in the Brangus breed by Paim et al. (2020). The reason could be that we investigated recent signatures of selection, i.e., those regions underlying selection following the breed formation, using multiple EHH-based methods. While runs of homozygosity, which is an indicator of genomic autozygosity, may arise due to several population phenomena like inbreeding, genetic drift, consanguineous mating, population bottleneck, as well as natural and artificial selection (Falconer and Mackay, 1996; Curik et al., 2014).

PPI network analysis of Brangus candidate genes revealed significant interactions with the Rac subfamily of the Rho family of GTPases, particularly RAC1, RAC2, and RAC3. Moreover, in PPI network analysis, we found multiple pathways which are connected to these genes, including Ras signaling pathway, regulation of actin cytoskeleton, Rho GTPase cycle, Rap1 signaling pathway, cAMP signaling pathway, and Ras protein signal transduction. Ras signaling is a critical intracellular signaling pathway that plays a vital role in cellular proliferation and differentiation, survival, and gene expression (Vojtek and Der, 1998). Rho GTPases of the Ras superfamily and Rap1 protein act as molecular switches to control a wide range of essential biochemical pathways in response to environmental stimuli through intracellular signal transduction pathways (Bar-Sagi and Hall, 2000). It is noteworthy that Rap1 activation is mediated by several second messengers, such as cAMP (Zwartkruis and Bos, 1999). Studies on cattle have shown the importance of Ras signaling pathway in meat quality (Li et al., 2020) and subcutaneous fat deposition (Taniguchi et al., 2008), Rap1 signaling pathway in muscle development (Zhan et al., 2018), and cAMP signaling pathway in lipid metabolism (Junjvlieke et al., 2020).

To the best of our knowledge, signatures of selection analysis have not been previously conducted on Santa Gertrudis. We identified 85 genomic regions with significant selection signals in this breed (**Figure 3**). The most significant region was located on BTA7 (7: 35.83–35.85 Mb; q-value = 0.00055). We found a candidate gene of Tumor Necrosis Factor-Alpha Inducible Protein 9 (*TNFAIP9*) in this region. *TNFAIP9* is a member of the STEAP (six transmembrane epithelial antigen of prostate) family vital for cellular iron uptake and homeostasis (Han et al., 2018). Studies in humans suggested that *TNFAIP9* might be involved in adipocyte development and metabolism (Gui et al., 2016). However, its function in cattle has not been comprehensively understood yet. We also detected several genomic regions with significant selection signals on BTA11 (11: 31.25–38.71 Mb) and BTA13 (13:25.80–44.95 Mb). In gene annotation study of these regions, several candidate genes have been found related to reproductive system function in cattle, including Follicle Stimulating Hormone Receptor (*FSHR*)

(Widmer et al., 2021), *KIAA1217* (Mohammadi et al., 2020), Rho GTPase Activating Protein 21 (*ARHGAP21*) (Wolf et al., 2021), Myosin IIIA (*MY O 3A*) (Mohammadi et al., 2020), and Glutamate Decarboxylase 2 (*GAD2*) (Mohammadi et al., 2020). Fertility is an essential element in beef cattle production because it directly relates to producing the offspring necessary to offset costs in production systems (Thundathil et al., 2016). This trait has been improved in CB breeds during the last decades through assisted reproductive technologies of artificial insemination and genetic selection (Cammack et al., 2009). We also detected a significant signal on BTA 3 (3: 80.75–80.92 Mb). A larger overlapped region has been previously detected to be under significant selection in the Sheko breed (3:80.10–80.93 Mb) (Bahbahani et al., 2018). We detected the candidate gene of Janus Kinase 1 (*JAK1*) in this region which is involved in cattle residual feed intake (Xi et al., 2015) and immune response to *Mycobacterium bovis* (Imai et al., 2003).

The PPI network analysis revealed the central role of two members of the aldehyde dehydrogenase (ALDH) gene superfamily, including Aldehyde Dehydrogenase 3 Family Member A2 (*ALDH3A2*) and Aldehyde Dehydrogenase 7 Family Member A1 (*ALDH7A1*). Recent studies have indicated the role of *ALDH7A1* in feed efficiency (De Oliveira et al., 2014) and growth rate (Buzanskas et al., 2014), but less information is available about the *ALDH3A2* role. Gene enrichment analysis indicated that Beefmaster candidate genes were significantly overrepresented in several metabolic pathways, particularly of amino acids metabolism (**Supplementary Figure S1**). Amino acids are the principal nutrition component for protein synthesis and meat, and rapid rates of muscle protein deposition is positively correlated with growth and efficient beef production. Lysin is a limiting amino acid for optimizing the growth of certain animals such as pigs and poultry, and it is one of the first three limiting amino acids (methionine, lysine, and threonine) in growing cattle diet (Richardson and Hatfield, 1978). Moreover, lean growth rate in cattle is positively associated with overall feed efficiency in growing and finishing animals (Archer et al., 1997). Therefore, the lysin metabolism might play an important role in selection of feed efficient animals with high growth rate.

Integration of Recent Signatures of Selection and Introgression Events

The overlap between the results of signatures of selection and introgression analyses deciphered that the selected regions in studied CB breeds were predominantly EBT in origin (**Figure 4**). However, the proportion of EBT ancestry in selected regions is higher in Beefmaster (100%) than Brangus (92.04%) and Santa Gertrudis (81.98%). Our results complement the previous study of Paim et al. (2020) that showed that homozygous regions in Brangus are mainly of Angus ancestry. Our results show how the combination of selection and complementary can shift the genetic architecture of CB populations following the breed formation.

Although a reasonable number of genotyped animals were available for this study, the sample size is still a limiting factor. Therefore, larger datasets could enable accurate genotype imputation analyses and therefore, the inclusion of a larger number of SNP markers in the analyses. In addition, it would be recommended to add other breeds that might have contributed to the formation of the three CB analyzed as well as other composite populations such as the Montana tropical Composite (Grigoletto et al., 2020).

CONCLUSION

Our study revealed human-mediated introgression events and genomic regions underlying selection in three CB breeds. We confirmed the low contribution of alleles with IBI origin in the CB cattle genome. The majority of selected genomic regions in CB cattle breeds came from EBT that can be in conjunction with the traits of interest for genetic improvement and selection. Our results demonstrate how complementarity and selection collaborate in shaping the genetic architecture of the CB breeds population. We showed that the overlaps between these two events were breed-specific, suggesting that differences in breeding objectives and selection intensities exist between CB breeds. Investigating the CB breeds' genomic architecture allows the estimation of genome-wide indicine and taurine genome proportions and demonstrates the locations within the genome where alleles with either taurine or indicine origin provide a selective advantage. Such findings provide the opportunity to control the breeding programs more efficiently.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://widde.toulouse.inra.fr/widde/widde/main.do?module=cattle.

AUTHOR CONTRIBUTIONS

SV conceived the study. SV carried out the analyses and drafted the manuscript. SS helped to carry out the analyses and write the manuscript. LB, SG, and MB critically edited and reviewed the manuscript. LB, SV, SM, and KP contributed to the final approval of the version to be published. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Guilhem Sempéré, Laurence Flori, and their colleagues for developing the freely accessible WIDDE database. A part of the statistical analyses was conducted in Canada using Compute Canada resources (www.computecanada.ca). The access to Compute Canada resources was provided by a sponsor at Dalhousie University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.844653/full#supplementary-material

REFERENCES

Ahmad, M., Srinivasula, S. M., Wang, L., Talanian, R. V., Litwack, G., Fernandes-Alnemri, T., et al. (1997). CRADD, a Novel Human Apoptotic Adaptor Molecule for Caspase-2, and FasL/tumor Necrosis Factor Receptor-Interacting Protein RIP. *Cancer Res.* 57 (4), 615–619.

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

Anand, L., and Lopez, C. M. R. (2022). ChromoMap: an R Package for Interactive Visualization of Multi-omics Data and Annotation of Chromosomes. *BMC Bioinform.* 23 (1), 1–9. doi:10.1186/s12859-021-04556-z

Aoto, J., Martinelli, D. C., Malenka, R. C., Tabuchi, K., and Südhof, T. C. (2013). Presynaptic Neurexin-3 Alternative Splicing Trans-synaptically Controls Postsynaptic AMPA Receptor Trafficking. *Cell* 154 (1), 75–88. doi:10.1016/j.cell.2013.05.060

Archer, J. A., Arthur, P. F., Herd, R. M., Parnell, P. F., and Pitchford, W. S. (1997). Optimum Postweaning Test for Measurement of Growth Rate, Feed Intake, and Feed Efficiency in British Breed Cattle. *J. Anim. Sci.* 75 (8), 2024–2032. doi:10.2527/1997.7582024x

Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J., and Strasser, A. (2018). How Does P53 Induce Apoptosis and How Does This Relate to P53-Mediated Tumour Suppression? *Cell Death Differ* 25 (1), 104–113. doi:10.1038/cdd.2017.169

Bahbahani, H., Afana, A., and Wragg, D. (2018). Genomic Signatures of Adaptive Introgression and Environmental Adaptation in the Sheko Cattle of Southwest Ethiopia. *PLoS One* 13 (8), e0202479. doi:10.1371/journal.pone.0202479

Bar-Sagi, D., and Hall, A. (2000). Ras and Rho GTPases. *Cell* 103 (2), 227–238. doi:10.1016/s0092-8674(00)00115-x

Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., et al. (2020). Adaptive Introgression from Indicine Cattle into white Cattle Breeds from Central Italy. *Sci. Rep.* 10 (1), 1–11. doi:10.1038/s41598-020-57880-4

Bassères, D. S., Tizzei, E. V., Duarte, A. A., Costa, F. F., and Saad, S. T. (2002). ARHGAP10, a Novel Human Gene Coding for a Potentially Cytoskeletal Rho-GTPase Activating Protein. *Biochem. Biophys. Res. Commun.* 294 (3), 579–585. doi:10.1016/S0006-291X(02)00514-4

Bélanger, J., and Pilling, D. (2019). *The State of the World's Biodiversity for Food and Agriculture*. (Rome: Food and Agriculture Organization of the United Nations FAO).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bickhart, D. M., Xu, L., Hutchison, J. L., Cole, J. B., Null, D. J., Schroeder, S. G., et al. (2016). Diversity and Population-Genetic Properties of Copy Number Variations and Multicopy Genes in Cattle. *DNA Res.* 23 (3), 253–262. doi:10.1093/dnares/dsw013

Blandin, G., Beroud, C., Labelle, V., Nguyen, K., Wein, N., Hamroun, D., et al. (2012). UMD-DYSF, a Novel Locus Specific Database for the Compilation and Interactive Analysis of Mutations in the Dysferlin Gene. *Hum. Mutat.* 33 (3), E2317–E2331. doi:10.1002/humu.22015

Bovine HapMap Consortium (2009). Genome-wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324 (5926), 528–532. doi:10.1126/science.1167936

Buzanskas, M. E., Grossi, D. A., Ventura, R. V., Schenkel, F. S., Sargolzaei, M., Meirelles, S. L., et al. (2014). Genome-wide Association for Growth Traits in Canchim Beef Cattle. *PLoS One* 9 (4), e94802. doi:10.1371/journal.pone.0094802

Buzanskas, M. E., Ventura, R. V., Seleguim Chud, T. C., Bernardes, P. A., Santos, D. J. de. A., Regitano, L. C. de. A., et al. (2017). Study on the Introgression of Beef Breeds in Canchim Cattle Using Single Nucleotide Polymorphism Markers. *PLoS One* 12 (2), e0171660. doi:10.1371/journal.pone.0171660

Cammack, K., Thomas, M., and Enns, R. (2009). Reproductive Traits and Their Heritabilities in Beef Cattle. *The Prof. Anim. Scientist* 25 (5), 517–528. doi:10.15232/s1080-7446(15)30753-1

Cartwright, T. (1970). Selection Criteria for Beef Cattle for the Future. *J. Anim. Sci.* 30 (5), 706–711. doi:10.2527/jas1970.305706x

Carvalho, M., Baldi, F., Alexandre, P., Santana, M., Ventura, R., Bueno, R., et al. (2019). Genomic Regions and Genes Associated with Carcass Quality in Nelore Cattle. *Genet. Mol. Res.* 18 (1), 1–15. doi:10.4238/gmr18226

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome Resequencing Reveals World-wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nat. Commun.* 9 (1), 1–13. doi:10.1038/s41467-018-04737-0

Cook, D. E., and Andersen, E. C. (2017). VCF-kit: Assorted Utilities for the Variant Call Format. *Bioinformatics* 33 (10), 1581–1582. doi:10.1093/bioinformatics/btx011

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and Runs of Homozygosity: A Possible Solution to an Old Problem. *Livestock Sci.* 166, 26–34. doi:10.1016/j.livsci.2014.05.034

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

Daugaard, M., Rohde, M., and Jäättelä, M. (2007). The Heat Shock Protein 70 Family: Highly Homologous Proteins with Overlapping and Distinct Functions. *FEBS Lett.* 581 (19), 3702–3710. doi:10.1016/j.febslet.2007.05.039

De Oliveira, P. S., Cesar, A. S., do Nascimento, M. L., Chaves, A. S., Tizioto, P. C., Tullio, R. R., et al. (2014). Identification of Genomic Regions Associated with Feed Efficiency in Nelore Cattle. *BMC Genet.* 15 (1), 1–10. doi:10.1186/s12863-014-0100-0

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genet.* 10 (3), e1004254. doi:10.1371/journal.pgen.1004254

Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2018). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* 18 (2), 623–632. doi:10.1021/acs.jproteome.8b00702

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Evans, R. D., Walsh, S. W., et al. (2020). Genomic Regions Associated with Muscularity in Beef Cattle Differ in Five Contrasting Cattle Breeds. *Genet. Selection Evol.* 52 (1), 1–18. doi:10.1186/s12711-020-0523-1

Dupin, E., Creuzet, S., and Le Douarin, N. M. (2006). The Contribution of the Neural Crest to the Vertebrate Body. *Neural Crest Induction Differ.* 589, 96–119. doi:10.1007/978-0-387-46954-6_6

Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Mol. Biol. Evol.* 28 (8), 2239–2252. doi:10.1093/molbev/msr048

Elzo, M., Johnson, D., Wasdin, J., and Driver, J. (2012). Carcass and Meat Palatability Breed Differences and Heterosis Effects in an Angus–Brahman Multibreed Population. *Meat Sci.* 90 (1), 87–92. doi:10.1016/j.meatsci.2011.06.010

Falconer, D., and Mackay, T. (1996). *Introduction to Quantitative Genetics*. Harlow, UK): Addison Wesley Longman.

Feder, M. E., and Hofmann, G. E. (1999). Heat-shock Proteins, Molecular Chaperones, and the Stress Response: Evolutionary and Ecological Physiology. *Annu. Rev. Physiol.* 61 (1), 243–282. doi:10.1146/annurev.physiol.61.1.243

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Mol. Biol. Evol.* 31 (5), 1275–1291. doi:10.1093/molbev/msu077

Fonseca, P. A., Suárez-Vega, A., and Cánovas, A. (2020). Weighted Gene Correlation Network Meta-Analysis Reveals Functional Candidate Genes Associated with High-And Sub-fertile Reproductive Performance in Beef Cattle. *Genes* 11 (5), 543. doi:10.3390/genes11050543

Freitas, P. H., Wang, Y., Yan, P., Oliveira, H. R., Schenkel, F. S., Zhang, Y., et al. (2021). Genetic Diversity and Signatures of Selection for thermal Stress in Cattle and Other Two Bos Species Adapted to Divergent Climatic Conditions. *Front. Genet.* 12, 102. doi:10.3389/fgene.2021.604823

Garner, J., Chamberlain, A., Vander Jagt, C., Nguyen, T., Mason, B., Marett, L., et al. (2020). Gene Expression of the Heat Stress Response in Bovine Peripheral white Blood Cells and Milk Somatic Cells *In Vivo*. *Scientific Rep.* 10 (1), 1–12. doi:10.1038/s41598-020-75438-2

Ghoreishifar, S. M., Eriksson, S., Johansson, A. M., Khansefid, M., Moghaddaszadeh-Ahrabi, S., Parna, N., et al. (2020). Signatures of Selection Reveal Candidate Genes Involved in Economic Traits and Cold Acclimation in Five Swedish Cattle Breeds. *Genet. Selection Evol.* 52 (1), 1–15. doi:10.1186/s12711-020-00571-5

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A Draft Sequence of the Neandertal Genome. *Science* 328 (5979), 710. doi:10.1126/science.1188021

Gregory, K., and Cundiff, L. (1980). Crossbreeding in Beef Cattle: Evaluation of Systems. *J. Anim. Sci.* 51 (5), 1224–1242. doi:10.2527/jas1980.5151224x

Gregory, K., Cundiff, L., Koch, R., Dikeman, M., and Koohmaraie, M. (1994). Breed Effects, Retained Heterosis, and Estimates of Genetic and Phenotypic Parameters for Carcass and Meat Traits of Beef Cattle. *J. Anim. Sci.* 72 (5), 1174–1183. doi:10.2527/1994.7251174x

Grigoletto, L., Ferraz, J., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. *Front. Genet.* 11, 123. doi:10.3389/fgene.2020.00123

Gui, L., Hu, Y., Hou, Z., and Wang, P. (2016). Expression and Clinical Significance of the Obesity-Related Gene TNFAIP9 in Obese Children. *Genet. Mol. Res.* 15 (3). doi:10.4238/gmr.15037995

Han, M., Xu, R., Wang, S., Yang, N., Ni, S., Zhang, Q., et al. (2018). Six-transmembrane Epithelial Antigen of Prostate 3 Predicts Poor Prognosis and Promotes Glioblastoma Growth and Invasion. *Neoplasia* 20 (6), 543–554. doi:10.1016/j.neo.2018.04.002

Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., et al. (2002). Selection and Use of SNP Markers for Animal Identification and Paternity Analysis in US Beef Cattle. *Mamm. Genome* 13 (5), 272–281. doi:10.1007/s00335-001-2146-3

Hedrick, P. W. (2013). Adaptive Introgression in Animals: Examples and Comparison to New Mutation and Standing Variation as Sources of Adaptive Variation. *Mol. Ecol.* 22 (18), 4606–4618. doi:10.1111/mec.12415

Hiendleder, S., Lewalski, H., and Janke, A. (2008). Complete Mitochondrial Genomes of *Bos taurus* and *Bos indicus* Provide New Insights into Intra-species Variation, Taxonomy and Domestication. *Cytogenet. Genome Res.* 120 (1–2), 150–156. doi:10.1159/000118756

Highfill, C., Esquivel-Font, O., Dikeman, M., and Kropf, D. (2012). Tenderness Profiles of Ten Muscles from F1 *Bos indicus* X *Bos taurus* and *Bos taurus* Cattle Cooked as Steaks and Roasts. *Meat Sci.* 90 (4), 881–886. doi:10.1016/j.meatsci.2011.11.022

Hoppe, S., Brandt, H. R., König, S., Erhardt, G., and Gauly, M. (2010). Temperament Traits of Beef Calves Measured under Field Conditions and Their Relationships to Performance. *J. Anim. Sci.* 88 (6), 1982–1989. doi:10.2527/jas.2008-1557

Imai, K., Kurita-Ochiai, T., and Ochiai, K. (2003). Mycobacterium Bovis bacillus Calmette-Gueérin Infection Promotes SOCS Induction and Inhibits IFN-γ-Stimulated JAK/STAT Signaling in J774 Macrophages. *FEMS Immunol. Med. Microbiol.* 39 (2), 173–180. doi:10.1016/s0928-8244(03)00231-1

Junjvlieke, Z., Khan, R., Mei, C., Cheng, G., Wang, S., Raza, S. H. A., et al. (2020). Effect of ELOVL6 on the Lipid Metabolism of Bovine Adipocytes. *Genomics* 112 (3), 2282–2290. doi:10.1016/j.ygeno.2019.12.024

Kester, H. A., Blanchetot, C., den Hertog, J., van der Saag, P. T., and van der Burg, B. (1999). Transforming Growth Factor-β-Stimulated Clone-22 Is a Member of a Family of Leucine Zipper Proteins that Can Homo-And Heterodimerize and Has Transcriptional Repressor Activity. *J. Biol. Chem.* 274 (39), 27439–27447. doi:10.1074/jbc.274.39.27439

Kuhn, M., Mavangira, V., and Sordillo, L. (2020). Invited Review: Cytochrome P450 Enzyme Involvement in Health and Inflammatory-Based Diseases of Dairy Cattle. *J. Dairy Sci.* 104 (2), 1276–1290. doi:10.3168/jds.2020-18997

Li, N., Yu, Q.-L., Yan, X.-M., Li, H.-B., and Zhang, Y. (2020). Sequencing and Characterization of miRNAs and mRNAs from the Longissimus Dorsi of Xinjiang Brown Cattle and Kazakh Cattle. *Gene* 741, 144537. doi:10.1016/j.gene.2020.144537

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., and Verity, B. (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

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

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., et al. (2015). Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater lake. *Science* 350 (6267), 1493. doi:10.1126/science.aac9927

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite-Fast D-statistics and Related Admixture Evidence from VCF Files. *Mol. Ecol. Resour.* 21 (2), 584–595. doi:10.1111/1755-0998.13265

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., et al. (2018). Whole-genome Sequences of Malawi Cichlids Reveal Multiple Radiations Interconnected by Gene Flow. *Nat. Ecol. Evol.* 2 (12), 1940–1955. doi:10.1038/s41559-018-0717-x

McTavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F., and Hillis, D. M. (2013). New World Cattle Show Ancestry from Multiple Independent Domestication Events. *Proc. Natl. Acad. Sci.* 110 (15), E1398–E1406. doi:10.1073/pnas.1303367110

Mohammadi, A., Alijani, S., Rafat, S. A., and Abdollahi-Arpanahi, R. (2020). Genome-wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle. *Ann. Anim. Sci.* 20 (3), 825–851. doi:10.2478/aoas-2020-0031

Morales-Cruz, A., Aguirre-Liguori, J. A., Zhou, Y., Minio, A., Riaz, S., Walker, A. M., et al. (2021). Introgression Among North American Wild Grapes (Vitis) Fuels Biotic and Abiotic Adaptation. *Genome Biol.* 22 (1), 1–27. doi:10.1186/s13059-021-02467-z

Morrison, D., Humes, P., and Koonce, K. (1989). Comparisons of Brahman and Continental European Crossbred Cows for Calving Ease in a Subtropical Environment. *J. Anim. Sci.* 67 (7), 1722–1731. doi:10.2527/jas1989.6771722x

Moskovitz, J., Bar-Noy, S., Williams, W. M., Requena, J., Berlett, B. S., and Stadtman, E. R. (2001). Methionine Sulfoxide Reductase (MsrA) Is a Regulator of Antioxidant Defense and Lifespan in Mammals. *Proc. Natl. Acad. Sci.* 98 (23), 12920–12925. doi:10.1073/pnas.231472998

Moubarak, A., and Rosenkrans, C., Jr (2000). Hepatic Metabolism of Ergot Alkaloids in Beef Cattle by Cytochrome P450. *Biochem. Biophysical Res. Commun.* 274 (3), 746–749. doi:10.1006/bbrc.2000.3210

Mucignat-Caretta, C. (2014). *Neurobiology of Chemical Communication*.

