EARLY DOMESTICATION AND ARTIFICIAL SELECTION OF ANIMALS

EDITED BY: Hai Xiang, Martijn Derks, Guoqiang Yi and Xingbo Zhao

PUBLISHED IN: Frontiers in Genetics and Frontiers in Veterinary Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-711-5
DOI 10.3389/978-2-88974-711-5

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

EARLY DOMESTICATION AND ARTIFICIAL SELECTION OF ANIMALS

Topic Editors:
Hai Xiang, Foshan University, China
Martijn Derks, Wageningen University and Research, Netherlands
Guoqiang Yi, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China
Xingbo Zhao, China Agricultural University, China

Citation: Xiang, H., Derks, M., Yi, G., Zhao, X., eds. (2022). Early Domestication and Artificial Selection of Animals. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-711-5

Table of Contents

04 ***Editorial: Early Domestication and Artificial Selection of Animals***
Hai Xiang, Martijn F. L. Derks, Guoqiang Yi and Xingbo Zhao

06 ***Genomic Analysis Reveals Human-Mediated Introgression From European Commercial Pigs to Henan Indigenous Pigs***
Kejun Wang, Lige Zhang, Dongdong Duan, Ruimin Qiao, Xiuling Li, Xinjian Li and Xuelei Han

16 ***Genetic Diversity of MHC B-F/B-L Region in 21 Chicken Populations***
Yiming Yuan, Huanmin Zhang, Guoqiang Yi, Zhen You, Chunfang Zhao, Haixu Yuan, Kejun Wang, Junying Li, Ning Yang and Ling Lian

26 ***Genomic Analyses Revealed the Genetic Difference and Potential Selection Genes of Growth Traits in Two Duroc Lines***
Desen Li, Min Huang, Zhanwei Zhuang, Rongrong Ding, Ting Gu, Linjun Hong, Enqin Zheng, Zicong Li, Gengyuan Cai, Zhenfang Wu and Jie Yang

35 ***Domestication and Feed Restriction Programming Organ Index, Dopamine, and Hippocampal Transcriptome Profile in Chickens***
Siyu Chen, Chao Yan, Jinlong Xiao, Wen Liu, Zhiwei Li, Hao Liu, Jian Liu, Xiben Zhang, Maojun Ou, Zelin Chen, Weibo Li and Xingbo Zhao

46 ***Genetic Variations and Differential DNA Methylation to Face Contrasted Climates in Small Ruminants: An Analysis on Traditionally-Managed Sheep and Goats***
Laure Denoyelle, Pierre de Villemereuil, Frédéric Boyer, Meidhi Khelifi, Clément Gaffet, Florian Alberto, Badr Benjelloun and François Pompanon

59 ***Ancient Mitogenomes Provide New Insights into the Origin and Early Introduction of Chinese Domestic Donkeys***
Linying Wang, Guilian Sheng, Michaela Preick, Songmei Hu, Tao Deng, Ulrike H. Taron, Axel Barlow, Jiaming Hu, Bo Xiao, Guojiang Sun, Shiwen Song, Xindong Hou, Xulong Lai, Michael Hofreiter and Junxia Yuan

67 ***Trait-specific Selection Signature Detection Reveals Novel Loci of Meat Quality in Large White Pigs***
Yu Shen, Haiyan Wang, Jiahao Xie, Zixuan Wang and Yunlong Ma

75 ***Ancient Mitogenomes Reveal the Domestication and Distribution of Cattle During the Longshan Culture Period in North China***
Xing Zhang, Liu Yang, Lingyun Hou, Hua Li, Hai Xiang and Xingbo Zhao

85 ***Review: Balancing Selection for Deleterious Alleles in Livestock***
Martijn F. L. Derks and Marije Steensma

frontiers in Genetics

EDITORIAL
published: 25 February 2022
doi: 10.3389/fgene.2022.841252

Editorial: Early Domestication and Artificial Selection of Animals

Hai Xiang[1]*, *Martijn F. L. Derks*[2], *Guoqiang Yi*[3] and *Xingbo Zhao*[4]

[1]Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, School of Life Science and Engineering, Foshan University, Foshan, China, [2]Animal Breeding and Genomics, Wageningen University and Research, Wageningen, Netherlands, [3]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [4]National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, China Agricultural University, Beijing, China

Keywords: animal, domestication, artificial selection, breed formation, adaptive evolution

Editorial on the Research Topic

Early Domestication and Artificial Selection of Animals

The domestication of several key farm animals during the early Holocene initiated the transition from the hunter-gatherer lifestyle to food production. Early domestication started by the recognition of biological characteristics in animals that would benefit human demands. Since their initial domestication, animals have been translocated worldwide and selected intensively, which has resulted in substantial genomic adaption of various traits such as tameness, growth, reproduction, etc.

Beyond the general interest in animal domestication, numerous important genetic loci selected by humans have been detected. These researches contribute to a better understanding of early domestication and subsequent artificial selection. However, the primary selective signals of domestication have been lost by migration and introgression of domesticated and wild animals.

In this research topic we cover genetic signals associated with early domestication. Moreover, we cover several aspects of human driven selection basis of the phenotypic differentiation and environmental adaptation and the effect of artificial selection on domestic animals.