Neupane, M., Geary, T. W., Kiser, J. N., Burns, G. W., Hansen, P. J., Spencer, T. E., et al. (2017). Loci and Pathways Associated with Uterine Capacity for Pregnancy and Fertility in Beef Cattle. *PLoS One* 12 (12), e0188997. doi:10.1371/journal.pone.0188997

Omura, T. (1999). Forty Years of Cytochrome P450. *Biochem. Biophysical Res. Commun.* 266 (3), 690–698. doi:10.1006/bbrc.1999.1887

Paim, T. do. P., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., Paiva, S. R., et al. (2020). Genomic Breed Composition of Selection Signatures in Brangus Beef Cattle. *Front. Genet.* 11, 710. doi:10.3389/fgene.2020.00710

Paredes-Sánchez, F. A., Sifuentes-Rincón, A. M., Casas, E., Arellano-Vera, W., Parra-Bracamonte, G. M., Riley, D. G., et al. (2020). Novel Genes Involved in the Genetic Architecture of Temperament in Brahman Cattle. *PLoS One* 15 (8), e0237825.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient Admixture in Human History. *Genetics* 192 (3), 1065–1093. doi:10.1534/genetics.112.145037

Pease, J. B., and Hahn, M. W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Syst. Biol.* 64 (4), 651–662. doi:10.1093/sysbio/syv023

Peletto, S., Strillacci, M., Capucchio, M., Biasibetti, E., Modesto, P., Acutis, P., et al. (2017). Genetic Basis of Lipomatous Myopathy in Piedmontese Beef Cattle. *Livestock Sci.* 206, 9–16. doi:10.1016/j.livsci.2017.09.027

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., Evans, R. D., and Berry, D. P. (2020). Breed-and Trait-specific Associations Define the Genetic Architecture of Calving Performance Traits in Cattle. *J. Anim. Sci.* 98 (5), skaa151. doi:10.1093/jas/skaa151

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.* 10 (2), e1004148. doi:10.1371/journal.pgen.1004148

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). G: Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). *Nucleic Acids Res.* 47 (W1), W191–W198. doi:10.1093/nar/gkz369

Rehman, K. S., Yin, S., Mayhew, B. A., Word, R. A., and Rainey, W. E. (2003). Human Myometrial Adaptation to Pregnancy: CDNA Microarray Gene Expression Profiling of Myometrium from Non-pregnant and Pregnant

Women. *MHR: Basic Sci. Reprod. Med.* 9 (11), 681–700. doi:10.1093/molehr/gag078

Rezende, F., Dietsch, G., and Peñagaricano, F. (2018). Genetic Dissection of Bull Fertility in US Jersey Dairy Cattle. *Anim. Genet.* 49 (5), 393–402. doi:10.1111/age.12710

Rhoad, A. O. (1949). The Santa Gertrudis Breed: The Genesis and the Genetics of a New Breed of Beef Cattle. *J. Hered.* 40 (5), 115–126. doi:10.1093/oxfordjournals.jhered.a106008

Richardson, C., and Hatfield, E. (1978). The Limiting Amino Acids in Growing Cattle. *J. Anim. Sci.* 46 (3), 740–745. doi:10.2527/jas1978.463740x

Ritchie, H. (2009). *Breeds of Beef and Multi-Purpose Cattle: Acknowledgements*.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., 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:10.1038/nature01140

Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE: a Web-Interfaced Next Generation Database for Genetic Diversity Exploration, with a First Application in Cattle. *BMC Genomics* 16 (1), 1–8.

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

Singh, A., Mehrotra, A., Gondro, C., da Silva Romero, A. R., Pandey, A. K., Karthikeyan, A., et al. (2020). Signatures of Selection in Composite Vrindavani Cattle of India. *Front. Genet.* 11. doi:10.3389/fgene.2020.589496

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31 (10), 2824–2827. doi:10.1093/molbev/msu211

Taniguchi, M., Basarab, J. A., Dodson, M. V., and Moore, S. S. (2008). Comparative Analysis on Gene Expression Profiles in Cattle Subcutaneous Fat Tissues. *Comp. Biochem. Physiol. D: Genomics Proteomics* 3 (4), 251–256. doi:10.1016/j.cbd.2008.06.002

Thundathil, J. C., Dance, A. L., and Kastelic, J. P. (2016). Fertility Management of Bulls to Improve Beef Cattle Productivity. *Theriogenology* 86 (1), 397–405. doi:10.1016/j.theriogenology.2016.04.054

Torres, R., Szpiech, Z. A., and Hernandez, R. D. (2018). Human Demographic History Has Amplified the Effects of Background Selection across the Genome. *PLoS Genet.* 14 (6), e1007387. doi:10.1371/journal.pgen.1007387

Ueda, S., Hosoda, M., Yoshino, K., Yamanoue, M., and Shirai, Y. (2021). Gene Expression Analysis Provides New Insights into the Mechanism of Intramuscular Fat Formation in Japanese Black Cattle. *Genes* 12 (8), 1107. doi:10.3390/genes12081107

Venables, W. N., and Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media.

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

Vilchez, J. J., Gallano, P., Gallardo, E., Lasa, A., Rojas-García, R., Freixas, A., et al. (2005). Identification of a Novel Founder Mutation in the DYSF Gene Causing Clinical Variability in the Spanish Population. *Arch. Neurol.* 62 (8), 1256–1259. doi:10.1001/archneur.62.8.1256

Villanueva, R. A. M., and Chen, Z. J. (2019). *ggplot2: Elegant Graphics for Data Analysis*.

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 (3), e72. doi:10.1371/journal.pbio.0040072

Vojtek, A. B., and Der, C. J. (1998). Increasing Complexity of the Ras Signaling Pathway. *J. Biol. Chem.* 273 (32), 19925–19928. doi:10.1074/jbc.273.32.19925

Wang, J., Liu, C., Chen, J., Bai, Y., Wang, K., Wang, Y., et al. (2020). Genome-Wide Analysis Reveals Human-Mediated Introgression from Western Pigs to Indigenous Chinese Breeds. *Genes* 11 (3), 275. doi:10.3390/genes11030275

Warwick, E. J. (1958). Fifty Years of Progress in Breeding Beef Cattle. *J. Anim. Sci.* 17 (4), 922–943. doi:10.2527/jas1958.174922x

Wheeler, T., Cundiff, L., Shackelford, S., and Koohmaraie, M. (2010). Characterization of Biological Types of Cattle (Cycle VIII): Carcass, Yield, and Longissimus Palatability Traits. *J. Anim. Sci.* 88 (9), 3070–3083. doi:10.2527/jas.2009-2497

Widmer, S., Seefried, F. R., von Rohr, P., Häfliger, I. M., Spengeler, M., and Drögemüller, C. (2021). A Major QTL at the LHCGR/FSHR Locus for Multiple Birth in Holstein Cattle. *Genet. Selection Evol.* 53 (1), 1–15. doi:10.1186/s12711-021-00650-1

Wilkins, A. S., Wrangham, R. W., and Fitch, W. T. (2014). The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics* 197 (3), 795–808. doi:10.1534/genetics.114.165423

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P., Brockmann, G. A., König, S., et al. (2021). Genome-Wide Association Study Using Whole-Genome Sequence Data for Fertility, Health Indicator, and Endoparasite Infection Traits in German Black Pied Cattle. *Genes* 12 (8), 1163. doi:10.3390/genes12081163

Wright, D., Henriksen, R., and Johnsson, M. (2020). *Trends in Ecology & Evolution*, S0169–S5347. doi:10.1016/j.tree.2020.08.009Defining the domestication syndrome: Comment on Lord et al. 2020

Xi, Y., Yang, Z., Wu, F., Han, Z., and Wang, G. (2015). Gene Expression Profiling of Hormonal Regulation Related to the Residual Feed Intake of Holstein Cattle. *Biochem. Biophysical Res. Commun.* 465 (1), 19–25. doi:10.1016/j.bbrc.2015.07.092

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). Rmvp: A Memory-Efficient, Visualization-Enhanced, and Parallel-Accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics & Bioinformatics* S1672-0229 (21), 00050–00054. doi:10.1016/j.gpb.2020.10.007

Yurchenko, A. A., Daetwyler, H. D., Yudin, N., Schnabel, R. D., Vander Jagt, C. J., Soloshenko, V., et al. (2018). Scans for Signatures of Selection in Russian Cattle Breed Genomes Reveal New Candidate Genes for Environmental Adaptation and Acclimation. *Scientific Rep.* 8 (1), 1–16. doi:10.1038/s41598-018-31304-w

Zhan, S., Zhao, W., Song, T., Dong, Y., Guo, J., Cao, J., et al. (2018). Dynamic Transcriptomic Analysis in Hircine Longissimus Dorsi Muscle from Fetal to Neonatal Development Stages. *Funct. Integr. Genomics* 18 (1), 43–54. doi:10.1007/s10142-017-0573-9

Zhang, Y., Hu, Y., Wang, X., Jiang, Q., Zhao, H., Wang, J., et al. (2020). Population Structure, and Selection Signatures Underlying High-Altitude Adaptation Inferred from Genome-wide Copy Number Variations in Chinese Indigenous Cattle. *Front. Genet.* 10, 1404. doi:10.3389/fgene.2019.01404

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), 1–10. doi:10.1186/gb-2009-10-4-r42

Zwartkruis, F. J., and Bos, J. L. (1999). Ras and Rap1: Two Highly Related Small GTPases with Distinct Function. *Exp. Cel Res.* 253 (1), 157–165. doi:10.1006/excr.1999.4695

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 July 2022
doi: 10.3389/fgene.2022.848951

Genomic Diversity and Selection Signatures for Weining Cattle on the Border of Yunnan-Guizhou

Yangkai Liu[1†], Haijian Cheng[1†], Shikang Wang[1], Xiaoyv Luo[1], Xiaohui Ma[1], Luyang Sun[1], Ningbo Chen[1], Jicai Zhang[2], Kaixing Qu[3], Mingjin Wang[4], Jianyong Liu[2], Bizhi Huang[2] and Chuzhao Lei[1]**

[1]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, [2]Yunnan Academy of Grassland and Animal Science, Kunming, China, [3]Academy of Science and Technology, Chuxiong Normal University, Chuxiong, China, [4]Bijie Animal Husbandry and Veterinary Science Institute, Bijie, China

Weining cattle is a Chinese indigenous breed influenced by complex breeding and geographical background. The multi-ethnic breeding culture makes Weining cattle require more attention as livestock resources for its genetic diversity. Here, we used 10 Weining cattle (five newly sequenced and five downloaded) and downloaded another 48 genome data to understand the aspects of Weining cattle: genetic diversity, population structure, and cold-adapted performance. In the current study, a high level of genetic diversity was found in Weining cattle, and its breed comprised two potential ancestries, which were *Bos taurus* and *Bos indicus*. The positive selective sweep analysis in Weining cattle was analyzed using composite likelihood ratio (CLR) and nucleotide diversity ($\theta\pi$), resulting in 203 overlapped genes. In addition, we studied the cold adaptation of Weining cattle by comparing with other Chinese cattle (Wannan and Wenshan cattle) by three methods (F_{ST}, $\theta\pi$-ratio, and XP-EHH). Of the top 1% gene list, *UBE3D* and *ZNF668* were analyzed, and these genes may be associated with fat metabolism and blood pressure regulation in cold adaptation. Our findings have provided invaluable information for the development and conservation of cattle genetic resources, especially in southwest China.

Keywords: Weining cattle, hybrid, genetic diversity, selection signatures, UBE3D

OPEN ACCESS

Edited by:
Tiago do Prado Paim, Goiano Federal Institute (IFGOIANO), Brazil

Reviewed by:
Qianjun Zhao, Institute of Animal Sciences (CAAS), China
Lingyang Xu, Institute of Animal Sciences (CAAS), China
Oscar Cortes Gardyn, Complutense University of Madrid, Spain
Yang Zhou, Huazhong Agricultural University, China

****Correspondence:***
Bizhi Huang
hbz@ynbp.cn
Chuzhao Lei
leichuzhao1118@126.com

†These authors have contributed equally to this work and share first authorship

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

Received: *05 January 2022*
Accepted: *30 May 2022*
Published: *07 July 2022*

Citation:
Liu Y, Cheng H, Wang S, Luo X, Ma X, Sun L, Chen N, Zhang J, Qu K, Wang M, Liu J, Huang B and Lei C (2022) Genomic Diversity and Selection Signatures for Weining Cattle on the Border of Yunnan-Guizhou. Front. Genet. 13:848951. doi: 10.3389/fgene.2022.848951

1 INTRODUCTION

Humpless cattle (*Bos taurus*) and humped cattle (*Bos indicus*) are two main sub-species of cattle (Decker et al., 2014), and they have been directed by multiple domestication events from the early requirement of labor to the current need for beef and milk. Today, a total of 55 indigenous cattle breeds are officially identified in China. Though the hybrids of indicine × taurine breed in Chinese cattle breeds are massive in number, the composition of the ancestors remains unclear for native breeds, particularly in southwest China. In recent studies, maternal and paternal genetic markers have shown that southwest Chinese cattle were complex but interesting, which consisted of an important node for inspiring the domestication history of Chinese cattle (Lei et al., 2006; Chen et al., 2009; Xia et al., 2019). From a whole-genome aspect, cattle identified as the most purebred could be used as a nucleus for recovering the native genetic background in the current admixed population. One finding supported that domestic cattle consist of five core groups, which were European taurine, Eurasian taurine, East Asian taurine, Chinese indicine, and Indian indicine (Chen et al., 2018). Based

on that information, other researchers have focused on native breeds by the whole-genome selection (WGS) studied from economic characters to adaptable ones (Kawahara-Miki et al., 2011; Kim et al., 2017; Shen et al., 2020).

To better understand the genetic basis of adapted traits in cattle, many studies have focused on kinds of breeds adapted to various environments, including Tibetan cattle at high elevations (Xin et al., 2019), Iraqi cattle in dry and hot environment (Alshawi et al., 2019), and cold acclimation to Swedish cattle breeds (Ghoreishifar et al., 2020). Herein, low-temperature stimulation can induce animal hormones and other environmental adaption, which has a direct impact on the reproduction efficiency and production level. To date, several cold environment types have been studied, and a number of candidate genes have been reported with major effects on cold adaptation in cattle. For example, *RETREG1* and *RPL7* were under strong selection in Yakut cattle, which originate from Eastern Siberia (Yurchenko et al., 2018). *FGF5*, a hair growth factor, was selected as a distinctive feature (long and dense hairs) of Yanbian cattle (Shen et al., 2020). Here, a unique alpine mountainous area with lower temperature and high humidity on the border of Yunnan–Guizhou is formed due to the uplift of the Qinghai–Tibet plateau. Few studies have been reported on the cattle adapted to cold and humid mountains in the Yunnan–Guizhou region.

Weining cattle were shaped both from multicultural zone and complex natural ecological environments. It has characteristics of rough feeding resistance, cold resistance, good climbing, and easy fattening (China National Commission of Animal Genetic Resources, 2011). Cattle in the border of Yunnan–Guizhou are a typical hybrid of *Bos taurus* × *Bos indicus*, and Weining cattle is one of them (Nie et al., 1999). Historically, multi-ethnic livestock breeding backgrounds from Yi, Miao, and Hui national minorities bred diverse native cattle breeds, which potentially contained complex genetic backgrounds in southwest China. For the long-term national autonomy management and advocating natural national culture, it makes the genetic improvement process tardy. Geographically, the uplift of altitude (2,800 m) in the Yunnan–Guizhou area caused Weining cattle to adapt to the cold (annual average temperature 10 °C) and humid (annual mean humidity 75%–80%) environment. In terms of physical characteristics, the sagging skin of the neck enhances the heat dissipation capacity of indicine cattle, but the sagging skin of Weining cattle is not developed in humid and cold environment. In the present scenario, the low socioeconomic benefits still push this cattle breed to gradually decrease (Li R. et al., 2019). Thereby, it is necessary to study the genetic diversity and adaptability of Weining cattle.

A scarce number of studies were carried out to explore the knowledge of genomic variation in Weining cattle at the genome level. We analyzed 58 whole-genome data of individuals (including five newly sequenced Weining cattle data) and identified single-nucleotide polymorphisms (SNPs) compared with those of commercial and native populations around the world based on the *Bos taurus* reference genome assembly (ARS-UCD1.2). This study may potentially reveal the ancestral components, population structure, and genetic diversity of Weining cattle.

2 MATERIAL AND METHODS

2.1 Ethics Statement

This study was approved by the Institutional Animal Care and Use Committee of Northwest A&F University following the recommendation of the Regulations for the Administration of Affairs Concerning Experimental Animals of China (Permit number: NWAFAC1019).

2.2 Sample Collection and Genome Re-sequencing

We sampled five Weining cattle from Bi'jie, Guizhou, China. These samples were collected from villages, and the farmers were interviewed in detail to ensure unrelatedness among the sampled individuals. Genomic DNA was extracted from ear tissues using the standard phenol–chloroform method (Reid, 1991). Paired-end libraries with an insert size of 500 bp were constructed for each individual, and whole-genome sequencing was performed using Illumina NovaSeq instruments at Novogene Bioinformatics Institute, Beijing, China.

In addition, another five published data of Weining cattle were downloaded, and we also downloaded genome-wide data of 48 cattle for comparison including Wenshan cattle (n = 5), Wannan cattle (n = 5), Guangfeng cattle (n = 4), Hanwoo cattle (n = 10), Brahman (n = 4), Gir (n = 2), Nelore (n = 1), Angus (n = 9), and Simmental (n = 8) (**Supplementary Table S1**). In total, 58 individuals were used from ten breeds in our analysis. More detailed information about all samples analyzed in this study is provided in the additional file: **Supplementary Table S1**. Raw FASTQ sequences have been deposited to the NCBI under the BioProject accession number PRJNA379859.

2.3 Read Mapping and SNP Calling

After obtaining the WGS data, all clean reads were aligned to the latest *Bos taurus* reference assembly ARS-UCD1.2 using BWA-MEM (0.7.13- r1126) with default parameters. The average mapping rate of these reads sequenced in this study was 99.28%, and the sequencing coverage was approximately 12 × per individual. Then, potential duplicate reads were filtered by Picard tools (REMOVE_DUPLICATES = true) (http://broadinstitute.github.io/picard). After that, the Genome Analysis Toolkit (GATK, version 3.8) was further used for SNP calling (McKenna et al., 2010). GATK, "variant Filtration" was implemented for all SNPs as follows: "DP < 235 (1/3-fold total sequence depth for all individuals), DP > 2,115 (3-fold of total sequence depth for all individuals), QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0 and SOR >3.0". Finally, all the high-quality SNPs were annotated by SnpEff software (v4.3T) (Cingolani et al., 2012).

2.4 Population Structure and Genetic Diversity Analysis

This study used Admixture, constructed an unrooted neighbor-joining (NJ) tree, and performed PCA by using genome-wide

SNPs within autosomes to determine the population genetic structure. Principal component analysis (PCA) was performed using the smartPCA program in the EIGENSOFT v5.0 software package (Patterson et al., 2006). Population structure was carried out using ADMIXTURE v1.3 with kinship (K) set from 2 to 4 (Alexander et al., 2009). NJ trees were constructed with PLINK using a pairwise genetic distance matrix and visualized with MEGA v5.0 and iTOL (v5.1.2) (https://itol.embl.de/) (Letunic and Bork, 2021).

VCFtools were used to estimate nucleotide diversity (θπ) for each breed or population (Danecek et al., 2011). The window size and step sizes were 50K and 20K bp, respectively. Linkage disequilibrium (LD) decay was calculated using PopLDdecay with default parameters (Zhang et al., 2019). Based on the number of autosomal SNPs, runs of homozygosity (ROHs) of each individual were calculated by PLINK (-homozyg-window-snp 50). We primarily calculated the total number of ROHs (0.5–1 Mb, 1–2 Mb, 2–4 Mb, and >4 Mb) per breed/population. Using ROHs to calculate the genomic inbreeding coefficient (Forutan et al., 2018), F_{ROH} can accurately calculate the number of inbreeding lines. F_{ROH} is calculated by calculating the ratio of the total length of ROH fragments in the genome to the total length (L_{ROH}) of the genome (L_{auto}). The formula is as follows: $F_{ROH} = \sum L_{ROH}/L_{auto}$.

2.5 Genome-wide Selective Sweep Identification

Within Weining cattle for genome scans, we have used the nucleotide diversity (θπ) and the composite likelihood ratio (CLR), two statistic methods. θπ was estimated based on a sliding window method with windows of 50 kb and a step of 20 kb using VCFtools (Danecek et al., 2011). The CLR test was calculated for sites in non-overlapping 50-kb windows by using SweepFinder2 (Nielsen et al., 2005). The top 1% of the windows of each method was considered as candidate signatures of selection.

According to the National Cattle Resources of China (China National Commission of Animal Genetic Resources, 2011), Weining cattle is well-adapted to the cold and humid environment of Yunnan–Guizhou Plateau. We also chose Wannan cattle and Wenshan cattle as reference populations for their resistance to higher temperatures (Yan et al., 2022). To identify genomic regions of selective sweeps associated with cold adaptation, fixation index (F_{ST}), nucleotide diversity ratio (θπ-ratio), and cross-population extended haplotype homozygosity (XP-EHH) methods were used to select positive natural regions in the Weining cattle genome. These statistics were calculated by using a sliding window approach: a 50-kb window and a step size of 20 kb. We calculated the average F_{ST}, θπ-ratio, and XP-EHH values of each SNP window and used the outlier method to obtain the windows with the top 1% values of each method. Finally, three gene clusters of three methods in these outlier windows were carried out.

In addition, candidate genes overlapped at least in two methods were taken for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analyses by KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/) (Bu et al., 2021). Finally, the pathway terms (p-value <0.05) were taken as significant terms for statistics.

3 RESULTS

3.1 Genome Sequencing, Mapping, and SNP Identification

Individual genomes of five Weining cattle were generated to 10.43 × coverage each and were jointly genotyped with the publicly available genome (ARS-UCD1.2) (**Supplementary Table S1**). To reveal the diversity of Weining cattle, 10 Weining cattle (five new and five download) were jointly genotyped with 48 publicly available genomes from four representative groups which were European taurine (Angus and Simmental), East Asian taurine (Hanwoo), Chinese indicine (Wenshan, Wannan, and Guangfeng) and Indian indicine (Nelore, Brahman, and Gir). The average alignment rate and sequencing depth of the final set reached 99.24% and ~12 × respectively. In total, 29, 541, 306 SNPs were kept out from 10 Weining cattle genomes by snpEff. Functional annotation of the polymorphic sites revealed that the vast majority of SNPs were present in either intergenic regions (58.99%) or intronic regions (37.97%) (**Supplementary Table S2**). Exons contained 0.83% of the total SNPs with 157,065 non-synonymous SNPs and 246,751 synonymous SNPs (**Figure 1A**). Meanwhile, the analyses of specific SNPs of each breed/population have displayed that the genetic diversity of Weining cattle was second only to Chinese indicine (**Figure 1B**).

3.2 Genetic Diversity and Population Structure of Weining Cattle

The specific SNPs of each population showed a basically consistent pattern in the different populations. Specifically, Weining cattle, Chinese indicine, and Indian indicine exhibited higher specific SNP numbers, whereas the opposite genomic variations were observed in the European taurine and East Asian taurine (**Figure 1B**). For genomic characteristics, the nucleotide diversity of Weining (mean θπ = 0.00315) was lower than that of Chinese indicine (mean θπ = 0.0038) but approximately two times higher than that of European breeds (mean θπ = 0.001–0.0013) (**Figure 1C**). The genetic diversity of Weining cattle is located between indicine and taurine two clusters, but it still has higher nucleotide polymorphisms found in the indicine population (Chinese indicine > Weining cattle > Indian indicine) (**Figure 1C**). On the contrary, Weining cattle have a low level of LD, and the taurine population (Hanwoo, Angus, and Simmental) exhibited a higher level (**Figure 1D**). ROH results showed that Chinese indicine, Indian indicine, and Weining had a low self-interbreeding degree, and the main length of ROH was distributed in the interval of 0.5–1 Mb, while medium (1–2 Mb) and long (2–4 Mb) ROH fragments were found in Angus, Hanwoo, and Simmental genomes (**Figure 1E**). F_{ROH} also showed that Angus and Simmental cattle breeds had significantly higher inbreeding coefficients than other populations (**Figure 1F**).

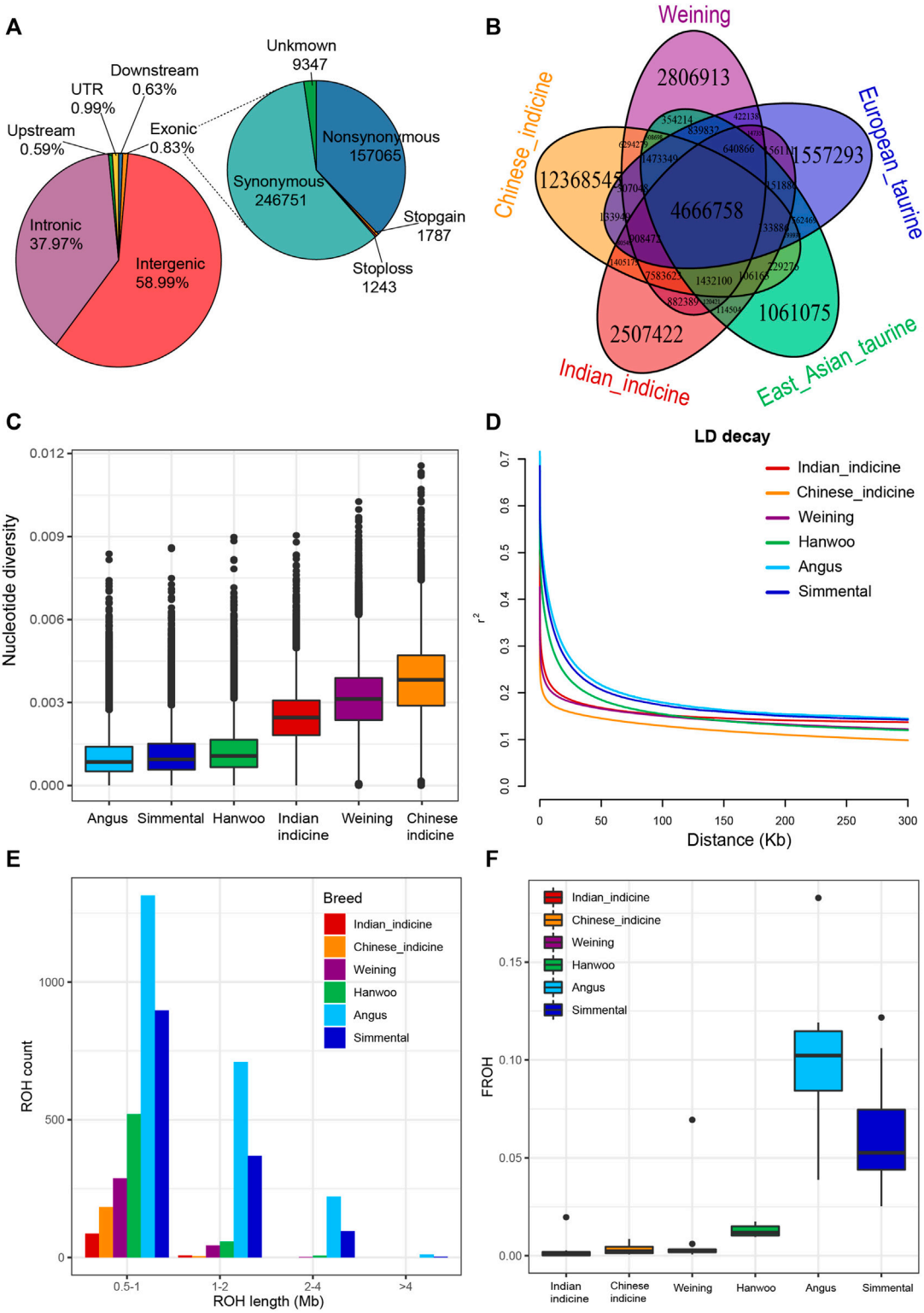

FIGURE 1 | Summary statistics for genomic variation. **(A)**. Functional classification of the detected SNPs. **(B)**. Specific and shared SNPs between Weining and other cattle groups. **(C)**. Genome-wide distribution of nucleotide diversity of each breed/population in 50-kb windows with 20 kb steps. The horizontal line inside the box indicates the median of this distribution; box limits indicate the first and the third quartiles, and points show outliers. Data points outside the whiskers can be considered outliers. **(D)**. Genome-wide average LD decay estimated from each breed/population; **(E)**. Distribution of the total number of ROH across chromosomes. **(F)**. F_{ROH} of each breed/population.

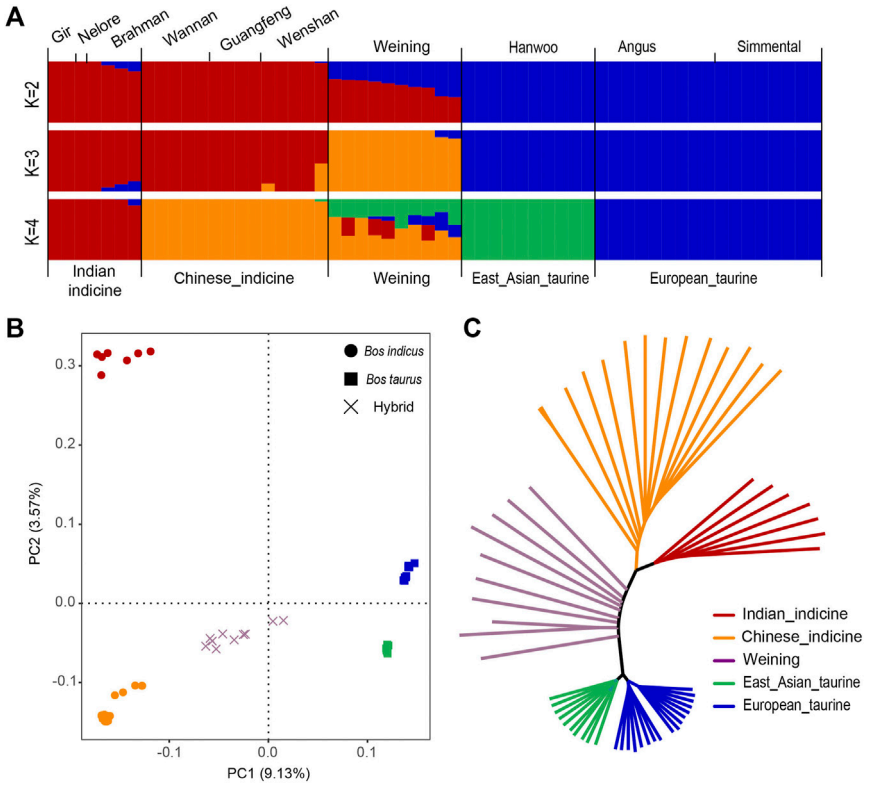

FIGURE 2 | Population structure and relationships of Weining in comparison to several possible ancestral breeds. **(A)**. Model-based clustering of cattle breeds using ADMIXTURE with K = 2 and K = 4. Breeds are colored by geographic regions and labeled with breed name. **(B)**. Principal component analysis of 10 cattle breeds. **(C)**. Neighbor-joining tree of the relationships between the ten cattle breeds (58 animals).

The admixture analysis revealed two clusters (K = 2 with the lowest cross-validation error), corresponding to taurine and indicine cattle lineages (**Figure 2A**). Similarly, the result of PCA showed that the first PC, explaining 9.13% of the total variation, was driven by the difference between indicine and taurine cattle. The second PC, explaining 3.57% of the total variation, separated South Asian indicine cattle (Nelore, Brahman, and Gir cattle) from Chinese indicine (Wannan, Wenshan, and Guangfeng cattle) and Weining cattle (**Figure 2B**). The same population classification was recovered in the NJ tree (**Figure 2C**).

3.3 Genome-wide Selective Sweep Test Within Weining Breed

To further uncover genomic region information, θπ and CLR were used to detect the genome print of Weining cattle (**Supplementary Figure S1A, B**). Totally, 1736 and 606 genes were annotated by θπ and CLR, respectively (**Supplementary Table S3, S4**). After overlapping these two gene clusters, 203 genes were analyzed (**Supplementary Figure S1C**). Of these genes, *DLG2*, *PRLR*, *MLH1*, *CFAP299*, *GOLGA4*, and *CCNH* were reported as reproductive trait–related candidate genes (Ortega et al., 2016; Li H. et al., 2019; Singh et al., 2019; Sweett et al., 2020; Shi et al., 2021; Turan et al., 2021). The age of first mating of Weining cattle is 22 months, which is 3–5 months later than that of Wannan cattle and Wenshan cattle (China National Commission of Animal Genetic Resources, 2011). *CCNH* has been reported to be associated with male fertility in humans, and it showed low nucleotide diversity at Chr7: 87037314-87067889 of the Weining cattle genome (**Supplementary Figure S1D**) (Singh et al., 2019).

We also found genes (*CYB5R4*, *UBE3D*, and *VGLL2*) related to muscle growth and fat deposition (Rovadoscki et al., 2018; Hou et al., 2021; Xia et al., 2021). Furthermore, KEGG terms for overlapped genes by KOBAS were carried out, whereas, the thyroid hormone signaling pathway, PI3K-Akt signaling pathway, hippo signaling pathway, TGF-beta signaling pathway, MAPK signaling pathway, sphingolipid signaling pathway, and FoxO signaling pathway were significantly enriched ($p < 0.05$) (**Supplementary Figure S1**, **Table S5, S6**).

3.4 Candidate Regions and Genes Under Positive Selection in Weining Cattle Related to Cold Adaptation

To detect the genome-wide selection signature related to cold climate adaptation, we compared Weining to the Chinese indicine population (Wannan and Wenshan cattle breeds), and significant signal regions (top 1%) were obtained by three

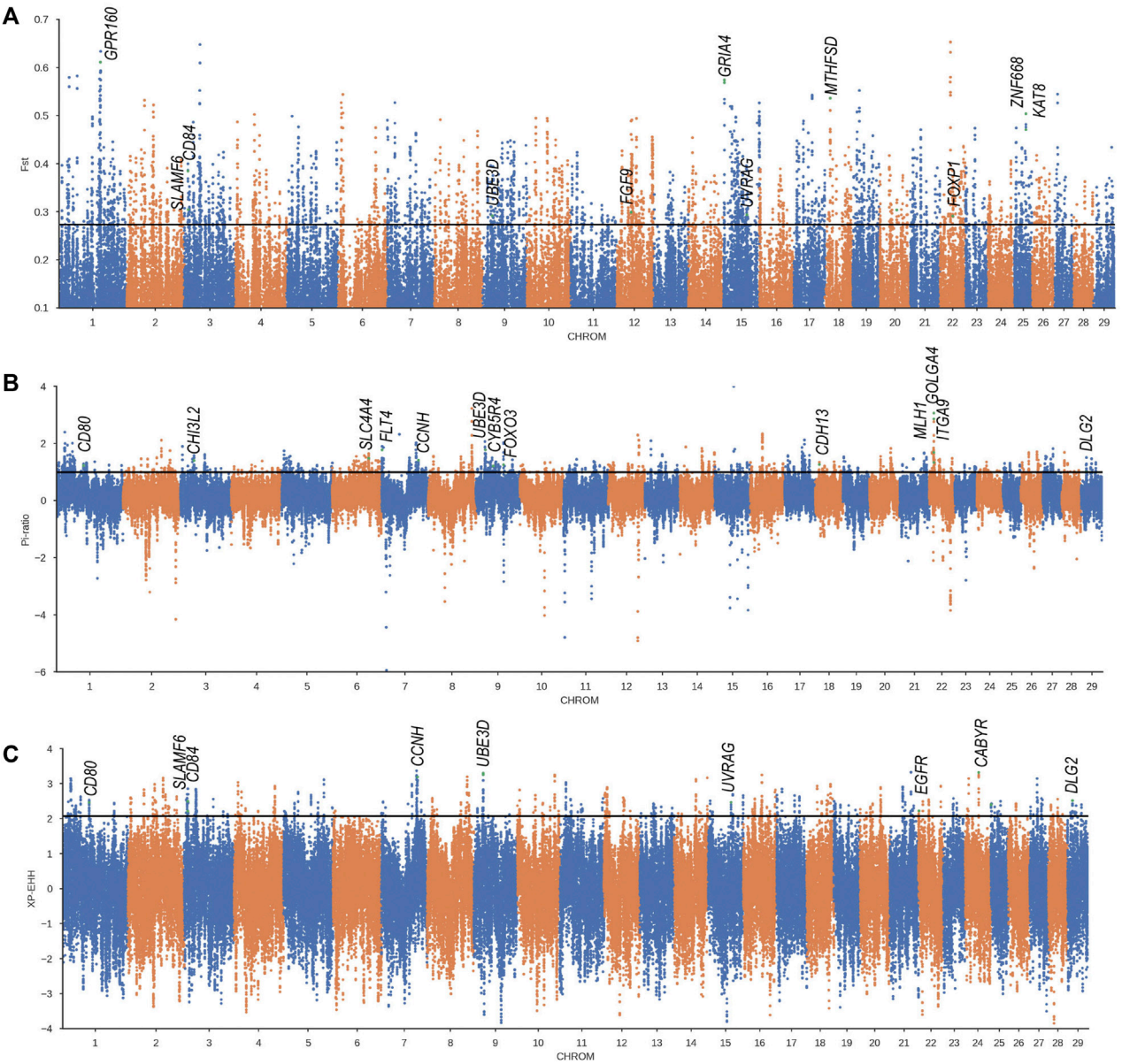

FIGURE 3 | Genome-wide selection scan in Weining cattle using sliding window analysis (50 kb window size, 20 kb step size, 99th percentile cutoff) **(A)**. Selection signatures in Weining cattle for F_{ST} (Weining-to-Wenshan & Wannan). **(B)**. Selection signatures in Weining cattle for π-ratio (Wenshan & Wannan/Weining). **(C)** Selection signatures in Weining cattle for XP-EHH (Weining-to-Wenshan & Wannan). The threshold (top 1%) of F_{ST}, π-ratio, and XP-EHH was marked with a horizontal black line.

methods (F_{ST}, XP-EHH, and θπ-ratio) (**Supplementary Table S7, S8, S9**) and annotated to 1803, 400, and 700 genes respectively; of these genes, 248 genes were obtained overlapping at least in two methods (**Figure 4A**). To determine the most likely cold-adapted ones among these overlapped genes, we performed haplotype and non-synonymous variation analyses and reviewed lots of research studies. We found a region (Chr 9: 22800001–22850000) with high F_{ST}, θπ-ratio, and XP-EHH values under strong selective scanning (**Figure 3**). This region was annotated to the *UBE3D* gene, and it also has a strong signal in CLR and θπ, which may be related to the cold adaptability of Weining cattle. In addition, *ZNF668* and *KAT8* genes were found with high F_{ST} values and low nucleotide diversity (**Figure 4B**).