OPEN ACCESS

Edited and reviewed by:
Pablo Orozco-terWengel, Cardiff University, United Kingdom

***Correspondence:**
Hai Xiang
vamyluo@126.com

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *22 December 2021*
Accepted: *04 February 2022*
Published: *25 February 2022*

Citation:
Xiang H, Derks MFL, Yi G and Zhao X (2022) Editorial: Early Domestication and Artificial Selection of Animals. Front. Genet. 13:841252. doi: 10.3389/fgene.2022.841252

EARLY DOMESTICATION

Using paleo genomic technologies, Wang et al. and Zhang et al. retrieved ancient genetic information to unravel the spatial and temporal histories of the appearance and spread of domesticated animals. By analyzing three near-complete mitochondrial genomes of donkey specimens (2,350–300 cal. BP), Wang et al. revealed that two maternal donkey lineages had been introduced into Midwestern China at least at the opening of Silk Road (approximately the first century BC). The further analyses confirmed that the two donkey lineages experienced somewhat different past demographic expansion histories, and donkeys were supposed to have undergone at least two independent domestication events. Meanwhile, Zhang et al. obtained three near-complete mitochondrial genomes from bovine remains dating back cal. 4,000 years in North China. For the first time at the mitogenome level, phylogenetic analyses confirmed the ~4,000-year-old bovines from North China as taurine cattle, which originated from the Near East. Ancient Chinese cattle had a genetic contribution to the modern taurine cattle of South China. Furthermore, a rapid decrease in the female effective population size cal. 4,650 years ago and a steep increase cal. 1,990 years ago occurred in Chinese taurine cattle.

GENETIC DIVERSITY AND INTROGRESSION

After the early domestication, contiguous breeding had been conducted on animals, leading to distinct genetic diversity between breeds or subspecies. By refining the genetic map of chicken MHC B-F/B-L region, Yuan et al. illustrated that Chinese domestic chicken breeds showed higher genetic diversity among 21 populations, suggesting their broad-spectrum resistance to pathogens. The highest and lowest genetic diversity of the MHC B-F/B-L region occurred in the indigenous chicken breeds and highly inbred line chickens, respectively, which might result from the long-term intense selection for decades.

Along with human-mediated dispersal and translocation, genetic introgression extensively happened between subspecies of domestic animals from different origins. Moreover, crossbreeding was often conducted to breed for specific traits of interest or to improve livestock's productivity. Wang et al. identified significant human-mediated introgression from European pigs into the genome of Chinese indigenous pigs. The introgression derived from Duroc pigs was captured for Henan indigenous pigs, especially in Nanyang black pigs. The introgressive *NR6A1*, *GPD2*, and *CSRNP3* genes were suggested to contribute to pig body size and feed conversion ratio, providing evidence for the effect of introgression and selection on reshaping the pig genome.

Finally, the review by Derks and Steensma explored a range of causative variants under balancing selection including loss-of-function variation and regulatory variation in domestic animals. This paper highlights some of the negative consequences associated with strong artificial selection. The authors marked various deleterious variants under balancing selection that had been discovered in various livestock species including pig (*Sus scrofa*), cow (*Bos taurus*), sheep (*Ovis aries*), chicken (*Gallus gallus*) and the horse (*Equus caballus*). The functional consequences at the molecular, phenotypic, and population level were described and summarized to provide a resource for further study.

SIGNATURES OF (ARTIFICIAL) SELECTION

Strong artificial selection has shaped the genome of domestic species. Given that animals had been subjected to artificial selection and improvement of various phenotypes with distinct purposes, great changes in production performance and adaptability have occurred in domestic populations. Li et al. performed genomic analyses on the genetic difference and growth traits in American and Canadian Duroc lines. The results showed apparent genetic differentiation between the two Duroc lines, and showed multiple selective regions and candidate genes associated with growth traits. By making phenotypic gradient differential population pairs, Shen et al. detected a total of 427 and 307 trait-specific selection signatures for meat quality in Large White pigs through different analyses. Several overlapping genes within the trait-specific selection signatures were responsible for the phenotypes including fat metabolism and muscle development.

TRANSCRIPTOMIC AND EPIGENOMIC RESPONSE TO SELECTION

With the aim to understand the genetic basis for adaptation and plasticity of contrasted climates in small ruminants, Denoyelle et al. characterized both epigenetic and genetic variations by contrasting 22 sheep and 21 goats from both sides of a climate gradient. For both species, a series of candidate genes in relation to environmental perception, immunity, reproduction and production were detected under selection with high vs. low temperature annual variations. Two differentially methylated genes, namely *AGPTA4* in goat and *SLIT3* in sheep were identified and are both related to milk production and muscle development. This work provided an explorative example to investigate the transcriptional regulatory elements and epigenetic modifications driving adaptations and selection of domesticated animals. On the other hand, Chen et al. tested the different phenotypic plasticity between slow- and fast-growing chicken in response to feed restriction. The results showed that feed restriction profoundly affected on the brain organ index and plasma dopamine in the slow- and fast-growing chickens. And the transcriptome profile analysis suggested that feed restriction might result in issues relating to cardiovascular and neurodegenerative diseases in the fast-growing breed tested, but not in the slow-growing breed.

In summary, this Research Topic collected extensive genetic evidence related to early and late domestication and in relation to current day artificial selection.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (31961133031, 32102538 and 32002150), the Guangdong Basic and Applied Basic Research Foundation (2019B030301010, 2019A1515110453 and 2020B1515120053).

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 June 2021
doi: 10.3389/fgene.2021.705803

Genomic Analysis Reveals Human-Mediated Introgression From European Commercial Pigs to Henan Indigenous Pigs

Kejun Wang, Lige Zhang, Dongdong Duan, Ruimin Qiao, Xiuling Li, Xinjian Li and Xuelei Han**

College of Animal Sciences and Technology, Henan Agricultural University, Zhengzhou, China

Introgression of genetic features from European pigs into Chinese pigs was reported possibly contributing to improvements in productivity traits, such as feed conversion efficiency and body size. However, the genomic differences from European pigs and the potential role of introgression in Henan indigenous pigs remains unclear. In this study, we found significant introgression from European pigs into the genome of Chinese indigenous pigs, especially in Henan indigenous pigs. The introgression in Henan indigenous pigs, particularly in the Nanyang black pig, was mainly derived from Duroc pigs. Most importantly, we found that the *NR6A1*, *GPD2,* and *CSRNP3* genes were introgressed and reshaped by artificial selection, and these may have contributed to increases in pig body size and feed conversion efficiency. Our results suggest that human-mediated introgression and selection have reshaped the genome of Henan pigs and improved several of their desired traits. These findings contribute to our understanding of the history of Henan indigenous pigs and provide insights into the genetic mechanisms affecting economically important traits in pig populations.

Keywords: pig, whole-genome resequencing, introgression, selection, genetic

OPEN ACCESS

Edited by:
Hai Xiang,
Foshan University, China

Reviewed by:
Lifan Zhang,
Nanjing Agricultural University, China
Yunlong Ma,
Huazhong Agricultural University, China

****Correspondence:***
Xinjian Li
lxjlongfei@163.com
Xuelei Han
hxl014@126.com

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

Received: *06 May 2021*
Accepted: *24 May 2021*
Published: *18 June 2021*

Citation:
Wang K, Zhang L, Duan D, Qiao R, Li X, Li X and Han X (2021) Genomic Analysis Reveals Human-Mediated Introgression From European Commercial Pigs to Henan Indigenous Pigs. Front. Genet. 12:705803. doi: 10.3389/fgene.2021.705803

INTRODUCTION

Livestock animals were domesticated from wild ancestors for human use, subsequently undergoing contiguous breeding and improvement through the generation of breeds with distinct phenotypes. Crossbreeding between breeds or subspecies is often conducted during the breeding process to improve livestock's productivity or desired appearance (Bosse et al., 2014). Examples of this selective breeding includes yellow skin, penciled feathers, and spotted comb in domesticated chickens resulting from introgressions from *Gallus sonneratii*, *Gallus varius,* and *Gallus lafayettii* (Morejohn, 1968; Eriksson et al., 2008; Fallahshahroudi et al., 2019). Human-mediated introgression of alleles plays an important role in altering the genome of domesticated animals, including pigs (Bosse et al., 2014).

Divergence between Asian wild and European wild pigs can been traced to approximately half a million years ago, prior to independent domestication (∼10,000 years ago) (Giuffra et al., 2000; Larson et al., 2005; Groenen et al., 2012). Subsequently, many domesticated pig populations were formed, each with distinct phenotypes. Since then, hybridization and introgression have reshaped the modern pig genome (Giuffra et al., 2000; Goedbloed et al., 2013). According to

historical records, the Chinese pig was imported into Europe in the mid-to-late 18th century (Chen et al., 2020). This introgression was likely to have played an important role in improving the reproductive traits of the Large White (LW) pig (Chen et al., 2020). Moreover, Chinese pigs were imported into America to improve the production performance of American indigenous pigs through crossbreeding, and this accounts for about 33% of the genome on average (Fang and Andersson, 2006). Several studies based on genetic data have revealed evidence for introgression between sub-populations (Yang et al., 2009, 2014; Jiang et al., 2012; Wang et al., 2020). However, the genomic differences and functions introduced by introgression are still not well understood. Henan is a central province of China with a rich diversity of indigenous pig breeds. Three indigenous pig breeds, the Nanyang Black pig (NY), the Huainan pig (HN), and the Queshan black pig (QS), have been named based on the classification scheme from the National Commission of Animal Genetics resource book (2011). These three pig breeds exhibited strong disease-resistance, high meat yield and early maturation (Qiao et al., 2019). Low density SNP analysis revealed that there has been admixture between Henan indigenous pigs and European commercial pigs (Qiao et al., 2019). However, the genomic differences and potential role of introgression in Henan indigenous pigs is still unclear.

Here, whole genome data of Henan indigenous pigs were generated and analyzed together with data for European pigs and other Chinese indigenous pigs. Furthermore, admixture and population structures were investigated. Subsequently, introgression from European pigs into Henan indigenous pigs was determined by multiple statistical analyses, and the proportion of introgression was estimated. Most importantly, common introgressed regions and selection regions were identified, which were possibly associated with high feed conversion efficiency and larger body size.