In addition, KEGG and GO terms ($p < 0.05$) were obtained. KEGG pathways are enriched in terms such as "Purine metabolism, bta00230," "MAPK signaling pathway, bta04010," "Wnt signaling pathway, bta04310," and "Endocrine resistance, bta01522" (**Supplementary Table S10**). Gene ontology (GO) terms showed that Weining cattle has increased GO categories involved in "negative regulation of chondrocyte differentiation, GO:0032331," "positive regulation of cartilage development, GO:

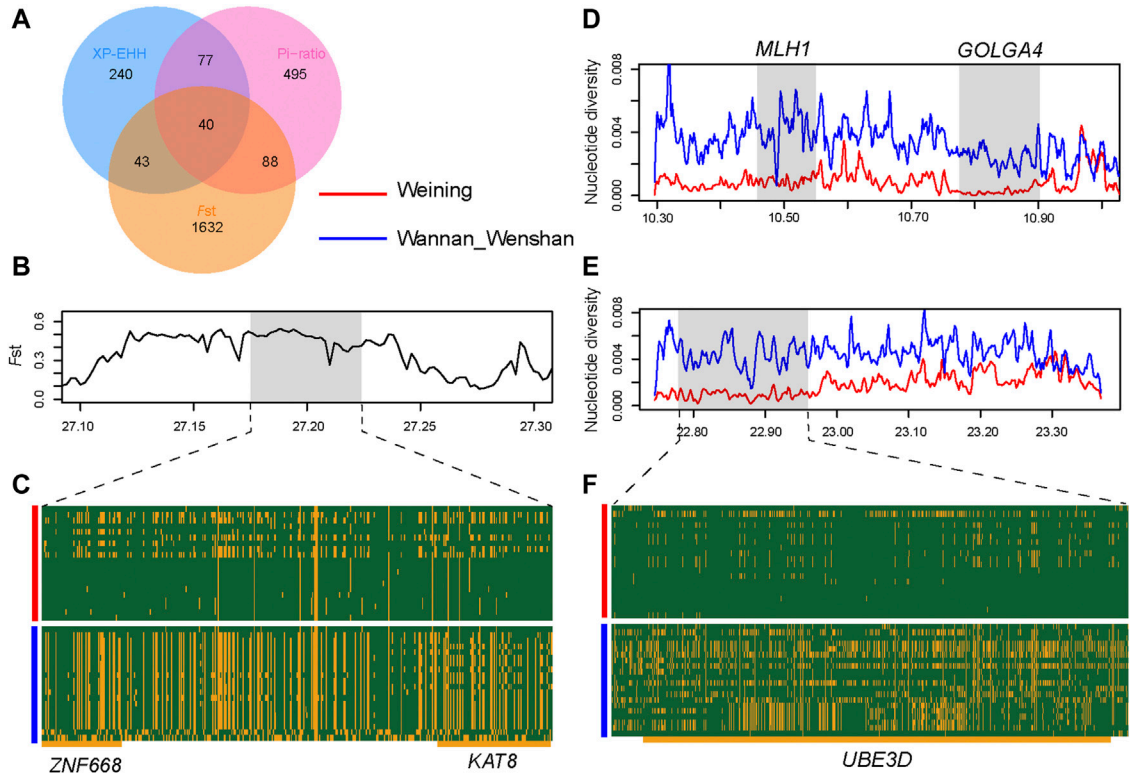

FIGURE 4 | Analysis of the signatures of positive selection in the genome of Weining. **(A)** Venn diagram showing the gene overlap among θπ-ratio, *F*ST, and XP-EHH. **(B)** F_{ST} at the *ZNF668* and *KAT8* gene region. **(C)** SNPs with minor allele frequencies >0.05 are used to construct haplotype patterns (Chr 25: 27.17–27.22 Mb). **(D)** Nucleotide diversity plots of the *MLH1* and *GOLGA4* genomic region. **(E)** Nucleotide diversity plots of the *UBE3D* genomic region. **(F)** SNPs with minor allele frequencies > 0.05 are used to construct haplotype patterns (Chr 9: 22.77–22.96 Mb). The major allele at each SNP position in Weining is colored in yellow, and the minor one in green.

0061036," "locomotory exploration behavior, GO:0035641," and "cellular response to hormone stimulus, GO:0032870" (**Supplementary Table S11**).

4 DISCUSSION

In this study, we analyzed 10 Weining cattle to well understand the complex population structure and high genetic diversity of Weining cattle. Mitochondrial and Y haplotypes showed that Weining cattle are a typical hybrid breed of *Bos taurus* × *Bos indicus* (Xia et al., 2019). The relatively high level of Y-chromosome variability was in accordance with the extensive mtDNA diversity in Weining cattle (Lei et al., 2006). Our result was also consistent with the non-autosomal genomic information of Weining cattle, that is, Weining cattle are hybrid of Chinese indicine and East Asian taurine, and in fact, this type of East Asian taurine and Chinese indicine dominates the crossbred type in China (Xia et al., 2021; Zhang et al., 2021). Multiple genetic backgrounds resulting from natural and artificial selection have occurred in Southwest China following the history of Chinese pastoralism. Furthermore, the pressure of rapid socioeconomic development has forced a sharp decrease in the number of groups in the Weining breed. According to breeding records, breeds such as Angus and Simmental have been introduced in the process of local breed improvement projects for the purpose of bringing more economic benefits to the locals. This partially explains the European taurine ancestry mixed into Weining cattle.

Genetic diversity in local breeds is a prerequisite for their continuous adaptation to the pressure of environmental changes. The relatively high level of genomic diversity found in Weining is likely the result of hybridization, which instills Weining cattle with not only components of *Bos taurus* but also components of *Bos indicus*. From genetic diversity analysis, Weining cattle have higher nucleotide diversity and lower inbreeding coefficient. Studies have shown that outer cattle breeds (bison, buffalo, yak, etc.) enriched the genetic diversity of Chinese indicine from the introgression (Chen et al., 2018). Interestingly, the Yunnan-Kweichow Plateau might act as an important channel for the Indian subcontinent breeds to enter China (Chen et al., 2020), making it possible for Indian indicine to mix into the genome of Weining cattle. As early as 2,500 years ago, indicus cattle extended to southern China along the east (Chen et al., 2009). More detailed information and history still need to be excavated and researched.

The fecundity of Weining cattle in mountainous areas is mainly manifested as late estrus time (China National

Commission of Animal Genetic Resources, 2011), which may be caused by many factors. Research shows that the admixture of *Bos taurus* × *Bos indicus* resulted in adaptability but caused a cost of reduced reproductive fitness due to genomic incompatibility (Kim et al., 2020). Within the Weining cattle genome, we calculated CLR and θπ-ratio to analyze the positive selection regions. In chromosome 22, *MLH1*, *GOLGA4*, *DLCK3*, and *ITGA9* genes may be affected by hitchhiking effect, resulting in selection signals. Although *GOLGA4* was highly expressed in the testis of mice, a study has already shown that the knockout mice did not show relevant male fertility disorders and the *GOLGA4* gene may exist only as a redundant one (Guo et al., 2020). In the adjacent region of the *GOLGA4* gene, we found that the *MLH1* gene also had a high signal, and it was one of the candidate genes related to heifer fertility (Shi et al., 2021). A close relationship between the cold environment and the poor fertility of Weining cattle was observed. Of course, this is our conjecture, and more theoretical supports were required to experiment.

Comparing two native cattle breeds (Wannan and Wenshan cattle) bred in hot (annual average temperature 15–20 °C) and humid environments (annual mean humidity 75%–85%), we found cold-resistant candidate genes in Weining cattle, for instance: *UBE3D* and *ZNF668*. In the current study, we have selected a region including the *UBE3D* gene by all the methods (Fst, XP-EHH, and Pi-ratio). *UBE3D*, ubiquitin-protein ligase E3D, is involved in intracellular physiological processes by regulating the ubiquitination process of regulatory proteins. The study has shown that Weining cattle had high unsaturated fatty acids (UFA) (linoleic acid, alpha-Linolenic acid, etc.) in southwest Chinese native breeds (Yang and Yang, 2010). Also, the function of the ubiquitin-proteasome system (UPS) might be regulated by fatty acids physiologically (Ando et al., 2004). Coincidentally, *UBE3D* was selected as the understanding of genetic mechanisms of fat composition in sheep (Ando et al., 2004). On the other hand, the UPS maintains the stability of endoplasmic reticulum function in brown adipose tissue (BAT) by degrading useless or damaged proteins, which plays an important role in cold-adapted metabolism (Bartelt et al., 2018). In addition, a number of zinc finger genes are reported to have an association between low ambient temperature and blood pressure (Lim et al., 2017; Xu et al., 2020). *ZNF536* was related to cold tolerance in Chantecler chickens (Xu et al., 2021). Herein, methylated changes of *ZNF668* might be involved in the elevation of blood pressure to hold body temperature when exposed to a cold environment (Xu et al., 2020). Compared with other cold regions, the genetic mechanisms underlying cold tolerance in Yanbian cattle in northern China (annual average temperature 2–6°C, annual mean humidity 68.6%) might explain the parallelism at the fatty acid point in Weining cattle (Yan et al., 2022).

Weining cattle are one of the potential beef cattle breeds in southwest China. The animal breeding culture of different ethnic minorities has contributed to the hybrid of Weining cattle. Meanwhile, the rich genetic diversity causes its remarkable adaptability in cold and humid mountains. Although the impact of the market economy forces the protection and development of Weining cattle under great pressure, the diversity and enrichment of genetic information in the genome of Weining cattle is a valued material for cattle breeding.

5 CONCLUSION

Analysis of genomic diversity and selection signatures of Weining cattle were carried out at a sequence level, and new insights into the genetic basis of crossbred cattle were provided. By comparing the Weining breed with other cattle populations, we introduce its genetic diversity, population genetic structure, and environmental adaptation characteristics. Moreover, a set of candidate genes were identified and may be related to cold adaptation, low fertility, and fatty acid composition in Weining cattle, although additional physiological and functional experiments are needed for verification. Overall, it is of great significance to understand the genetic diversity and adaptation of cattle breeds in southwest China.

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 Institutional Animal Care and Use Committee of Northwest A&F University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

YL performed the analysis of genome data and drafted the manuscript. HC carried out biological interpretation from the results and wrote the manuscript. SW and XL provided major assistance to the experiment and data analysis of the manuscript. XM and LS wrote and corrected the manuscript. JZ, JL, and MW collected samples, generated data from the sample, and contributed to the interpretation of the results. NC and KQ supervised and managed the whole study. BH and CL designed the experiment and provided the funding for this research. All authors read and approved the final manuscript.

FUNDINGS

This work was supported by the National Natural Science Foundation of China (No. 31872317), the China Agriculture Research System of MOF and MARA (No. CARS-37), the Program of Yunling Scholar and the Young and Middle-Aged

Academic Technology Leader Backup Talent Cultivation Program in Yunnan Province, China (No. 2018HB045), and the Yunnan Provincial Major S&T Project (No. 2019ZG007 and 2019ZG011).

ACKNOWLEDGMENTS

We would like to thank High-Performance Computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.848951/full#supplementary-material.

Supplementary Figure S1 | Analysis of the signatures of selection in the genome of Weining cattle. **(A)** CLR Manhattan plot of selective sweeps in Weining cattle. **(B)** θπ Manhattan plot of selective sweeps in Weining cattle. **(C)** Venn diagram showing the genes overlap among CLR and θπ. **(D)** Nucleotide diversity plots of the *CCNH* genomic region (Chr7: 87037314-87067889). **(E)** Plot of KEGG pathway analysis of Weining cattle candidate genes overlapped by θπ and CLR methods. Threshold (top 1%) of CLR and π were marked with a horizontal black line.

REFERENCES

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

Alshawi, A., Essa, A., Al-Bayatti, S., and Hanotte, O. (2019). Genome Analysis Reveals Genetic Admixture and Signature of Selection for Productivity and Environmental Traits in Iraqi Cattle. *Front. Genet.* 10, 609. doi:10.3389/fgene.2019.00609

Ando, H., Watabe, H., Valencia, J. C., Yasumoto, K., Furumura, M., Funasaka, Y., et al. (2004). Fatty Acids Regulate Pigmentation *via* Proteasomal Degradation of Tyrosinase: A New Aspect of Ubiquitin-Proteasome Function. *J. Biol. Chem.* 279 (15), 15427–15433. doi:10.1074/jbc.M313701200

Bartelt, A., Widenmaier, S. B., Schlein, C., Johann, K., Goncalves, R., Eguchi, K., et al. (2018). Brown Adipose Tissue Thermogenic Adaptation Requires Nrf1-Mediated Proteasomal Activity. *Nat. Med.* 24 (3), 292–303. doi:10.1038/nm.4481

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 (W1), W317–W325. doi:10.1093/nar/gkab447

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-Genome Resequencing Reveals World-Wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nat. Commun.* 9, 2337. doi:10.1038/s41467-018-04737-0

Chen, N., Ren, L., Du, L., Hou, J., Mullin, V. E., Wu, D., et al. (2020). Ancient Genomes Reveal Tropical Bovid Species in the Tibetan Plateau Contributed to the Prevalence of Hunting Game until the Late Neolithic. *Proc. Natl. Acad. Sci. U. S. A.* 117 (45), 28150–28159. doi:10.1073/pnas.2011696117

Chen, S., Lin, B.-Z., Baig, M., Mitra, B., Lopes, R. J., Santos, A. M., et al. (2009). Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic. *Mol. Biol. Evol.* 27 (1), 1–6. doi:10.1093/molbev/msp213

China National Commission of Animal Genetic Resources (2011). *Animal Genetic Resources in China Bovines*. Beijing: Chinese Agricultural Press. (In Chinese).

Cingolani, P., Platts, A., Wang, L. 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 M* Strain W1118; Iso-2; Iso-3. *Fly* 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

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genet.* 10, e1004254. doi:10.1371/journal.pgen.1004254

Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F. S., and Sargolzaei, M. (2018). Inbreeding and Runs of Homozygosity before and after Genomic Selection in North American Holstein Cattle. *BMC Genomics* 19 (1), 98. doi:10.1186/s12864-018-4453-z

Ghoreishifar, S. M., Eriksson, S., Johansson, A. M., Khansefid, M., Moghaddaszadeh-Ahrabi, S., Parna, N., et al. (2020). Signatures of Selection Reveal Candidate Genes Involved in Economic Traits and Cold Acclimation in Five Swedish Cattle Breeds. *Genet. Sel. Evol.* 52 (1), 52. doi:10.1186/s12711-020-00571-5

Guo, S., Chunyu lv, L., Ouyang, S., Wang, X., Liao, A., and Yuan, S. (2020). GOLGA4, a Golgi Matrix Protein, Is Dispensable for Spermatogenesis and Male Fertility in Mice. *Biochem. Biophys. Res. Commun.* 529 (3), 642–646. doi:10.1016/j.bbrc.2020.05.170

Hou, D., Qin, P., Niu, X., Li, T., Chen, B., Wei, C., et al. (2021). Genome-Wide Identification Evolution and Expression of Vestigial-Like Gene Family in Chicken. *Anim. Biotechnol.*, 1–11. doi:10.1080/10495398.2021.1920425

Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., et al. (2011). Whole-Genome Resequencing Shows Numerous Genes with Nonsynonymous SNPs in the Japanese Native Cattle Kuchinoshima-Ushi. *BMC Genomics* 12 (1), 103. doi:10.1186/1471-2164-12-103

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The Genome Landscape of Indigenous African Cattle. *Genome Biol.* 18 (1), 34. doi:10.1186/s13059-017-1153-y

Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., et al. (2020). The Mosaic Genome of Indigenous African Cattle as a Unique Genetic Resource for African Pastoralism. *Nat. Genet.* 52 (10), 1099–1110. doi:10.1038/s41588-020-0694-2

Lei, C. Z., Chen, H., Zhang, H. C., Cai, X., Liu, R. Y., Luo, L. Y., et al. (2006). Origin and Phylogeographical Structure of Chinese Cattle. *Anim. Genet.* 37 (6), 579–582. doi:10.1111/j.1365-2052.2006.01524.x

Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) V5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* 49 (W1), W293–W296. doi:10.1093/nar/gkab301

Li, H., Dai, Y., Luo, Z., and Nie, D. (2019). Cloning of a New Testis-Enriched Gene C4orf22 and its Role in Cell Cycle and Apoptosis in Mouse Spermatogenic Cells. *Mol. Biol. Rep.* 46 (2), 2029–2038. doi:10.1007/s11033-019-04651-8

Li, R., Li, C., Chen, H., Liu, X., Xiao, H., and Chen, S. (2019). Genomic Diversity and Admixture Patterns Among Six Chinese Indigenous Cattle Breeds in Yunnan. *Asian-Australas J. Anim. Sci.* 32 (8), 1069–1076. doi:10.5713/ajas.18.0605

Lim, Y. H., Han, C., Bae, S., and Hong, Y. C. (2017). Modulation of Blood Pressure in Response to Low Ambient Temperature: The Role of DNA Methylation of Zinc Finger Genes. *Environ. Res.* 153, 106–111. doi:10.1016/j.envres.2016.11.019

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

Nie, L., Yu, Y., Zhang, X.-Q., Yang, G.-F., Wen, J.-K., and Zhang, Y.-P. (1999). Genetic Diversity of Cattle in South china as Revealed by Blood Protein Electrophoresis. *Biochem. Genet.* 37 (7), 257–265. doi:10.1023/A:1018798924778

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic Scans for Selective Sweeps Using SNP Data. *Genome Res.* 15 (11), 1566–1575. doi:10.1101/gr.4252305

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

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genet.* 2 (12), e190. doi:10.1371/journal.pgen.0020190

Reid, G. (1991). Molecular Cloning: A Laboratory Manual, 2nd Edn by J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbor Laboratory Press, 1989. $115.00 (3 Vols; 1659 Pages) ISBN 0 87969 309 6. *Trends Biotechnol.* 9 (1), 213–214. doi:10.1016/0167-7799(91)90068-s

Rovadoscki, G. A., Pertile, S. F. N., Alvarenga, A. B., Cesar, A. S. M., Pértille, F., Petrini, J., et al. (2018). Estimates of Genomic Heritability and Genome-Wide Association Study for Fatty Acids Profile in Santa Inês Sheep. *BMC Genomics* 19 (1), 375. doi:10.1186/s12864-018-4777-8

Shen, J., Hanif, Q., Cao, Y., Yu, Y., Lei, C., Zhang, G., et al. (2020). Whole Genome Scan and Selection Signatures for Climate Adaption in Yanbian Cattle. *Front. Genet.* 11, 94. doi:10.3389/fgene.2020.00094

Shi, R., Brito, L. F., Liu, A., Luo, H., Chen, Z., Liu, L., et al. (2021). Genotype-by-Environment Interaction in Holstein Heifer Fertility Traits Using Single-Step Genomic Reaction Norm Models. *BMC Genomics* 22 (1), 193. doi:10.1186/s12864-021-07496-3

Singh, V., Jaiswal, D., Singh, K., Trivedi, S., Agrawal, N. K., Gupta, G., et al. (2019). Azoospermic Infertility is Associated with Altered Expression of DNA Repair Genes. *DNA Repair* 75, 39–47. doi:10.1016/j.dnarep.2019.01.006

Sweett, H., Fonseca, P. A. S., Suárez-Vega, A., Livernois, A., Miglior, F., and Cánovas, A. (2020). Genome-Wide Association Study to Identify Genomic Regions and Positional Candidate Genes Associated with Male Fertility in Beef Cattle. *Sci. Rep.* 10 (1), 20102. doi:10.1038/s41598-020-75758-3

Turan, I., Demir, K., Mengen, E., Kotan, L. D., Gürbüz, F., Yüksel, B., et al. (2021). *DLG2* Mutations in the Etiology of Pubertal Delay and Idiopathic Hypogonadotropic Hypogonadism. *Horm. Res. Paediat.* 94, 364–368. doi:10.1159/000520409

Xia, X., Yao, Y., Li, C., Zhang, F., Qu, K., Chen, H., et al. (2019). Genetic Diversity of Chinese Cattle Revealed by Y-SNP and Y-STR Markers. *Anim. Genet.* 50 (1), 64–69. doi:10.1111/age.12742

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22 (1), 43. doi:10.1186/s12864-020-07340-0

Xin, J.-W., Chai, Z.-X., Zhang, C.-F., Zhang, Q., Zhu, Y., Cao, H.-W., et al. (2019). Transcriptome Profiles Revealed the Mechanisms Underlying the Adaptation of Yak to High-Altitude Environments. *Sci. Rep.* 9 (1), 7558. doi:10.1038/s41598-019-43773-8

Xu, R., Li, S., Guo, S., Zhao, Q., Abramson, M. J., Li, S., et al. (2020). Environmental Temperature and Human Epigenetic Modifications: A Systematic Review. *Environ. Pollut. (Barking, Essex : 1987)* 259, 113840. doi:10.1016/j.envpol.2019.113840

Xu, N., Si, W., Li, M., Gong, M., Larivière, J. M., Nanaei, H. A., et al. (2021). Genome-Wide Scan for Selective Footprints and Genes Related to Cold Tolerance in Chantecler Chickens. *Zoolog. Res.* 42 (6), 710–720. doi:10.24272/j.issn.2095-8137.2021.189

Yan, C.-L., Lin, J., Huang, Y., Lin, J., Huang, Y.-Y., Gao, Q.-S., et al. (2022). Population Genomics Reveals that Natural Variation in PRDM16 Contributes to Cold Tolerance in Domestic Cattle. *Zool. Res.* 43 (2), 275–284. doi:10.24272/j.issn.2095-8137.2021.360

Yang, Z., and Yang, Z. (2010). Study on Nutrient Characteristics and Amino Acid Model of Meat Quality of 18 Guizhou Local Yellow Cattle. *Chin. J. Animal Husb.* 46 (09), 4–7. (In Chinese).

Yurchenko, A. A., Daetwyler, H. D., Yudin, N., Schnabel, R. D., Vander Jagt, C. J., Soloshenko, V., et al. (2018). Scans for Signatures of Selection in Russian Cattle Breed Genomes Reveal New Candidate Genes for Environmental Adaptation and Acclimation. *Sci. Rep.* 8 (1), 12984. doi:10.1038/s41598-018-31304-w

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, X., Qu, K., Jia, P., Zhang, J., Liu, J., Lei, C., et al. (2021). Assessing Genomic Diversity and Productivity Signatures in Dianzhong Cattle by Whole-Genome Scanning. *Front. Genet.* 12, 719215. doi:10.3389/fgene.2021.719215

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: 18 July 2022
doi: 10.3389/fgene.2022.858970

Genetic and Genomic Characterization of a New Beef Cattle Composite Breed (Purunã) Developed for Production in Pasture-Based Systems

*Henrique Alberto Mulim[1], Luiz F. Brito[2], Luís Fernando Batista Pinto[1], José Luis Moletta[3], Lilian Regina Da Silva[4] and Victor Breno Pedrosa[1,2,5]**

[1]Department of Animal Science, Federal University of Bahia, Salvador, Brazil, [2]Department of Animal Sciences, Purdue University, West Lafayette, IN, United States, [3]Agronomic Institute of Paraná—IAPAR, Ponta Grossa, Brazil, [4]NEOGEN Corporation, Pindamonhangaba, Brazil, [5]Department of Animal Science, State University of Ponta Grossa, Ponta Grossa, Brazil

Purunã is a composite beef cattle breed, developed in Southern Brazil by crossing the Angus, Charolais, Canchim, and Caracu breeds. The goal of this study was to perform the first genetic characterization of the Purunã breed, based on both pedigree and genomic information. For this, 100 randomly selected animals were genotyped, and 11,205 animals born from 1997 to 2019 had pedigree information. The genetic analyses performed were principal component analysis, admixture, phylogenic tree, pedigree and genomic inbreeding, linkage disequilibrium (LD), effective population size (Ne), consistency of the gametic phase, runs of homozygosity (ROH), heterozygosity-enriched regions (HERs), and functional analyses of the ROH and HER regions identified. Our findings indicate that Purunã is more genetically related to the Charolais, Canchim, and Angus breeds than Caracu or Nellore. The levels of inbreeding were shown to be small based on all the metrics evaluated and ranged from –0.009 to 0.029. A low (–0.12–0.31) correlation of the pedigree-based inbreeding compared to all the genomic inbreeding coefficients evaluated was observed. The LD average was 0.031 (±0.0517), and the consistency of the gametic phase was shown to be low for all the breed pairs, ranging from 0.42 to 0.27 to the distance of 20 Mb. The Ne values based on pedigree and genomic information were 158 and 115, respectively. A total of 1,839 ROHs were found, and the majority of them are of small length (<4 Mb). An important homozygous region was identified on BTA5 with pathways related to behavioral traits (sensory perception, detection of stimulus, and others), as well as candidate genes related to heat tolerance (*MY O 1A*), feed conversion rate (*RDH5*), and reproduction (*AMDHD1*). A total of 1,799 HERs were identified in the Purunã breed with 92.3% of them classified within the 0.5–1 Mb length group, and 19 HER islands were identified in the autosomal genome. These HER islands harbor genes involved in growth pathways, carcass weight (*SDCBP*), meat and carcass quality (*MT2A*), and marbling deposition (*CISH*). Despite the genetic relationship between Purunã and the founder breeds, a multi-breed genomic evaluation is likely not feasible due to their population structure and low consistency of the gametic phase among them.

Keywords: beef cattle, genomic diversity, inbreeding coefficient, persistency of the gametic phase, runs of heterozygosity, runs of homozygosity

OPEN ACCESS

Edited by:
Lingyang Xu,
Chinese Academy of Agricultural Sciences, China

Reviewed by:
Jose Bento Sterman Ferraz,
University of São Paulo, Brazil
Ali Esmailizadeh,
Shahid Bahonar University of Kerman, Iran
Yaokun Li,
South China Agricultural University, China

****Correspondence:***
Victor Breno Pedrosa
vbpedrosa@uepg.br

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

Received: *20 January 2022*
Accepted: *10 June 2022*
Published: *18 July 2022*

Citation:
Mulim HA, Brito LF, Batista Pinto LF, Moletta JL, Da Silva LR and Pedrosa VB (2022) Genetic and Genomic Characterization of a New Beef Cattle Composite Breed (Purunã) Developed for Production in Pasture-Based Systems. Front. Genet. 13:858970. doi: 10.3389/fgene.2022.858970

1 INTRODUCTION

The characterization of the population structure and genetic diversity is essential for the understanding of the genetic background of environmental adaptation and conservation of cattle genetic resources (Xia et al., 2021). Such characterization and diversity assessment need to be considered when designing or updating breeding programs and conservation strategies that can be applied in purebred and crossbred populations.

The Purunã breed is a composite population developed in Southern Brazil by crossing Angus, Charolais, Canchim, and Caracu, in identical proportions. This was performed to improve key traits of interest and exploit the complementarity among the breeds (Otto et al., 2021), especially for production in pasture-based systems. The background research to generate the Purunã breed started at the Agronomic Institute of Paraná (IAPAR; Ponta Grossa, Paraná, Brazil) at the beginning of the 1980s, when IAPAR researchers estimated the heterosis in the crossbred progenies of Charolais x Caracu and Angus x Canchim (Perotto et al., 2000a; Perotto et al., 2000b). Almost 15 years later, the first results were obtained from this experiment, where the heterosis retained from those crosses resulted in higher hot carcass weight, hot carcass yield, rib-eye area, better carcass conformation from Charolais x Caracu (Perotto et al., 2000b), and higher average weight daily gain in different ages with the crosses of Angus x Canchim (Perotto et al., 2000a).

Based on the first results, IAPAR researchers conducted a second mating to generate another set of animals using the progenies resulting from the previous F1 population. The goal at that point was to combine all favorable characteristics in a composite that presented a heavyweight and produce a high-quality carcass. The hypothesis for using the breeds mentioned earlier was to capture a particular contribution from each breed to create a composite population with higher productive performance and adapted to the tropical and subtropical regions of Brazil. The Angus breed provided traits related to precocity, more docile temperament, and high meat quality (Cristiana and Mirela, 2018; Taye et al., 2018); Charolais provided a higher weight gain and carcass yield (Jahuey-Martínez et al., 2019), and finally, Caracu and Canchim contributed with rusticity, heat tolerance, and parasite resistance (Urbinati et al., 2016; Pires et al., 2021). Such animals are very well adapted to tropical environmental conditions and showed good potential to gain weight (Ito et al., 2010). The characterization of Purunã, defined by the Brazilian Purunã Cattle Breed Association (Ponta Grossa, Paraná, Brazil), is that the animals must present short hair with a shiny aspect, admitting variation on the coat color (red, white, black, and bay), medium-to-large size, and good muscle distribution as shown in **Figure 1**. Additionally, the animals are expected to be docile and prolific, with sexual precocity and fast carcass finishing.

Some studies in Brazil have evaluated the performance of the Purunã breed for carcass traits (Ito et al., 2010), meat production and quality traits (Missio et al., 2015), growth (Moura et al., 2014), and weight at different time points (Otto et al., 2021). These studies indicate animals slaughtered at 24 months of age, weighed an average of 460 kg, possessed a fat thickness close to 3 mm, and had a substantial concentration of fatty acids in the meat. In addition, estimates of genetic parameters for growth traits have demonstrated heritable estimates (0.05–0.21) for body weights measured at different ages (Otto et al., 2021). However, no previous research to date has evaluated the genetic diversity and population structure of the Purunã breed. Although the crossbreeding increases the genetic variability of a population, their development history and population management across generations could have impacted the genetic diversity of the population formed (Peripolli et al., 2020). Genetic diversity studies play an important role in the constitution of a crossbreeding program since how the variability is controlled may interfere with the heterosis produced and impact the expected hybrid vigor.

Genetic diversity studies are crucial in the initial phase, called pre-breeding, in which it is possible to regenerate, characterize, explore, and promote the conservation of variability of the population (Pontes et al., 2020). The parameters estimated include inbreeding coefficients of the individual animals, the genetic relationship between animals, and overall levels of homozygous and heterozygous regions in the genome as well as their distribution along the chromosomes (Biscarini et al., 2020). Furthermore, linkage disequilibrium needs to be estimated for better implementing genomic selection and for identifying conserved segments of the genome among breeds (Larmer et al., 2014). All these metrics contribute to a better understanding of the genetic events that happen in the population, the impact of decisions made in the past, and the strategies that will be taken in the future. Our goal with this study was to characterize the genetic and genomic diversity and the population structure of a new composite beef cattle breed—Purunã, based on genomic and pedigree information.

2 MATERIAL AND METHOD

One hundred animals of the Purunã breed were randomly sampled and genotyped using the GGP Bovine 100K array (GGP, 2021) containing over 100,000 single nucleotide polymorphisms (SNPs). The genetic material was provided by the Agronomic Institute of Paraná (IAPAR, Ponta Grossa, Paraná, Brazil). For the genotype quality control (QC), only autosomal chromosomes were retained and a QC was performed separately for each analysis. For runs of homozygosity (ROH) and heterozygous-enriched regions (HER), we removed SNPs with call rate lower than 0.90, duplicated position, non-autosomes, or without a known position (Ferenčaković et al., 2013; Biscarini et al., 2020). For the other analyses, minor allele frequency (MAF <0.05) and extreme departure from the Hardy-Weinberg equilibrium (HWE $<10^{-6}$) parameters were also used to filter out SNPs.

For the pedigree database, information from 11,205 animals born between 1997 and 2019 was considered, where 5,224 were males and 5,981 females, and the base population was formed by 3,999 animals. These data were used to create the pedigree database including information on individual animals, sire, dam, sex, and birth date.

2.1 Population Stratification

2.1.1 Principal Component Analysis

To assess the similarities between the Purunã breed and Angus, Canchim, Charolais, and Nellore breeds, we performed a principal component analysis (PCA) by PLINK v1.9 software (Purcell et al., 2007). The genotypes of the Angus, Charolais, and Canchim were retrieved from the WIDDE database (Sempéré et al., 2015), and the Nellore breed genotypes were provided by the Katayama Agropecuaria Ltda breeding company. The PCA was estimated based on the standardized variance of the genomic relationship matrix (**G**) where the covariance of each SNP was divided by the respective variance, using only the SNPs in common for all breeds (after the QC), as the following equation proposed by (VanRaden, 2008):

$$\boldsymbol{G} = \frac{(\boldsymbol{M} - 2\boldsymbol{P})(\boldsymbol{M} - 2\boldsymbol{P})'}{2\Sigma p_i(1 - p_i)}, \tag{1}$$

where **M** is a matrix of counts of allele A, p_i is the frequency of allele A of *i*th SNP, and **P** is a matrix with each row containing the p_i values.

2.1.2 Admixture Analysis

The admixture analysis was performed using the ADMIXTURE software (Alexander et al., 2015) to assess the evolutionary history between the Purunã breed and its founder breeds (Angus, Charolais, Canchim, and Nellore). This analysis estimates ancestries by efficiently computing maximum likelihood estimates in a parametric model as (Alexander and Lange, 2011):

$$\mathcal{L}(\boldsymbol{Q}, \boldsymbol{F}) = \sum_{ij} \left\{ n_{ij} \middle| np_{ij} + \left(2 - n_{ij}\right) \middle| n\left(1 - p_{ij}\right) \right\}, \tag{2}$$

where p_{ij} is the success probability in the binomial distribution n_{ij} ~ Bin(2, p_{ij}) depending on the fraction q_{ik} of i's ancestry attributable to population *k* and on the frequency f_{kj} of allele 1 in the population *k*. The matrices **Q** = (q_{ik}) and **F** = (f_{kj}).

The success of the analysis is dependent on the correct choice of *K*, which represents the number of ancestral populations. We evaluated K equal to 1 until 20, but only K = 2 and 3 were chosen to be shown here, which have more biological interpretation and K = 3 had the smallest cross-validation error. The "pong" package (Behr et al., 2016) was used to cluster the results and visualize the population structure.

2.1.3 Phylogenetic Tree

To estimate the distance among the populations, we used the hapFLK software (Fariello et al., 2013) based on the approach described by Bonhomme et al. (2010). The neighbor-joining tree was built from the Reynolds' genetic distances (Reynolds et al., 1983) between pairs of populations. Reynold's distance was estimated using the co-ancestry coefficient, where this coefficient is the probability that a random pair of genes at the same locus within a randomly chosen population is identical-by-descent, providing a natural measure of genetic drift. It is assumed that the allele frequency is equal to

$$\hat{p}_0 = \frac{1'n\mathcal{F}^{-1}p}{1'n\mathcal{F}^{-1}1_n}, \tag{3}$$

where p is the frequency, $\mathcal{F}$ is the co-ancestry matrix, and $\hat{p}_0$ is the unbiased linear estimate with minimum variance, with 1′n denoting the n-vector made of 1's.

2.2 Population Structure

2.2.1 Inbreeding Metrics

Six models of inbreeding coefficient estimates were analyzed. The first model was based on pedigree information (F_{PED}), using the ENDOG v4.8 software (Gutiérrez et al., 2010), following the method proposed by Meuwissen and Luo (1992) in which the average F of a given generation t (F_t) was calculated as follows:

$$F_t = 1 - (1 - \Delta F)^t, \tag{4}$$

in which ΔF is the change in the inbreeding rate from one generation to another, as the following equation:

$$\Delta F = \frac{(F_t - F_{t-1})}{(1 - F_{t-1})}, \tag{5}$$

in which F_t and F_{t-1} represent the average inbreeding estimates for the current and the previous generation (Falconer and Mackay, 1996).

The second method was based on the homozygous genotypes observed and expected (F_{HOM1}), calculated as follows (Purcell et al., 2007):

$$F_{HOM1} = \frac{H_{exp} - H_{obs}}{H_{exp}}, \tag{6}$$

where H_{exp} is the expected value (proportion) for homozygous genotypes, and H_{obs} is the observed value for the homozygous genotypes.

The third method was based on genotype additive variance (F_{GRM}), using the following model (VanRaden, 2008):

$$F_{GRM} = \frac{[x_i - 2p_i]^2}{h_i - 1} \text{ in which } h_i = 2p_i(1 - p_i), \tag{7}$$

where x_i is the number of reference allele copies of the *i*th SNP, and p_i is the reference allele frequency in the population. Similar to the second method, the methodology F_{HOM2} was based on homozygous genotypes following the model:

$$F_{HOM2} = 1 - \frac{x_i * (2 - x_i)}{h_i}. \tag{8}$$

The aforementioned models are all dependent on the genotype allele frequency, and for this reason, a fifth model was a test based on the correlation between uniting gametes (F_{UNI}) using the following model Yang et al. (2010):

$$F_{UNI} = \frac{\left[x_i^2 - (1 + 2p_i) * x_i + 2p_i^2\right]}{h_i}. \tag{9}$$

The last method was based on the sum of ROH individual length divided by the total length of the autosomal genome (F_{ROH}) using the following equation (McQuillan et al., 2008):

$$F_{ROH} = \frac{\sum_{i=i}^{n} f(ROH_i)}{\sum_{j=1}^{A} h(j)}, \tag{10}$$

FIGURE 1 | Purunã animals from the Agronomic Institute of Paraná (IAPAR, Ponta Grossa, Parana, Brazil).

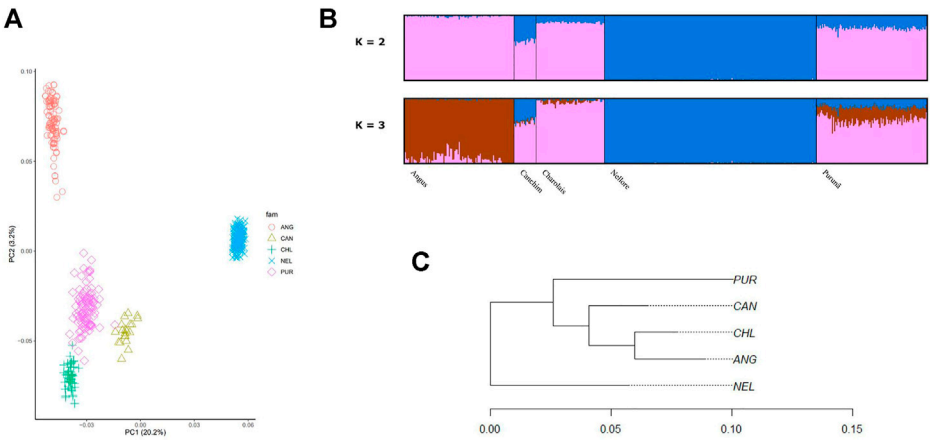

FIGURE 2 | Population stratification of the Purunã breed. **(A)** Principal component analysis (PCA) including Purunã, Angus, Canchim, Charolais, and Nellore breed animals. **(B)** Admixture analysis of Purunã, Angus, Canchim, Charolais, and Nellore breeds. **(C)** Phylogenetic tree using Reynold's distance for the Purunã (PUR), Angus (ANG), Canchim (CAN), Charolais (CHL), and Nellore (NEL) populations.

where $f\,(ROH_i)$ is the ROH length of individual ith, n is the total number of homozygous genomic regions of each individual, $h(j)$ is the length of chromosome jth, and A is the number of autosomal chromosomes (A = 29). Still, for each class of ROH (<2 Mb, 2–4 Mb, 4–8 Mb, 4–16 Mb, >16 Mb, <8 Mb, and >8 Mb), inbreeding estimates were obtained by dividing the total sum of ROH segments by the total length of the cattle autosomal genome covered by SNPs. All the genomic inbreeding coefficients were calculated using the PLINK v1.9 software (Purcell et al., 2007). The PROC CORR option of the SAS statistical software (SAS Institute Inc., 2013) was used to correlate the inbreeding coefficient estimates. A heatmap was created for better visualization of the results through the "plotly" package (Sievert, 2020).

2.2.2 Linkage Disequilibrium

The linkage disequilibrium (r^2) was estimated by PLINK v1.9 software. To observe the r^2 decrease along with the increase in the marker distance, we used the binning approach estimating the r^2 average of each distance from 10 to 100 kb in each 10 kb, and after the distance of 100 kb in each 100 kb until the distance of 1,000 kb (1 Mb). As a preliminary analysis, we defined that the bins reported in this study were required to have at least 50 pairwise markers to estimate the binned average of r^2.