RESULTS AND DISCUSSION

Whole-Genome Resequencing of Henan Indigenous Pigs

A set of whole-genome sequencing data comprising 102 individuals was collected, from nine Chinese indigenous pig breeds (CD), four European commercial pig breeds (EUD), one Chinese wild pig population (CW), one European wild pig population (EUW), and one Java warty pig (*Sus verrucosus*, an outgroup) (**Supplementary Table 1**). Of these, we sequenced 30 pigs from three Henan indigenous pigs (CDh), generating whole-genome data with an average ∼15x sequencing depth. Approximately 66.3 million autosomal biallelic SNPs were identified. After filtering minor allelic frequencies <0.01 and call rates >0.9, about 29.9 million autosomal biallelic SNPs were ultimately retained and used in subsequent analyses.

Inference of Population Structure

Divergence between Asian wild and European wild pigs was traced to approximately half a million years prior to subsequent independent domestication (∼10,000 years ago) (Giuffra et al., 2000; Larson et al., 2005). Evidence from principal component analysis (PCA) revealed obvious differentiation between Chinese pigs and European pigs, representing the genetic differences between these sub-species (**Figure 1A**). The first principal component (PC1) captured 17.6% of the total eigenvalue, dividing individuals into two clades (European pigs and Chinese pigs). The second principal component (PC2) captured 7.57% of the total eigenvalue, dividing Chinese pigs into Southwestern indigenous pigs (the Tibetan, TT, and Wuzhishan pigs, WZS), Chinese wild pigs, and other indigenous pigs. Interestingly, we found that a cluster of three Henan indigenous pig (CDh) breeds showed a closer relatedness to European pigs (**Figure 1A**). When rooting using the Java warty pig (*Sus verrucosus*), a phylogenetic tree based on a GTR model demonstrated that different subgroups formed obvious clades (**Figure 1B**). Interestingly, we also observed that Henan indigenous pigs (CDh) clustered with European pigs. These results imply admixture between Henan indigenous pigs (CDh) and European pigs. According to the methods found in Evanno et al. 2005, the best estimate for the number of clusters (*K*-value) was equal to 2 (**Figure 1C**). When $K = 2$ in the admixture graph, two ancestral lineages separated Chinese pigs and European pigs from the population in which admixture occurred (**Figure 1C**). Obviously, CDh and two CD pigs (HTDE and MIN) showed significant admixture with a European lineage, especially in Nanyang pigs (NY). Collectively, these findings revealed the clear signature of admixture of Chinese indigenous pigs with European pigs, especially in Henan indigenous pigs (CDh). This result was also consistent with previous evidence of European pig introgression with MIN pigs (MIN), Laiwu black pigs (LWH), Hetao pigs (HTDE), and Tibetan pigs (TT) (Chen et al., 2020; Wang et al., 2020).

Gene Flow of Duroc Pigs to Henan Indigenous Pigs

To obtain direct introgression evidence, *f4* statistics and *D-statistics* were calculated for each combination of European and Chinese indigenous pigs. As expected, we found evident gene flow of European pigs in seven Chinese indigenous pig breeds according to *f4* statistics, with the exceptions of WZS and EHL pigs (**Figure 2A**, **Supplementary Figure 1**). The order of introgression intensity from high to low was NY > QS > HN > MIN > HTDE > LWH > TT. Evidence from *D-statistics* showed the same pattern as *f4* statistics did. We also calculated *f3* statistics (CD; European pig, CW) to validate the admixture of Chinese indigenous pigs from European pigs and Chinese wild pigs. We only observed significant admixture for NY pigs, whereas HN and QS pigs were close to the threshold value (**Figure 2**, **Supplementary Figure 1**). This implied that *f3* statistics (CD, CW, European pigs) were insensitive to introgression at low proportions. To identify which breed of European pig was responsible for the introgression detected, the outgroup-*f3* statistics were explored. The Duroc (DU) pig was more genetically similar to the NY, QS, and HN pigs than to European pigs, while LW pigs were closer to other Chinese indigenous pigs. This result might be explained by historical evidence indicating that Chinese pigs were introduced into European breeds to improve the fertility and immunity of LW