2.2.3 Effective Population Size

Two methodologies were used to estimate the effective population size (Ne). The first method used pedigree information through the following equation:

$$Ne = \frac{1}{2}\Delta F, \text{Where} \Delta F = \frac{(F_t - F_{t-1})}{(1 - F_{t-1})}, \quad (11)$$

where F_t and F_{t-1} are the average inbreeding of offspring and their parents, respectively (Falconer and Mackay, 1996). The estimate was performed using the POPREP software (Groeneveld et al., 2009).

The second method was performed using genomic information, and investigated with the relationship method between LD variances and Ne through the following formula (Corbin et al., 2012):

$$N_{e(T)} = \left(4f(c_t)^{-1}\left(E\left[r^{2|ct}\right]^{-1} - \alpha\right)\right), \quad (12)$$

where Ne is the effective population size at the *t*th generation, c_t is the recombination rate for the physical distance between the markers, α is the probability for the occurrence of mutation, and r^2 is the LD value.

2.2.4 Consistency of the Gametic Phase

The consistency of the gametic phase (CGP) was taken by the square root of r^2 values adding the sign from the disequilibrium metric (D), as:

$$D = p(ab) - p(a)p(b), \quad (13)$$

where *p(a)* is the frequency of the haplotype-a, *p(b)* is the frequency of the haplotype-b, and p_{ab} is the haplotype frequency with allele *a* on the first locus and allele *b* on the second locus. The CGP was assumed as the Person correlation between each founder breed and Purunã using the signed-squared-root values. To estimate the CGP, only the SNPs in common (after the quality control) between each breed pair were used to estimate the CGP based on the same distance and bin described in the LD section.

2.3 Proportion of Polymorphic SNPs and Distributions of SNPs by the MAF Range

The proportion of polymorphic SNPs, after QC, was calculated based on the MAF. The distributions of SNPs were calculated on 10 MAF ranges from 0 to 0.5 defined every 0.05 points in MAF.

2.4 Runs of Homozygosity

The PLINK v1.9 software was used for the ROH identification based on the following criteria:

- One heterozygous and one missing SNP were allowed;
- The window of the threshold used was 0.05;
- The gap between consecutive SNPs could not be higher than 1,000 kb;
- The minimum length of an ROH was 500 kb;
- The minimum number of consecutive SNPs that create an ROH must be equal to or greater than 30;
- The density of 1 SNP used in at least 50 kb;
- A sliding genomic window was used with 50 SNPs.

ROHs were classified in the following classes: <2 Mb, 2–4 Mb, 4–8 Mb, 4–16 Mb, and >16 Mb (Lozada-Soto et al., 2021; Mulim et al., 2022). A region found in 36% of the population was considered for future analysis (functional and phylogenetic analysis).

2.5 Heterozygosity-Enriched Regions

The detectRUNS package (Biscarini et al., 2019) was used for the detection of HER following the consecutive-SNPs method. For the SNPs' consecutive analysis, the following parameters were considered:

- a minimum number of 20 consecutive SNPs constitutes an HER;
- a minimum length of 500 kb;
- a minimum of two homozygous and one missing SNP is allowed; and
- the maximum gap between consecutive SNPs could not be higher than 1,000 kb.

The genomic regions that showed at least 10% of the animals with HER were included in the subsequent functional analyses and phylogenetic tree.

2.6 Functional Analyses

The genomic regions considered as ROH and HER islands were used for genomic annotations. The GALLO package (Fonseca P. A. S. et al., 2020) was used for the annotation of genes in these regions, with the annotated data for *Bos taurus* from the Ensembl database (www.ensembl.org/Bos_taurus/Info/Index), version ARS-UCD1.2 (Rosen et al., 2020). Subsequently, the WebGestaltR package (Wang et al., 2020) was used for the Gene Ontology (GO) analyses to identify biological processes, molecular functions, and cellular components in which the positional candidate genes are involved in.

3 RESULTS

3.1 Population Stratification

3.1.1 Principal Component Analysis

The PCA among the populations of Purunã, Angus, Canchim, Charolais, and Nellore is presented in **Figure 2A**. The first principal component (PC1) explained 20.2% of the variation among the populations, while the second principal component (PC2) accounted for 3.2%. As shown in **Figure 2A**, the animals are grouped within breeds, with no clear mixture between groups, even for composite populations such as Purunã. The breeds closer to the Purunã are Charolais, Canchim, and Angus.

3.1.2 Admixture Analysis

Figure 2B presented the admixture analysis for Purunã, Angus, Canchim, Charolais, and Nellore populations for K = 2 and 3. For K = 2, two groups were observed and the mixture between them indicates that two distinct founder populations (*Bos taurus taurus* and *Bos taurus indicus*) were used when developing the Purunã breed. In average, Angus had 99.2% and 0.9%, Canchim 60.6% and 39.4%, Charolais 89.6% and 10.4%, Nellore 99.8% and 0.2%, and Purunã had 80.8% and 19.2% from ancestral population

TABLE 1 | Inbreeding coefficient estimates with different methodologies for animals of the Purunã breed.

	N	Mean	Std Dev	Minimum	Maximum
F_{PED1}	11,205	0.002	0.019	0.000	0.375
F_{PED2}	100	0.007	0.023	0.000	0.125
F_{HOM1}	100	−0.009	0.027	−0.052	0.163
F_{GRM}	100	−0.009	0.041	−0.092	0.095
F_{HOM2}	100	−0.009	0.032	−0.060	0.171
F_{UNI}	100	−0.009	0.023	−0.052	0.133
F_{ROH}	100	0.029	0.024	0.004	0.190
$F_{<2MB}$	100	0.004	0.002	0.000	0.010
$F_{2\text{-}4MB}$	100	0.007	0.003	0.001	0.023
$F_{4\text{-}8MB}$	100	0.007	0.005	0.000	0.021
$F_{8\text{-}16MB}$	100	0.006	0.007	0.000	0.032
$F_{>16MB}$	100	0.005	0.016	0.000	0.127
$F_{<8MB}$	100	0.018	0.007	0.002	0.045
$F_{>8MB}$	100	0.011	0.020	0.000	0.145

N: number of individuals analyzed; Mean: average of inbreeding coefficient; Std Dev: standard deviation.
F_{PED1}: inbreeding coefficient based on the pedigree for all individual in the Purunã breed.
F_{PED2}: inbreeding coefficient based on the pedigree for Purunã genotyped individuals.
F_{HOM1}: inbreeding coefficient based on the number of observed and expected homozygous genotypes.
F_{GRM}: inbreeding coefficient based on additive genotypic variance.
F_{HOM2}: inbreeding coefficient based on homozygosity of genotypes.
F_{UNI}: inbreeding coefficient based on the correlation between uniting gametes.
F_{ROH}: inbreeding coefficient based on the length of the ROH's and the total length of the autosomal genome.

1 and 2, respectively. For K = 3, three groups were observed to affect the admixture analysis for the population in the study. This result (K = 3) indicates more contribution from Charolais and Canchim in the Purunã breed, following the Angus breed and a small proportion of the Nellore breed.

3.1.3 Phylogenetic Tree

Figure 2C shows the genomic population tree for the breeds Purunã (PUR), Angus (ANG), Charolais (CHL), Canchim (CAN), and Nellore (NEL). There is a division into groups on the tree but the distance from one group to another is not high (0.05). Nellore appears in one section while Purunã, Canchim, Charolais, and Angus are situated in three other nodes, grouping in accord with clades of breed proximity.

3.2 Population Structure

3.2.1 Inbreeding

The averages of inbreeding coefficients are presented in **Table 1**.

The average for the inbreeding coefficient estimated based on pedigree for all Purunã individuals (F_{PED2}) was 0.002. The methods F_{HOM1}, F_{HOM2}, F_{UNI}, and F_{GRM} were the methods showing the lowest average values (−0.009), while the highest inbreeding coefficient average was obtained by the F_{ROH} metric (0.029). The correlations among the inbreeding coefficients method are presented in **Figure 3**.

Strong correlations were found between the methods: F_{HOM1}-F_{HOM2} (0.97), F_{HOM1}-F_{ROH} (0.93), F_{HOM1}-$F_{>16MB}$ (0.85), F_{HOM1}-$F_{>8MB}$ (0.89), F_{HOM2}-F_{ROH} (0.90), F_{HOM2}-$F_{>16MB}$ (0.79), and F_{HOM2}-$F_{>8MB}$ (0.85). Low values were found for all correlations among F_{PED} and the other methods. The F_{GRM} method shows a very weak correlation for almost all the methods, except for the $F_{UNI,}$ when the correlation was classified as moderate (0.74). Negative correlations were found for the methods: F_{GRM}-F_{HOM2} (−0.21), F_{GRM}-$F_{<2MB}$ (−0.24), F_{GRM}-F_{2-4MB} (−0.08), F_{GRM}-F_{4-8MB} (−0.01), F_{GRM}-F_{8-16MB} (−0.21), F_{GRM}-$F_{<8MB}$ (−0.11), and F_{GRM}-F_{PED} (−0.12).

3.2.2 Linkage Disequilibrium

The average LD ranged from 0.43 to 0.04, with a distance between two markers of 10 to 1,000 kb, respectively. The general average of LD was 0.031 (±0.0517) at the average distribution of the markers 4.856 (±2.8890) Mb. The decrease in LD with an increase in the marker distance can be observed in **Supplementary Figure S1**.

3.2.3 Effective Population Size

The effective population size based on pedigree was 158 for the current generation. On the other hand, the genomic-based Ne differed based on the generation and software used. The SNeP software (Barbato et al., 2015) and the PLINK software enabled the estimation of Ne up to the 13th and 5th generation back, respectively. For the SNeP, in the 13th generation, the Ne was 229, while for the PLINK, the result for the same generation was 207. The Ne estimated for the 5th generation on PLINK was equal to 115.

3.2.4 Consistency of the Gametic Phase

The consistency of the gametic phase between Purunã and Angus, Canchim, Charolais, and Nellore is presented in **Table 2**.

The highest correlation, at 20 kb between SNP pairs, between Purunã and the other breeds was found with Charolais (0.43), followed by Canchim (0.42), Angus (0.40), and Nellore (0.27). The distance of 10 kb showed a lower number of pairwise markers than the threshold (<50) used as a criterion. Therefore, these results were not presented.

3.3 Proportion of Polymorphic SNPs and Distribution of SNPs by MAF Range

The proportion of polymorphic SNPs based on the MAF category were as follows: $MAF_{0.00\text{-}0.05}$ 3,232 (3.67%); $MAF_{0.05\text{-}0.10}$ 2,967 (3.37%); $MAF_{0.10\text{-}0.15}$ 4,242 (4.82%); $MAF_{0.15\text{-}0.20}$ 5,792 (6.59%); $MAF_{0.20\text{-}0.25}$ 7,255 (8.25%); $MAF_{0.25\text{-}0.30}$ 9,314 (10.59%); $MAF_{0.30\text{-}0.35}$ 11,287 (12.83%); $MAF_{0.35\text{-}0.40}$ 13,283 (15.10%); $MAF_{0.40\text{-}0.45}$ 14,932 (16.98%); and $MAF_{0.45\text{-}0.50}$ 15,652 (17.80%).

3.4 Runs of Homozygosity

A total of 1,839 ROHs were found for the Purunã breed. The distribution along all autosomal genomes can be observed in **Figure 4A** and the ROH length size division. The length of ROH observed here can be classified as 37.4% for <2 Mb; 25.3% as 2–4 Mb; 17.1% as 4–8 Mb; 7.6% as 8–16 Mb; and only 2.6% ROH greater than 16 Mb. The chromosome that presented the highest amount of ROHs was the BTA5, followed by the BTA1, where the concentration of ROHs >16 Mb was superior compared to all other autosomes. The chromosomes that showed the smallest number of ROHs were the BTA27 and BTA25, representing a small fraction of regions in ROH.

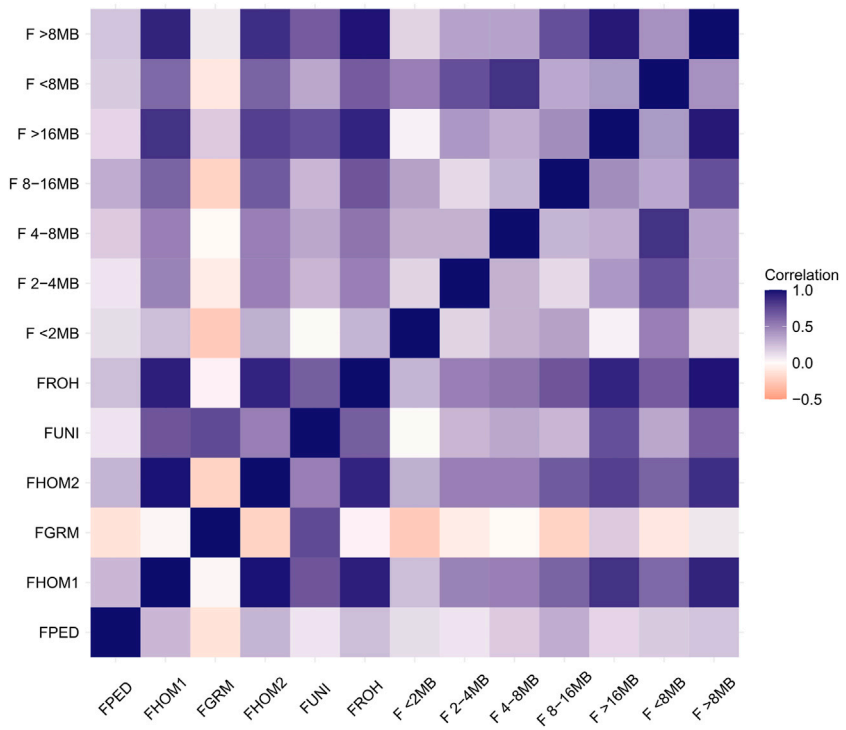

FIGURE 3 | Correlation among inbreeding estimation methods.

TABLE 2 | Consistency of the gametic phase based on Pearson correlation, between the Purunã breed and its founder breeds: Angus, Canchim, Charolais, and Nellore breeds.

Distance (kb)	Angus	Canchim	Charolais	Nellore
20	0.40	0.42	0.43	0.27
30	0.36	0.40	0.42	0.18
40	0.30	0.36	0.41	0.17
50	0.29	0.33	0.39	0.14
60	0.29	0.33	0.39	0.13
70	0.26	0.33	0.35	0.11
80	0.24	0.30	0.33	0.08
90	0.24	0.30	0.33	0.04
100	0.20	0.27	0.31	0.04
200	0.18	0.24	0.25	0.04
300	0.12	0.20	0.18	0.03
400	0.10	0.16	0.16	0.02
500	0.10	0.13	0.12	0.02
600	0.10	0.13	0.12	0.02
700	0.09	0.11	0.11	0.01
800	0.09	0.11	0.10	0.01
900	0.08	0.11	0.10	0.00
1,000	0.07	0.09	0.10	0.00

3.5 Heterozygous-Enriched Regions

In total, 1,799 HERs were found in the Purunã breed. The HER pattern distribution along with all autosomal genomes is shown in **Figure 4B**. Around 92.3% of the HERs found were classified in the length of 0.5–1.0 Mb; 7.0% as 1.0–1.5 Mb; and 0.7% as 1.5–2.0 Mb. No HER greater than 2 Mb was found for Purunã. The chromosome that presented the highest amount of HER was the BTA1, while BTA25 had the smallest number of HER.

3.6 ROH and HER Islands and Functional Analyses

3.6.1 Runs of Homozygosity

With the ROH analysis, we found a common region in homozygosity present in 36% of the animals, despite the fact that Purunã is a recently-developed composite breed. This region is located on BTA5 between 54, 304, 681 bp and 62, 031, 799 bp, and has a length of 7.73 Mb, where 131 SNPs are present in this region. This region is responsible for coding 220 genes, with 181 protein-coding genes, seven pseudogenes, five long non-coding RNA, nine microRNA, ten miscellaneous RNA, six small nucleolar RNA, and two small nuclear RNA. The list of all the genes found in this region is presented in **Supplementary Table S1**. The significant Gene Ontology (GO) terms ($p < 0.05$) in which these genes are part of are presented in **Table 3**.

Ten biological processes, three molecular functions, and six cellular components were identified in the significant pathways. Interestingly, pathways linked to animal behavior were found in this region, including sensory perception (GO: 0007600), detection of stimulus (GO:0051606), response to extracellular stimulus (GO:0009991), olfactory receptor activity (GO:0004984), and others. To track the origin of this homozygous region in Purunã, we performed a

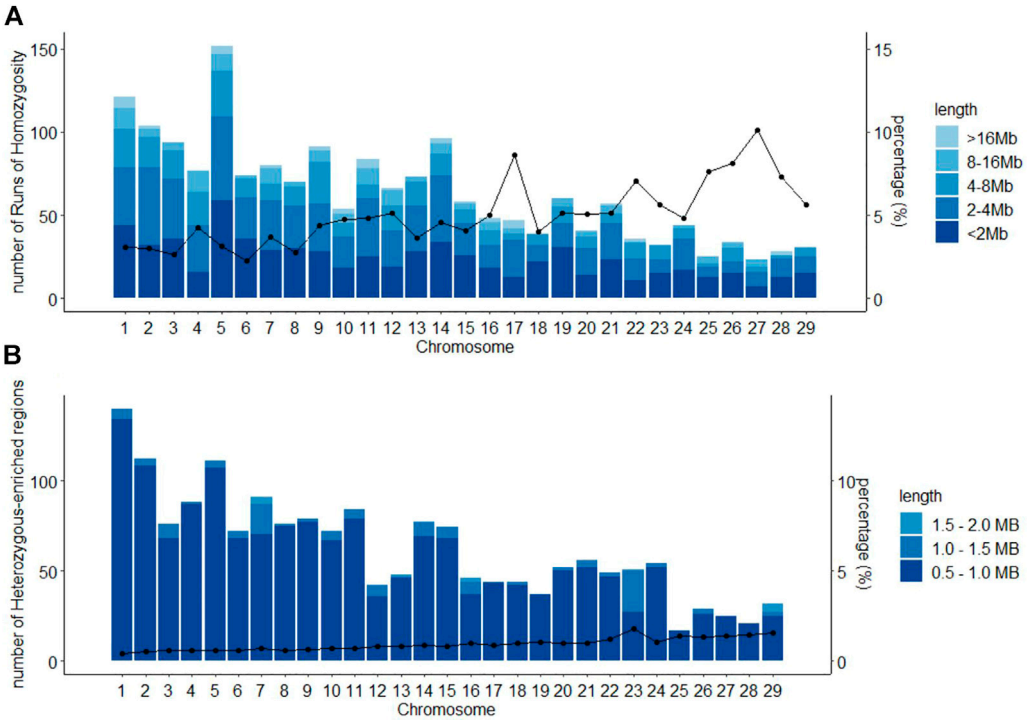

FIGURE 4 | Classification of runs of homozygosity ROH **(A)** and heterozygous-enriched regions HER **(B)**, by chromosome, according to the length size in the Purunã breed, and the average percentage of chromosome covered by ROHs/HERs.

phylogenetic tree analysis, using only the SNPs allocated in this region. **Figure 5A** shows the phylogenetic tree for the homozygous region found in BTA5. In this particular region, the breeds Purunã, Charolais, and Angus are closer together in comparison to the Canchim and Nellore breed, indicating that Charolais and Angus might have contributed to key behavioral characteristics observed in the Purunã breed.

3.6.2 Heterozygous-Enriched Regions

For the HER analysis, the regions identified in at least 10% of the animals were considered as HER islands and used to verify the candidate genes and pathways. **Table 4** presents the HER island found in the Purunã breed.

We found 19 HERs distributed in 17 chromosomes, where the BTA5 and BTA14 presented two HERs in each chromosome. The most frequent HER (27% of the population) was found in BTA23. The longest HER was found in BTA22 with a length size of 3.67 Mb, and the smallest HER was found in the BTA15 at 0.60 Mb.

All these regions are responsible for coding 413 genes, including 363 protein-coding, six pseudogenes, 13 long non-coding RNAs, seven microRNAs, two miscellaneous RNAs, one small nucleolar RNA, 15 small nuclear RNAs, three processed pseudogenes, and three ribosomal RNAs. The list of all the genes found in these regions is presented in **Supplementary Table S2**. The significant GO terms ($p < 0.05$) and their related genes are presented in **Table 5**.

In total, we found 17 significant GO terms involved in biological processes, eight in molecular functions, and six in cellular components. Interesting regions related to the growth pathways (GO:0,040,007) were found in the heterozygous-enriched regions in BTA10, BTA14, BTA18, BTA19, and BTA20, where 14 genes are acting in higher variability in the population. The genes are *PPIB* (peptidylprolyl isomerase B), *SDCBP* (syndecan binding protein), *MT2A* and *MT3* components of metallothionein, *RAI1* (retinoic acid-induced 1), *FLCN* (folliculin), *DCAF1* (DNA damage-binding protein 1), *CISH* (cytokine-inducible SH2), *HYAL1* and *HAYAL2* components of hyaluronidase, *SEMA3B* and *SEMA3F* components of semaforin, *ARIH2* (ariadne RBR E3 ubiquitin protein ligase 2), and IP6K2 (inositol hexakisphosphate kinase 2). To track the origin of these heterozygous regions in Purunã, we used a phylogenetic tree analysis. **Figure 5B** shows the phylogenetic tree for the regions related to the growth pathway found in the heterozygous-enriched regions.

4 DISCUSSION

Our main goal in this study was to genetically characterize the Purunã breed by estimating genetic diversity and population structure parameters based on both genomic and pedigree information. This breed was developed by crossing, in the same proportion, Charolais, Canchim, Angus, and Caracu breeds, while

TABLE 3 | Significant ($p < 0.05$) Gene Ontology (GO) terms for the genes located within runs of homozygosity regions in the Purunã breed.

	Description	*p*-Value	Genes
		Biological process	
GO: 0,043,648	Dicarboxylic acid metabolic process	0.001	*SHMT2; GLS2; AMDHD1; HAL*
GO: 0,007,600	Sensory perception	0.004	*MY O 1A; MIP; RDH5; OR10P1; ENSBTAG00000047825; ENSBTAG00000046778; ENSBTAG00000048295; ENSBTAG00000002913; OR10A7; ENSBTAG00000037629*
GO: 0,043,473	Pigmentation	0.008	*DCTN2; PMEL; CD63*
GO: 0,006,520	Cellular amino acid metabolic process	0.014	*MARS1; SHMT2; GLS2; AMDHD1; HAL*
GO: 0,006,091	Generation of precursor metabolites and energy	0.015	*NDUFA4L2; SHMT2; PTGES3; CS; COQ10A; BLOC1S1*
GO: 0,044,282	Small molecule catabolic process	0.019	*CYP27B1; SHMT2; GLS2; AMDHD1; HAL*
GO: 0,051,606	Detection of the stimulus	0.022	*OR10P1; ENSBTAG00000047825; ENSBTAG00000046778; ENSBTAG00000048295; ENSBTAG00000002913; OR10A7; ENSBTAG00000037629*
GO: 0,007,422	Peripheral nervous system development	0.023	*NAB2; ERBB3*
GO: 0,006,766	Vitamin metabolic process	0.041	*CYP27B1; SHMT2*
GO: 0,009,991	Response to the extracellular stimulus	0.047	*CYP27B1; DDIT3; MARS1; SLC39A5*
		Molecular function	
GO: 0,004,984	Olfactory receptor activity	0.005	*OR10P1; ENSBTAG00000047825; ENSBTAG00000046778; ENSBTAG00000048295; ENSBTAG00000002913; OR10A7; ENSBTAG00000037629*
GO: 0,016,741	Transferase activity, transferring one-carbon groups	0.018	*EEF1AKMT3; METTL1; SHMT2; METTL7B*
GO: 0,000,049	tRNA binding	0.032	*METTL1; MARS1*
		Cellular component	
GO: 0,009,295	Nucleoid	0.017	*SHMT2; ATP5F1B*
GO: 0,016,328	Lateral plasma membrane	0.018	*MY O 1A; ERBB3*
GO: 0,045,177	Apical part of the cell	0.022	*MY O 1A; MIP; ERBB3; NEDD1*
GO: 0,005,759	Mitochondrial matrix	0.031	*TSFM; SHMT2; ATP5F1B; CS; BLOC1S1*
GO: 0,098,687	Chromosomal region	0.034	*DCTN2; MBD6; NABP2; CDK2*
GO: 0,005,788	Endoplasmic reticulum lumen	0.045	*OS9; RDH5*

Canchim is also a composite breed that has Charolais and Nellore as the main founder breeds. Therefore, the average genetic proportion for Purunã is 13/32 Charolais, 8/32 Caracu, 8/32 Angus, and 3/32 Nellore.

Based on the results from the population stratification section, the Purunã breed seems to be genetically closer to Charolais, Canchim, and Angus, with the highest genomic contribution from the Charolais breed (**Figure 2B**, K = 3). Yet, our findings indicate that Purunã is closer to the *Bos taurus taurus* than *Bos taurus indicus* breeds. This was expected due to the greater contribution of taurine breeds in the formation of the Purunã breed.

4.1 Inbreeding Metrics

The maintenance of low levels of inbreeding is also desirable in composite breeds, once an advantage of crossbreeding is heterosis. Such heterosis is influenced by the genetic distance between the parental breeds and the level of inbreeding in the population, which can affect the degree of heterosis retention (Peripolli et al., 2020). As shown in **Table 1**, the inbreeding coefficient for all the metrics estimated in this study ranged from −0.009 (±0.041) to 0.029 (±0.024). These results are expected as the Purunã breed is a recently developed composite breed. The low level of inbreeding, with more emphasis on negative values (outbreeding), indicates that the probability of the two homologous genes within an individual being identical-by-descendent is smaller than two homologous genes drawn at random from the reference population and the ancestry shared into the population is small (Wang, 2014). However, in terms of gain or loss variability to a reference base population (Villanueva et al., 2021), the values indicated

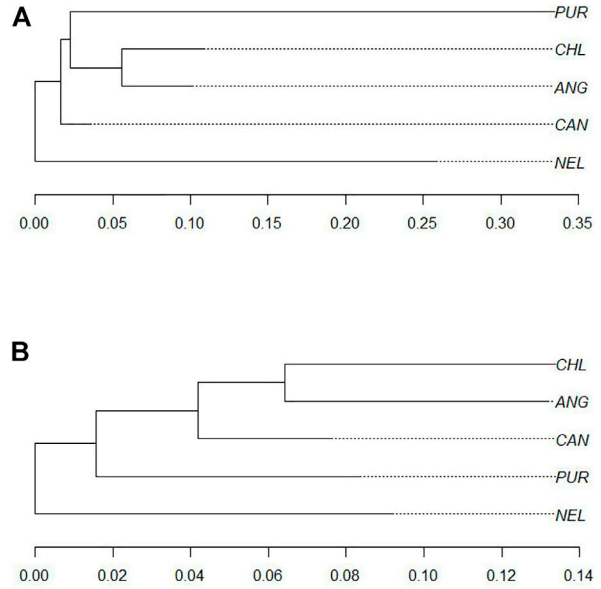

FIGURE 5 | Phylogenetic trees. **(A)** Phylogenetic tree for homozygous regions comparing the Purunã (PUR), Angus (ANG), Canchim (CAN), Charolais (CHL), and Nellore (NEL) populations. **(B)** Phylogenetic tree for growth pathway in heterozygous-enriched regions comparing the Purunã (PUR), Angus (ANG), Canchim (CAN), Charolais (CHL), and Nellore (NEL) populations.

TABLE 4 | Heterozygous-enriched regions (HER) which appear in at least 10% of Purunã individuals.

CHR	%	BP1	BP2	nSNP	Length
BTA1	11	26,505,838	29,555,484	22	3,049,646
BTA2	17	42,384,465	43,575,039	23	1,190,574
BTA3	10	8,435,805	9,838,978	23	1,403,173
BTA5	14	70,752,944	72,012,890	23	1,259,946
BTA5	14	75,043,240	75,983,135	26	939,895
BTA6	11	27,154,761	28,275,511	21	1,120,750
BTA7	13	8,562,310	10,432,630	24	1,870,320
BTA10	11	44,820,482	46,032,038	25	1,211,556
BTA11	14	67,243,961	69,096,131	22	1,852,170
BTA12	10	40,237,435	41,970,427	25	1,732,992
BTA14	17	24,167,298	25,953,073	24	1,785,775
BTA14	16	50,608,626	51,640,291	23	1,031,665
BTA15	10	1,215,097	1,819,862	30	604,765
BTA18	10	23,515,690	24,470,198	23	954,508
BTA19	12	34,233,799	35,283,135	25	1,049,336
BTA20	10	44,099,958	45,220,153	21	1,120,195
BTA22	10	48,961,009	52,638,988	21	3,677,979
BTA23	27	26,021	1,697,122	27	1,671,101
BTA24	10	40,975,659	41,855,725	21	880,066

CHR: chromosome.
%: percentage of the population that presented this island.
BP1: position in the base pair where the HER start.
BP2: position in the base pair where the HER end.
nSNP: number of SNPs, that HER covers.
Length: HER, length.

that some variability had been gained (through migration or gene flow from other populations) or, in the cases where the inbreeding coefficient was positive, a slight loss of variability.

An accurate measure of F_{PED} could be expected when a complete, deep (many generations recorded) and no (or few) errors in the pedigree files. In the case of the Purunã breed, the information in the pedigree files goes to, on average, 2.35 generations. As the Purunã pedigree is shallow, the use of genomic information to estimate the inbreeding coefficients is a great alternative to access the inbreeding levels of the individuals. These inbreeding metrics, in addition, are not dependent on pedigree information, taking into account the Mendelian sampling variation (Doekes et al., 2019), the stochastic nature of recombination (Ferenčaković et al., 2013), and correcting the pedigree failure to assume that the founders of a population are unrelated (Rebelato et al., 2018). Yet, some metrics not only measure the levels of overall inbreeding but also give an estimate of when the inbreeding was created, as in the case of F_{ROH}.

The F_{ROH} captures the highest level of inbreeding, especially because the F_{ROH} metric is capable of capturing both recent and more ancient inbreeding (Ghoreishifar et al., 2020). As shown in **Table 1**, the value for the ancient inbreeding coefficient (F < 8 MB) is higher than more recent inbreeding (F > 8 MB). Such ancient inbreeding could be provided by ancient generations in ancient mating and still be in the population passing it through generations. This division between ancient and recent inbreeding is helpful to manage the diversity in the population. As not all inbreeding is expected to be equally unfavorable, recent inbreeding is expected to have more negative effects than ancient inbreeding (Doekes et al., 2019), therefore maintaining a low level of recent inbreeding coefficient is a desirable goal. **Figure 3** illustrates the correlation among the inbreeding coefficient metrics. All the metrics showed a low correlation with F_{PED}. Some authors have already mentioned that the genomic inbreeding metrics are more accurate in assessing individual inbreeding (Curik et al., 2014; Marras et al., 2015; Doekes et al., 2019). This happens due to the particularities mentioned before about the pedigree estimation, but as the F_{PED}, each metric used to calculate the genomic inbreeding coefficient has its specificities and captures a different type of inbreeding that was originally defined by Wright (1922) and/or Malécot (1948).

The genomic metrics vary according to the weight that each marker gets to find the **G** matrix or the allele frequency for each marker (Howard et al., 2017). This affects how the inbreeding is calculated for each individual and the correlation among the metrics. The metrics F_{HOM} and F_{ROH} weigh all the alleles equally, while the metrics F_{UNI} and F_{GRM} give more weight to rare alleles (Alemu et al., 2021). This could explain why the metrics F_{HOM} and F_{ROH}, and F_{GRM} and F_{UNI} show moderate to strong correlation, while the F_{GRM} and F_{HOM} or the F_{ROH} classes had a negative correlation.

4.2 Linkage Disequilibrium, Effective Population Size, and Consistency of the Gametic Phase

Higher LD values were observed for markers located closer to each other and a faster decreased LD values were found as the

TABLE 5 | Significant ($p < 0.05$) Gene Ontology (GO) terms, to biological process, to heterozygous-enriched regions found in the Purunã breed.

	Description	*p*-value	Genes
	Biological process		
GO: 0,051,270	Regulation of cellular component movement	0.003	*SLAMF1; RAC2; ZNF609; SDCBP; FLCN; MAP2K3; IQCF1; HYAL2; HYAL1; SEMA3B; SEMA3F; MST1; DAG1; RHOA; ELP6; PTPRM*
GO: 0,010,563	Negative regulation of the phosphorus metabolic process	0.012	*PWP1; ELFN2; RTRAF; FLCN; HYAL2; INKA1; DAG1; RHOA; QARS1; PRKAR2A*
GO: 0,090,407	Organophosphate biosynthetic process	0.013	*CD244; PIGM; LPCAT2; PRPSAP2; PEMT; FLCN; IP6K1; IMPDH2; IPGK2; TREX1; NME6*
GO: 0,072,521	Purine-containing compound metabolic process	0.018	*ATP1A2; PTGDR; PRPSAP2; SHMT1; FLCN; RHOA; IMPDH2; UQCRC1; TREX1; NME6; NDUFV2*
GO: 0,040,007	Growth	0.019	*PPIB; SDCBP; MT2A; MT3; RAI1; FLCN; DCAF1; CISH; HYAL2; HYAL1; SEMA3B; SEMA3F; ARIH2; IP6K2*
GO: 0,007,187	G protein-coupled receptor signaling pathway, coupled to the cyclic nucleotide second messenger	0.019	*PTGDR; GNA O 1; GRM2; GNAI2; GNAT1; PTH1R*
GO: 0,055,086	Nucleobase-containing small molecule metabolic process	0.020	*ATP1A2; PTGDR; PRPSAP2; SHMT1; NT5M; FLCN; GMPPB; RHOA; IMPDH2; UQCRC1; TREX1; NME6; NDUFV2*
GO: 0,031,647	Regulation of protein stability	0.038	*PEX19; PPIB; MT3; COPS3; USP4; TREX1*
GO: 1,901,657	Glycosyl compound metabolic process	0.040	*PTGDR; PRPSAP2; IMPDH2; NME6*
GO: 0,001,505	Regulation of neurotransmitter levels	0.046	*ATP1A2; KCNJ10; SYN3; SLC6A2; SHMT1; AMT*
GO: 0,001,667	Ameboidal-type cell migration	0.048	*MAP2K3; HYAL2; HYAL1; SEMA3B; SEMA3F; RHOA; PTPRM*

distance between the markers increased, as observed in other crossbred or composite populations (Prieur et al., 2017; Deng et al., 2019). The extent of LD is strongly influenced by the population history, particularly in domestic animal populations, which have undergone bottlenecks during both domestication and the subsequent formation of breeds (Brito et al., 2015). Such LD is directly related to genomic selection, where the number of markers required to accurately predict breeding values depends on the LD (Larmer et al., 2014). Following the proposed equation by (McKay et al., 2007), the number of markers required for accurate genomic selection will be around 95,000 markers (2.67 GB/30 kb at LD = 0.2) for the Purunã breed. However, it is essential to highlight that for an implementation of genomic selection in Purunã, it is crucial that a sizable training population needs to be generated, to provide accurate genomic predictions of breeding values and selection.

Analysis of LD plays a central role in many areas of population genetics, including the determination of genetic maps, ascertainment of levels of recombination at the population level, and Ne estimation (D'Ambrosio et al., 2019). Based on all the metrics, the Ne estimates for Purunã are higher than 100 in the current generations, which is a threshold proposed by Meuwissen (2009) to ensure long-term population sustainability. The Ne estimate based on LD was able to be detected up to the fifth generation ago using the PLINK software (Purcell et al., 2007). We observed a slight divergence between the results from the SNeP and PLINK software, but not as high as reported by Barbato et al. (2015).

Understanding the LD levels, population structure, and CGP across breeds are crucial for implementing genomic selection (Brito et al., 2015). The CGP for all the evaluated breeds, including the Purunã, resulted in a low correlation, as shown in **Table 2**. These results indicate that the markers' phase (or the phase between markers and QTL) is not consistent across breed pairs. In this context, the possible use of a multi-breed training population for genomic evaluations using these breeds (Purunã, Charolais, Canchim, and Angus) might not result in more accurate genomic breeding values. As the markers are not in the same phase across breeds, the ability to use one breed to determine the effects of SNP to aid in the selection of another population becomes less likely (Larmer et al., 2014).

4.3 Runs of Homozygosity

Figure 4A shows the number of ROHs found by chromosome in Purunã. The BTA5 showed a higher number of ROHs, as also observed in other beef cattle studies (Peripolli et al., 2018; Peripolli et al., 2020). The majority of ROH found (62.7%) were classified as short ROHs, and as the length of ROH is negatively correlated with the co-ancestry (Mastrangelo et al., 2018), the ROH found in Purunã were conceived in a more ancient generation. Taking the length of ROH and using the studies that estimate the ROH and correlate with the generation, as the work of Howrigan et al. (2011), the majority of ROH found in this study was created between 10 and 20 generations ago.

The ROH can be used for genome characterization and a better understanding of the implications of selection pressure (Marras et al., 2018). An interesting region was identified in the BTA5, which contains significant pathways related to behavioral traits. The first pathway was the sensory

perception associated with a series of events required for an organism to receive a sensory stimulus, convert it to a molecular signal, and recognize and characterize the signal (AmiGO, 2021c). The second was the detection of a stimulus pathway related to a stimulus received by a cell or organism. This pathway converts a signal into a response to an extracellular stimulus, associating any movement, secretion, enzyme production, or gene expression in an extracellular stimulus (AmiGO, 2021a). The third pathway was olfactory receptor activity, a pathway related to combining with an odorant and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity in response to the detection of smell (DeMaria and Ngai, 2010).

To track the origin of such a region, a phylogenetic tree (**Figure 5A**) was made to evaluate which breed could provide this region. As shown in **Figure 5A**, the Purunã, Charolais, and Angus animals seem to be genetically closer, and therefore, Angus and Charolais might have contributed to this region. Some studies have previously reported the same region with a high incidence of homozygous sequence (Szmatoła et al., 2019; Fabbri et al., 2021). Some interesting genes, already mentioned in the literature as candidate genes were identified in this region, as *MY O 1A* (Myosin IA) related to bovine heat-tolerance (Jia et al., 2019), *RDH5* (11-cis retinol dehydrogenase 5) associated with feed conversion (de Almeida Santana et al., 2016), and *AMDHD1* (amidohydrolase domain containing 1) related to reproduction (Moravčíková et al., 2019).

4.4 Heterozygous-Enriched Region

Maintaining diversity at a locus may be advantageous for fitness and could be subject to balancing or countervailing selection (Williams et al., 2016). These heterozygous-enriched regions are single nucleotide differences observed between paternal and maternal chromosomes and can reveal much about the population structure and demographic history (Santos et al., 2021). As shown in **Figure 4B**, the majority of HER found in this study were classified as shorter HER. Interestingly, one region is already mentioned as a conserved region for beef and dairy cattle in the BTA14 (Zhao et al., 2015). This region is variable in at least 17% of the Purunã individuals, demonstrating that even in more conservative regions, the crossbred could provide some variability to the animals.

Although some studies have shown that the majority of HER islands are related to immunity to diseases (Williams et al., 2016), survival rate, and fertility (Biscarini et al., 2020), an interesting pathway was found in our study related to growth. The growth pathway is a biological process related to the increase in size or mass of an entire organism, a part of an organism, or a cell (AmiGO, 2021b). In the case of *PPIB*, a gene used as a reference gene in studies of gene expression (Costa et al., 2013; da Costa et al., 2013) or the *SDCBP* gene, a possible candidate gene related to carcass weight in Hanwoo, a Korean native breed (Lee et al., 2013) and Montana Tropical Composite, a composite beef cattle population developed in Brazil (Grigoletto et al., 2020). Another gene mentioned as a possible candidate gene for meat quality and carcass yield was the *MT2A*, which is involved in glucocorticoid response and with metal and antioxidant biological responds (Haegeman et al., 2003). Although the *CISH* gene is directly related to insulin metabolism, Fonseca L. F. S. et al. (2020) indicated that this gene could play an essential role in marbling deposition. Our tracking of this pathway was not possible to define a unique breed responsible to provide such a region of HER to Purunã breed or even the breeds where the population is closer, as shown in **Figure 5B**. This means that such variability is not provided by a unique or small group of breeds, but by the mixture of the breeds used in the creation of the Purunã breed.