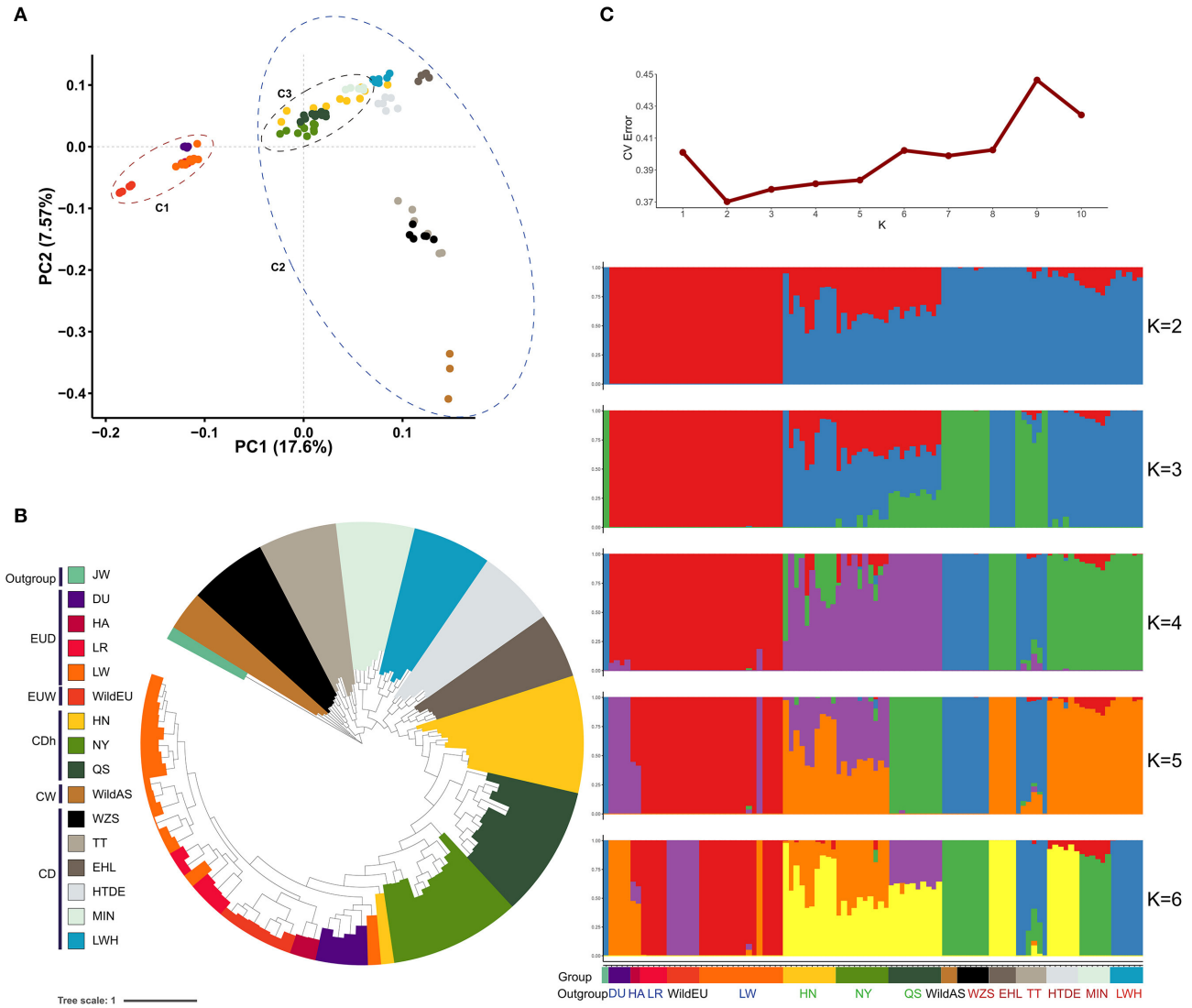

FIGURE 1 | Population structure and relationships. **(A)** Principal component analysis of Chinese and European pigs. C1, European pigs; C2, Chinese pigs; C3, Henan indigenous pigs. **(B)** Phylogenetic tree of Chinese and European pigs based on the GTR model. Group colors used in **(A)** are the same as **(B)**. **(C)** Population structure of Chinese and European pigs revealed by admixture analysis. JW, Java warty pig; DU, Duroc pig; HA, Hampshire pig; LR, Landrace pig; LW, Large White pig; WildEU, European Wild pig; HN, Huainan pig; NY, Nanyang pig; QS, Queshan pig; WildAS, Chinese wild pig; WZS, Wuzhishan pig; TT, Tibetan pig; EHL, Erhualian pig; HTDE, Hetao daer pig; MIN, Min pig; and LWH, Laiwuhei pig. All pigs were also classified as Chinese indigenous pig breeds (CD), European commercial pig breeds (EUD), Chinese wild pig populations (CW), European wild pigs (EUW), or Java warty pigs (*Sus verrucosus*, the outgroup).

pigs (Chen et al., 2020). Alternatively, introgression of LW into some Chinese indigenous pigs could not be ruled out.

To obtain more detailed introgression evidence, TreeMix analysis was conducted. The TreeMix analysis presented the same results as above, i.e., that admixture and division between European and Chinese pigs had occurred (**Figure 2B**). Increasing the number of migration events marginally enhanced the fitness of the model testing. When $m = 7$, the model explained 0.9936 of the variance for the autosomal SNP data, while when $m =$ 8 or 9, the model explained 0.9957 or 0.9963 of the variance, respectively. Evidence from TreeMix also indicated significant gene flow from DU into CDh (HN, QS, and NY), consistent with the results of our outgroup-*f3* analysis. We also observed weak introgression from DU to MIN, HTDE, TT, and LWH, which was consistent with the previous introgression results based on SNP chip data (Wang et al., 2020).

Identification of Introgression and Selective Regions in Henan Indigenous Pigs

To locate the detailed genomic regions for the gene flow from DU to CDh, we calculated the f-statistic (*f_d*) value for each combination of DU and CDh (HN, QS, and NY)

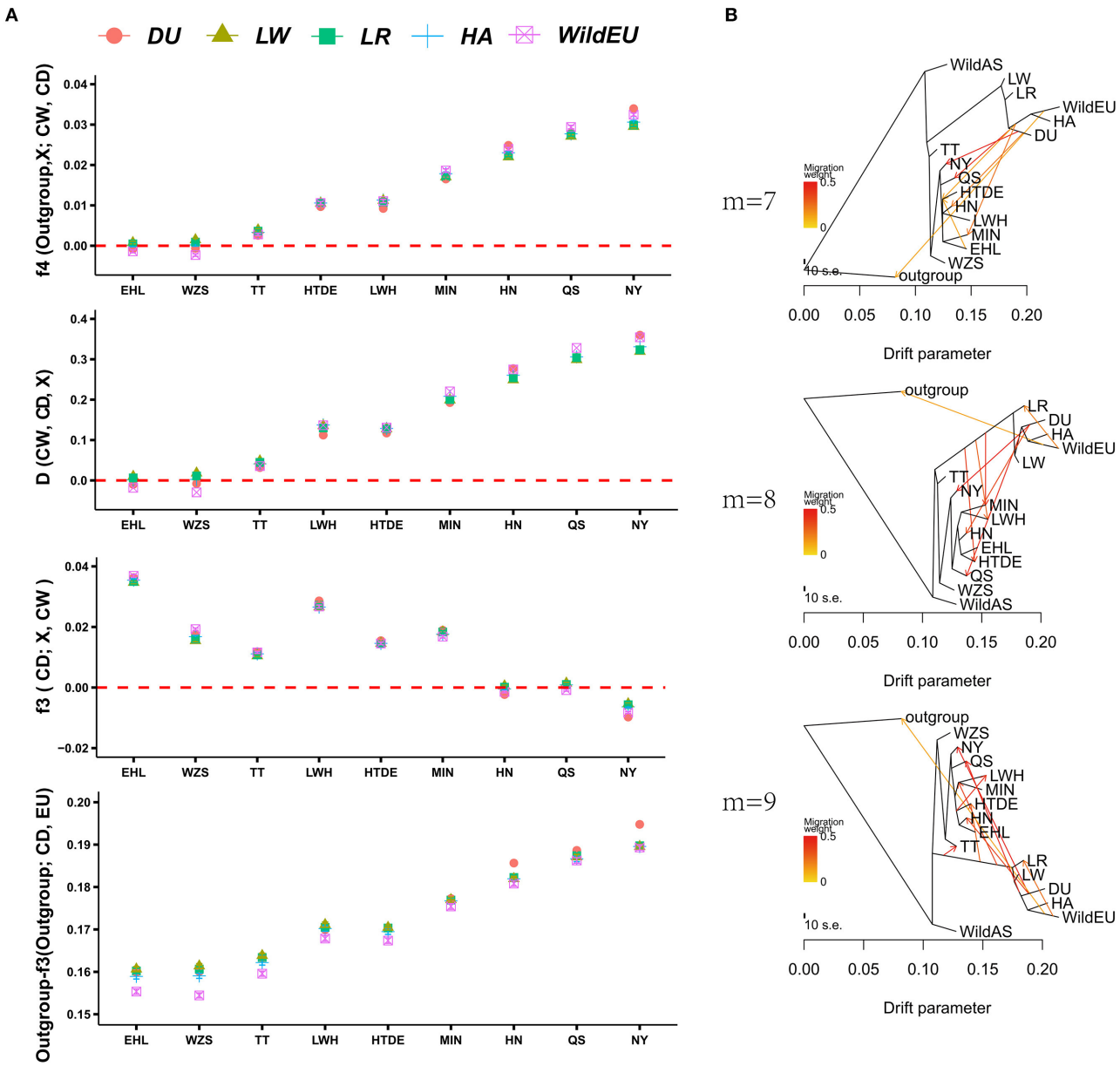

FIGURE 2 | Admixture of Chinese indigenous pigs and European pigs. **(A)** Evidence from *f4* statistics, *D-statistics*, *f3* statistics, and *outgroup-f3* statistics showing admixture of Chinese and European pigs. **(B)** TreeMix analysis revealing gene flow from European to Chinese pigs.

by the sliding-window method. After Z-transformation, the top 5% of the windows were considered to be outlier introgression regions (**Figure 3**, **Supplementary Table 2**). Totals of 506, 487, and 522 regions passed this threshold (**Supplementary Table 2**). Combining the overlapping regions, these identified introgression regions accounted for 12.53% of NY, 10.61% of HN, and 9.43% of the QS genome. The overlapping regions of CDh were located on three chromosomes, comprised of chr1:262,974,058-267,766,091, chr9:123,030,904-123,611,236, and chr15:61,503,802-74,093,823. Genes located in these overlapped regions were retrieved from the Ensemble database and are shown in **Supplementary Table 3**. Of these, the *NR6A1*, *CSRNP3*, and *GPD2* genes were considered to be putative important introgression genes (**Figure 3**).

To assess the importance of domestication and post-domestication selection on Henan indigenous pigs in the face of gene flow, we conducted a whole-genome scan for selective signatures via computing the XP-CLR of each CDh relative to Chinese wild boars (**Supplementary Figure 2**). A total of 342 selective regions were commonly identified in these three

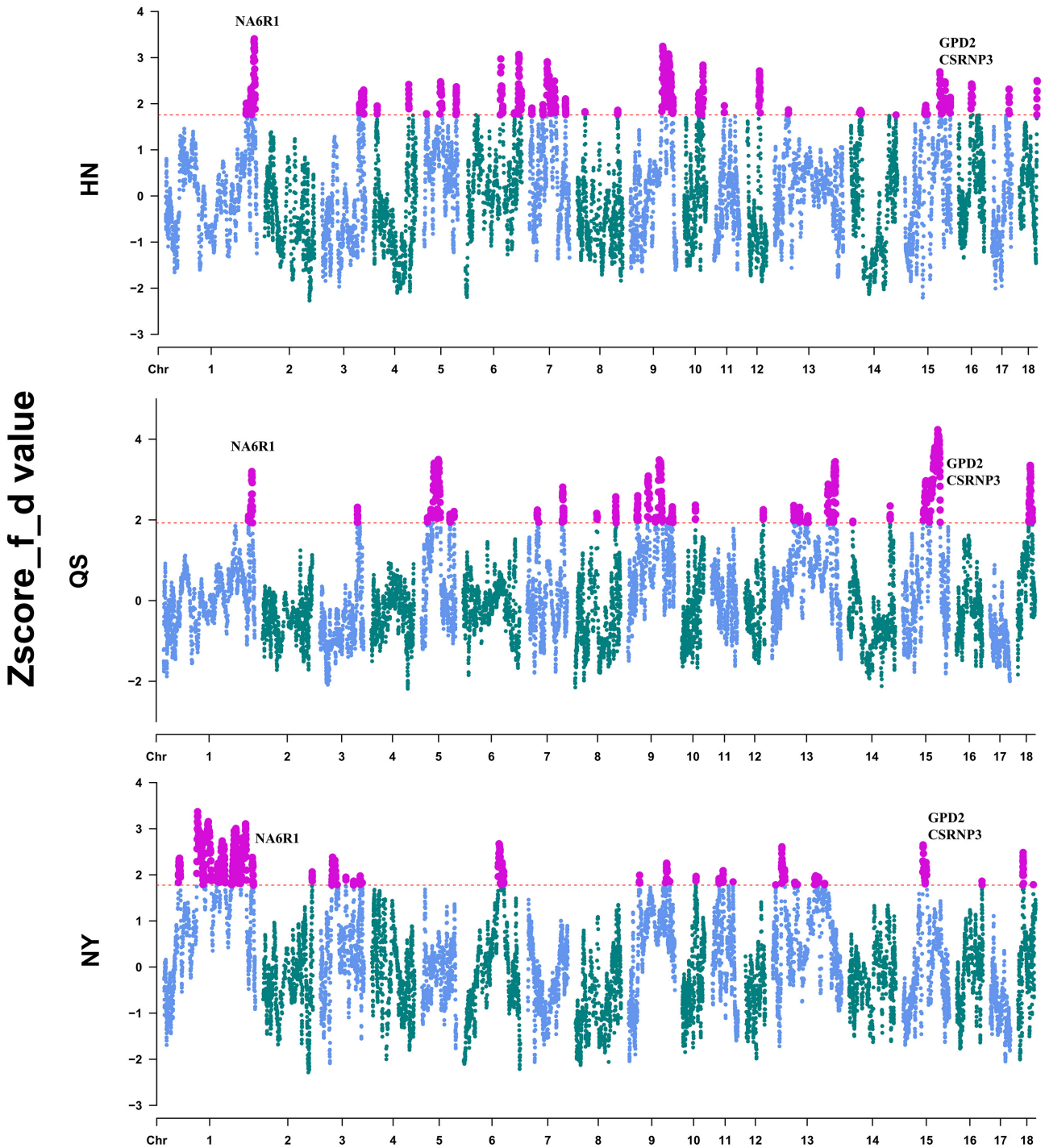

FIGURE 3 | Identification of the introgression region of Duroc pigs to Henan indigenous pigs. The red dashed line shows the 5% cutoff used to define the introgression region.

CDh (**Supplementary Table 4**). Genes located in the selective regions are shown in **Supplementary Table 5**. Two of three introgression regions were identified as being under selection and were divided into seven small genome regions: (chr1: 264,350,001–264,500,000, chr1: 265,300,001–265,600,000, chr1: 266,150,001–266,250,000, chr15: 63,100,001–63,250,000, chr15: 63,700,001–63,800,000, chr15: 64,600,001–64,700,000, and chr15: 72,100,001–72,200,000). A total of 10 coding genes and

2 non-coding genes were included, in these regions, consisting of *NR6A1*, *NR5A1*, *MAPKAP1*, *OLFML2A*, *DENND1A*, *CSRNP3*, *GPD2*, *ENSSSCG00000042870*, *ENSSSCG00000047320, ENSSSCG00000043788*, *ssc-miR-181a-2,* and *ssc-miR-181b-2*.

Introgression and Selection Reshaped the *NR6A1* Locus of Henan Indigenous Pigs

The *NR6A1* gene, nuclear receptor subfamily six group A member 1, encodes an orphan nuclear receptor. It has been widely reported to have been under selection during domestication and was shown to be associated with the number of vertebrae (Ribani et al., 2019; Zhang et al., 2019, 2020). A variant allele of p.Pro192Leu of *NR6A1* was correlated with an increase in the number of pig thoracic and dorsal vertebrae, from an average of 19 to 21–23 (Mikawa et al., 2007). An increased number of vertebrae plays a positive role in increasing body length and overall meat production (Burgos et al., 2015). Variation in the *NR6A1* gene is under selection and nearly fixed in European commercial pigs relative to European wild boars (Rubin et al., 2012). This implies the positive effect of the *NR6A1* gene was associated with this human desired production trait. In this study, we found that a region encompassing the *NR6A1* gene was significantly introgressed in Henan indigenous pigs from Duroc pigs based on *f_d* values (**Figure 3**). Moreover, we observed a remarkably high XP-CLR value in CDh pigs relative to CD pigs (**Figure 4**). Results from genotyping of European pigs and Chinese pigs for SNPs in the *NR6A1* region also showed that CDh pigs showed significant differences from Chinese pigs and greater similarity to EUD pigs. This region was reported to be fixed in European commercial pigs relative to European wild pigs and Chinese pigs represented by Meishan pigs (Rubin et al., 2012). Additionally, the *NR6A1* locus in European pigs was also reported to have introgressed into Chinese Licha pigs and has been associated with increasing vertebral number (Yang et al., 2009). Therefore, it can be inferred that this region was first introgressed into Henan indigenous pigs from European commercial pigs and was then subjected to conscious or unconscious artificial selection, since the *NR6A1* gene functions in increasing body length and meat production that are human-desired traits.

Introgression and Selection Reshaped the *GPD2* and *CSRNP3* Loci of Henan Indigenous Pigs

European commercial pigs demonstrated significantly better production performance than Chinese indigenous pigs in traits such as higher feed conversion efficiency and larger body size. In this study, we found that the loci *GPD2* and *CSRNP3* were introgressed from Duroc pigs into three Henan indigenous pig breeds (**Figure 3**). Moreover, these regions presented significant selection signals in Henan indigenous pig breeds relative to Chinese wild pigs (**Figure 5**). *GPD2*, encoding glycerol-3-phosphate dehydrogenase, catalyzes the unidirectional conversion of glycerol-3-phosphate to dihydroxyacetone phosphate and eventually to glycolysis (Gerbitz et al., 1996). The *GPD2* gene was shown to be significantly associated with residual feed intake traits, and this locus was significantly down-regulated in individuals with lower residual feed intake relative to those with higher residual feed intake (Lkhagvadorj et al., 2010). *CSRNP3* encodes a transcription factor, cysteine and serine rich nuclear protein 3, and the expression level of *CSRNP3* was associated with pig residual feed intake and feed conversion ratio traits (Vincent et al., 2015; Messad et al., 2019). In cattle, SNPs located in *CSRNP3* were significantly associated with chest width, hip width, or rump width (Doyle et al., 2020). We also compared these genotypes between European and Chinese pigs, and these showed obvious similarity to European commercial pigs (EUD) compared to Chinese wild pigs and other Chinese indigenous pigs (**Figure 5**). As is well-known, European commercial pigs have a higher feed conversion efficiency and a larger body size than Chinese native pigs. Functions of the *CSRNP3* and *GPD2* genes in feed conversion efficiency and body size may have driven the alleles from European commercial pigs to near fixation after introgression into Henan indigenous pigs under human subjective selection.

CONCLUSIONS

Collectively, we found significant introgression from European pigs into Chinese indigenous pigs, especially in Henan indigenous pigs. The introgression derived from Duroc pigs was captured for Henan indigenous pigs, especially in Nanyang black pigs. Most importantly, we found that the *NR6A1*, *GPD2,* and *CSRNP3* genes were introgressed and selected factors that may contribute to pig body size and feed conversion ratio traits. This suggests that human-mediated introgression and selection could have reshaped the pig genome and improved several desired traits. These findings further deepen our understanding of the history of Henan indigenous pigs and provide insights into the genetic mechanisms affecting economically important pig traits.

MATERIALS AND METHODS

Ethics Statement

Ear tissues of three indigenous pigs from Henan province were collected in strict accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Henan Agricultural University.

Sample Collection

Genome resequencing data for 102 pigs was collected in this study, of which 30 individuals were derived from three Henan indigenous breeds and 10 Yorkshire pigs were generated in our previous study (Li et al., 2020). We downloaded 68 accessions chosen from recent studies from the National Center for Biotechnology Information (NCBI) Sequence Read Archive database, which included from 13 breeds (**Supplementary Table 1**). Multiple statistics were used to improve the accuracy of the results in this study, but we could not rule out sampling error caused by differences in sample size.