5 CONCLUSION

As observed in the admixture analyses, the Purunã breed received a more significant genetic contribution to its formation from Charolais, Canchim, and Angus. The inbreeding levels for Purunã were small based on multiple inbreeding metrics. Higher LD values were observed for markers with small distances and a faster decrease associated with an increase in the distance between the markers (ranging from 0.43 to 0.04 with distance of 10–1,000 Kb), indicating that a denser panel of markers is necessary to achieve higher levels of accuracy in a genomic selection of Purunã. A high Ne (>100) was observed in all metrics evaluated and the consistency of gametic phase analyses resulted in a small correlation among all breeds, which determines that a multi-breed genetic evaluation for Purunã might not be advantageous. An interesting homozygous region was found in the BTA5 with significant pathways related to behavior and genes related to traits such as heat-tolerance (*MY O 1A*), feed conversion rate (*RDH5*), and reproduction (*AMDHD1*). This could indicate a possible pressure of selection in such regions. For the heterozygosity, the number of HER was elevated, but this was expected since Purunã is a composite breed. Among HER regions, an interesting pathway related to growth was identified with higher variability, containing genes previously associated with carcass weight (*SDCBP*), meat and carcass quality (*MT2A*), and marbling deposition (*CISH*).

DATA AVAILABILITY STATEMENT

The Purunã genotypes data are available in the OSF Repository (https://osf.io/7p6wt/). Genotypes from Angus, Canchim, and Charolais are available in the WIDDE database (http://widde.toulouse.inra.fr/widde/). The Nellore datasets presented in this article are not readily available because genotypes from

databases Katayama are not publicly available but can be obtained through a reasonable request via the corresponding author. Requests to access the datasets should be directed to vbpedrosa@uepg.br.

ETHICS STATEMENT

Ethical review and approval were not required for the animal study because no new animals were handled in this experiment.

AUTHOR CONTRIBUTIONS

HM, LFBR, and VP conceived, designed, and conducted the data analyses. JM, LS, and VP contributed to the data acquisition. HM, LFBR, LD, and VP wrote and edited the manuscript. All authors reviewed and contributed to editing of the manuscript and approved its final publication.

ACKNOWLEDGMENTS

The authors thank the Agronomic Institute of Paraná (IAPAR), for providing the pedigree and hair samples for DNA extraction, Katayama Ltd. and NEOGEN Corporation for providing the genomic information. The first author also acknowledges the Research Support Foundation of the State of Bahia (FAPESB) for the doctoral scholarship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.858970/full#supplementary-material

REFERENCES

Alemu, S. W., Kadri, N. K., Harland, C., Faux, P., Charlier, C., Caballero, A., et al. (2021). An Evaluation of Inbreeding Measures Using a Whole-Genome Sequenced Cattle Pedigree. *Heredity* 126, 410–423. doi:10.1038/s41437-020-00383-9

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinforma.* 12, 1–6. doi:10.1186/1471-2105-12-246

Alexander, D. H., Shringarpure, S. S., Novembre, J., and Lange, K. (2015). *Admixture 1.3 Software Manual*. Cold Spring Harbor Lab.

AmiGO (2021a). AmiGO 2: Term Details for "Detection of Stimulus" (GO: 0051606). Available at: http://amigo.geneontology.org/amigo/term/GO:0051606 (Accessed November 18, 2021).

AmiGO (2021b). AmiGO 2: Term Details for "Growth" (GO:0040007). Available at: http://amigo.geneontology.org/amigo/term/GO:0040007 (Accessed November 24, 2021).

AmiGO (2021c). AmiGO 2: Term Details for "Sensory Perception" (GO:0007600). Available at: http://amigo.geneontology.org/amigo/term/GO:0007600 (Accessed November 18, 2021).

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

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data. *Bioinformatics* 32, 2817–2823. doi:10.1093/BIOINFORMATICS/BTW327

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. (2019). *detectRUNS: An R Package to Detect Runs of Homozygosity and Heterozygosity in Diploid Genomes*. Milan: CRAN.R.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., and Ciampolini, R. (2020). Insights into Genetic Diversity, Runs of Homozygosity and Heterozygosity-Rich Regions in Maremmana Semi-feral Cattle Using Pedigree and Genomic Data. *Animals* 10, 2285. doi:10.3390/ani10122285

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics* 186, 241–262. doi:10.1534/GENETICS.110.117275

Brito, L. F., Jafarikia, M., Grossi, D. A., Kijas, J. W., Porto-Neto, L. R., Ventura, R. V., et al. (2015). Characterization of Linkage Disequilibrium, Consistency of Gametic Phase and Admixture in Australian and Canadian Goats. *BMC Genet.* 16, 67. doi:10.1186/S12863-015-0220-1

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. Anim. Breed. Genet.* 129, 257–270. doi:10.1111/j.1439-0388.2012.01003.x

Costa, A. S. H., Silva, M. P., Alfaia, C. P. M., Pires, V. M. R., Fontes, C. M. G. A., Bessa, R. J. B., et al. (2013). Genetic Background and Diet Impact Beef Fatty Acid Composition and Stearoyl-CoA Desaturase mRNA Expression. *Lipids* 48, 369–381. doi:10.1007/s11745-013-3776-4

Cristiana, A. D., and Mirela, C. S. (2018). Study on Technological Aspects Related to the Growth of the Aberdeen Angus Breed in Romania. *Ann. Food Sci. Technol.* 19, 1–5.

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

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., et al. (2019). Genome-wide Estimates of Genetic Diversity, Inbreeding and Effective Size of Experimental and Commercial Rainbow Trout Lines Undergoing Selective Breeding. *Genet. Sel. Evol.* 51, 26–15. doi:10.1186/S12711-019-0468-4/FIGURES/6

da Costa, A. S. H., Pires, V. M. R., Fontes, C. M. G. A., and Mestre Prates, J. A. (2013). Expression of Genes Controlling Fat Deposition in Two Genetically Diverse Beef Cattle Breeds Fed High or Low Silage Diets. *BMC Vet. Res.* 9, 118. doi:10.1186/1746-6148-9-118

de Almeida Santana, M. H., Junior, G. A. O., Cesar, A. S. M., Freua, M. C., da Costa Gomes, R., da Luz e Silva, S., et al. (2016). Copy Number Variations and Genome-wide Associations Reveal Putative Genes and Metabolic Pathways Involved with the Feed Conversion Ratio in Beef Cattle. *J. Appl. Genet.* 57, 495–504. doi:10.1007/s13353-016-0344-7

DeMaria, S., and Ngai, J. (2010). The Cell Biology of Smell. *J. Cell Biol.* 191, 443–452. doi:10.1083/jcb.201008163

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

Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., and Windig, J. J. (2019). Inbreeding Depression Due to Recent and Ancient Inbreeding in Dutch Holstein-Friesian Dairy Cattle. *Genet. Sel. Evol.* 51, 54. doi:10.1186/s12711-019-0497-z

Fabbri, M. C., Dadousis, C., Tiezzi, F., Maltecca, C., Lozada-Soto, E., Biffani, S., et al. (2021). Genetic Diversity and Population History of Eight Italian Beef Cattle Breeds Using Measures of Autozygosity. *PLoS ONE* 16, e0248087. doi:10.1371/journal.pone.0248087

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th ed. New York: Longman Essex.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting Signatures of Selection through Haplotype Differentiation Among

Hierarchically Structured Populations. *Genetics* 193, 929–941. doi:10.1534/GENETICS.112.147231

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

Fonseca, L. F. S., dos Santos Silva, D. B., Gimenez, D. F. J., Baldi, F., Ferro, J. A., Chardulo, L. A. L., et al. (2020). Gene Expression Profiling and Identification of Hub Genes in Nellore Cattle with Different Marbling Score Levels. *Genomics* 112, 873–879. doi:10.1016/j.ygeno.2019.06.001

Fonseca, P. A. S., Suárez-Vega, A., Marras, G., and Cánovas, Á. (2020). GALLO: An R Package for Genomic Annotation and Integration of Multiple Data Sources in Livestock for Positional Candidate Loci. *Gigascience* 9, 1–9. doi:10.1093/gigascience/giaa149

GGP (2021). *Neogen GeneSeek® Genomic Profiler™ Bovine 100K*. NEOGEN.

Ghoreishifar, S. M., Moradi-Shahrbabak, H., Fallahi, M. H., Jalil Sarghale, A., Moradi-Shahrbabak, M., Abdollahi-Arpanahi, R., et al. (2020). Genomic Measures of Inbreeding Coefficients and Genome-wide Scan for Runs of Homozygosity Islands in Iranian River Buffalo, Bubalus Bubalis. *BMC Genet.* 21, 16. doi:10.1186/s12863-020-0824-y

Grigoletto, L., Ferraz, J. B. S., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical Composite Beef Cattle. *Front. Genet.* 11, 123. doi:10.3389/fgene.2020.00123

Groeneveld, E., Westhuizen, B. V. D., Maiwashe, A., Voordewind, F., and Ferraz, J. B. S. (2009). POPREP: a Generic Report for Population Management. *Genet. Mol. Res.* 8, 1158–1178. doi:10.4238/vol8-3gmr648

Gutiérrez, J., Goyache, F., and Cervantes, I. (2010). *Endog v4.8 – a Computer Program for Monitoring Genetic Variability of Populations Using Pedigree Information*. User's Guide. Universidad Complutense de Madrid. Madrid, España, 45p.

Haegeman, A., Williams, J. L., Law, A., Van Zeveren, A., and Peelman, L. J. (2003). Mapping and SNP Analysis of Bovine Candidate Genes for Meat and Carcass Quality. *Anim. Genet.* 34, 349–353. doi:10.1046/j.1365-2052.2003.01008.x

Howard, J. T., Pryce, J. E., Baes, C., and Maltecca, C. (2017). Invited Review: Inbreeding in the Genomics Era: Inbreeding, Inbreeding Depression, and Management of Genomic Variability. *J. Dairy Sci.* 100, 6009–6024. doi:10.3168/jds.2017-12787

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

Ito, R. H., Prado, I. N. d., Visentainer, J. V., Prado, R. M. d., Fugta, C. A., and Pires, M. C. O. (2010). Carcass Characteristics, Chemical and Fatty Acid Composition of Longissimus Muscle of Purunã Bulls Slaughtered at 18 or 24 Months of Age. *Acta Sci. Anim. Sci.* 32, 299–307. doi:10.4025/actascianimsci.v32i3.7274

Jahuey-Martínez, F. J., Parra-Bracamonte, G. M., Sifuentes-Rincón, A. M., and Moreno-Medina, V. R. (2019). Signatures of Selection in Charolais Beef Cattle Identified by Genome-wide Analysis. *J. Anim. Breed. Genet.* 136, 378–389. doi:10.1111/jbg.12399

Jia, P., Cai, C., Qu, K., Chen, N., Jia, Y., Hanif, Q., et al. (2019). Four Novel SNPs of MYO1A Gene Associated with Heat-Tolerance in Chinese Cattle. *Animals* 9, 964. doi:10.3390/ANI9110964

Larmer, S. G., Sargolzaei, M., and Schenkel, F. S. (2014). Extent of Linkage Disequilibrium, Consistency of Gametic Phase, and Imputation Accuracy within and across Canadian Dairy Breeds. *J. Dairy Sci.* 97, 3128–3141. doi:10.3168/JDS.2013-6826

Lee, S. H., Choi, B. H., Lim, D., Gondro, C., Cho, Y. M., Dang, C. G., et al. (2013). Genome-Wide Association Study Identifies Major Loci for Carcass Weight on BTA14 in Hanwoo (Korean Cattle). *PLoS ONE* 8, e74677. doi:10.1371/journal.pone.0074677

Lozada-Soto, E. A., Maltecca, C., Lu, D., Miller, S., Cole, J. B., and Tiezzi, F. (2021). Trends in Genetic Diversity and the Effect of Inbreeding in American Angus Cattle under Genomic Selection. *Genet. Sel. Evol.* 53, 1–15. doi:10.1186/S12711-021-00644-Z/FIGURES/6

Malécot, G. (1948). *Les mathématiques de l'hérédité*. Paris: Masson.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., et al. (2015). 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

Marras, G., Wood, B. J., Makanjuola, B., Malchiodi, F., Peeters, K., Van As, P., et al. (2018). "Characterization of Runs of Homozygosity and Heterozygosity-Rich Regions in a Commercial turkey (*Meleagris gallopavo*) Population," in *Proceedings of the 11th World Congress of Genetics Applied to Livestock Production* (Auckland), 763–767.

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., and Portolano, B. (2018). Runs of Homozygosity Reveal Genome-wide Autozygosity in Italian Sheep Breeds. *Anim. Genet.* 49, 71–81. doi:10.1111/AGE.12634

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole Genome Linkage Disequilibrium Maps in Cattle. *BMC Genet.* 8, 74. doi:10.1186/1471-2156-8-74

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

Meuwissen, T. H. (2009). Accuracy of Breeding Values of 'unrelated' Individuals Predicted by Dense SNP Genotyping. *Genet. Sel. Evol.* 41, 35. doi:10.1186/1297-9686-41-35

Meuwissen, T., and Luo, Z. (1992). Computing Inbreeding Coefficients in Large Populations. *Genet. Sel. Evol.* 24, 305–313. doi:10.1186/1297-9686-24-4-305

Missio, R. L., Restle, J., Moletta, J. L., Kuss, F., Neiva, J. N. M., Elejalde, D. A. G., et al. (2015). Slaughter Weights on Animal Performance, Carcass Commercial Cuts and Meat Characteristics of Cull Cows. *Sem. Ci. Agr.* 36, 3827. doi:10.5433/1679-0359.2015v36n6p3827

Moravčíková, N., Trakovická, A., Kadlečík, O., and Kasarda, R. (2019). Genomic Signatures of Selection in Cattle through Variation of Allele Frequencies and Linkage Disequilibrium. *J. Central Eur. Agric.* 20, 576–580. doi:10.5513/JCEA01/20.2.2552

Moura, I. C. F., Kuss, F., Moletta, J. L., Menezes, L. F. G., Henrique, D. S., Lipinski, L. C., et al. (2014). Desempenho de vacas de corte Purunã submetidas a diferentes manejos de amamentação. *Pesq. Agropec. Bras.* 49, 49–56. doi:10.1590/S0100-204X2014000100007

Mulim, H. A., Brito, L. F., Pinto, L. F. B., Ferraz, J. B. S., Grigoletto, L., Silva, M. R., et al. (2022). Characterization of Runs of Homozygosity, Heterozygosity-Enriched Regions, and Population Structure in Cattle Populations Selected for Different Breeding Goals. *BMC Genomics* 23, 209–218. doi:10.1186/s12864-022-08384-0

Otto, P. I., Santos, A. L., Perotto, D., Oliveira, S. N., Granzotto, F., Gobo, D. O. R., et al. (2021). Estimation of Genetic Parameters for Weaning and Yearling Weights in a Composite Population Used to Form the Purunã Breed. *Rev. Bras. Zootec.* 50, 1–11. doi:10.37496/RBZ5020180224

Peripolli, E., Metzger, J., de Lemos, M. V. A., Stafuzza, N. B., Kluska, S., Olivieri, B. F., et al. (2018). Autozygosity Islands and ROH Patterns in Nellore Lineages: Evidence of Selection for Functionally Important Traits. *BMC Genomics* 19, 680. doi:10.1186/s12864-018-5060-8

Peripolli, E., Stafuzza, N. B., Amorim, S. T., Lemos, M. V. A., Grigoletto, L., Kluska, S., et al. (2020). Genome-wide Scan for Runs of Homozygosity in the Composite Montana Tropical Beef Cattle. *J. Anim. Breed. Genet.* 137, 155–165. doi:10.1111/jbg.12428

Perotto, D., Cubas, A. C., Moletta, J. L., and Lesskiu, C. (2000a). Heterose sobre os pesos de bovinos Canchim e Aberdeen Angus e de seus cruzamentos recíprocos. *Pesq. Agropec. Bras.* 35, 2511–2520. doi:10.1590/s0100-204x2000001200022

Perotto, D., Moletta, J. L., and Cubas, A. C. (2000b). Características quantitativas da carcaça de bovinos Charolês, Caracu e cruzamentos recíprocos terminados em confinamento. *R. Bras. Zootec.* 29, 117–124. doi:10.1590/S1516-35982000000100016

Pires, B. V., Stafuzza, N. B., de Freitas, L. A., Mercadante, M. E. Z., Ramos, E. S., and Paz, C. C. P. (2021). Expression of Candidate Genes for Residual Feed Intake in Tropically Adapted *Bos taurus* and *Bos indicus* Bulls under Thermoneutral and Heat Stress Environmental Conditions. *J. Therm. Biol.* 99, 102998. doi:10.1016/J.JTHERBIO.2021.102998

Pontes, D. S., Rosado, R. D. S., Cruz, C. D., Nascimento, M., Oliveira, A. M. C., and Pensky, S. M. (2020). Trait Selection Using Procrustes Analysis for the Study of Genetic Diversity in Conilon Coffee. *Acta Sci. Agron.* 42, e43195–12. doi:10.4025/actasciagron.v42i1.43195

Prieur, V., Clarke, S. M., Brito, L. F., McEwan, J. C., Lee, M. A., Brauning, R., et al. (2017). Estimation of Linkage Disequilibrium and Effective Population Size in New Zealand Sheep Using Three Different Methods to Create Genetic Maps. *BMC Genet.* 18, 68. doi:10.1186/s12863-017-0534-2

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

Rebelato, A. B., Caetano, A. R., Rebelato, A. B., and Caetano, A. R. (2018). Runs of Homozygosity for Autozygosity Estimation and Genomic Analysis in Production Animals. *Pesq. Agropec. Bras.* 53, 975–984. doi:10.1590/s0100-204x2018000900001

Reynolds, J., Weir, B. S., and Cockerham, C. C. (1983). Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics* 105 (3), 767–779. doi:10.1093/genetics/105.3.767

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., et al. (2020). De Novo assembly of the Cattle Reference Genome with Single-Molecule Sequencing. *Gigascience* 9, 1–9. doi:10.1093/gigascience/giaa021

Santos, W., Schettini, G., Fonseca, M. G., Pereira, G. L., Chardulo, L. A., Neto, O., et al. (2021). Fine-scale Estimation of Inbreeding Rates, Runs of Homozygosity and Genome-wide Heterozygosity Levels in the Mangalarga Marchador Horse Breed. *J. Anim. Breed. Genet.* 138, 161–173. doi:10.1111/jbg.12508

SAS Institute Inc. (2013). *SAS 9.1.3 Help and Documentation*. Cary: ADABAS.

Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE: A Web-Interfaced Next Generation Database for Genetic Diversity Exploration, with a First Application in Cattle. *BMC Genomics* 16, 1–8. doi:10.1186/s12864-015-2181-1

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman and Hall/CRC.

Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., et al. (2019). A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types. *Animals* 9, 1024. doi:10.3390/ani9121024

Taye, M., Yoon, J., Dessie, T., Cho, S., Oh, S. J., Lee, H.-K., et al. (2018). Deciphering Signature of Selection Affecting Beef Quality Traits in Angus Cattle. *Genes Genom* 40, 63–75. doi:10.1007/s13258-017-0610-z

Urbinati, I., Stafuzza, N. B., Oliveira, M. T., Chud, T. C., Higa, R. H., Regitano, L. C., et al. (2016). Selection Signatures in Canchim Beef Cattle. *J. Anim. Sci. Biotechnol.* 7, 29–9. doi:10.1186/S40104-016-0089-5/TABLES/3

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., et al. (2021). The Value of Genomic Relationship Matrices to Estimate Levels of Inbreeding. *Genet. Sel. Evol.* 53, 1–17. doi:10.1186/S12711-021-00635-0/FIGURES/7

Wang, J., Liao, Y., Jaehnig, E., Shi, Z., and Sheng, Q. (2020). *Gene Set Analysis Toolkit WebGestaltR*. CRAN.R.

Wang, J. (2014). Marker-based Estimates of Relatedness and Inbreeding Coefficients: An Assessment of Current Methods. *J. Evol. Biol.* 27, 518–530. doi:10.1111/jeb.12315

Williams, J. L., Hall, S. J. G., Del Corvo, M., Ballingall, K. T., Colli, L., Ajmone Marsan, P., et al. (2016). Inbreeding and Purging at the Genomic Level: the Chillingham Cattle Reveal Extensive, Non-random SNP Heterozygosity. *Anim. Genet.* 47, 19–27. doi:10.1111/age.12376

Wright, S. (1922). Coefficients of Inbreeding and Relationship. *Am. Nat.* 56, 330–338. doi:10.1086/279872

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing Genomic Diversity and Signatures of Selection in Jiaxian Red Cattle Using Whole-Genome Sequencing Data. *BMC Genomics* 22, 43. doi:10.1186/s12864-020-07340-0

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat. Genet.* 42, 565–569. doi:10.1038/ng.608

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, 49. doi:10.1186/s12711-015-0127-3

Conflict of Interest: Author LD is employed by the company NEOGEN Corporation.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership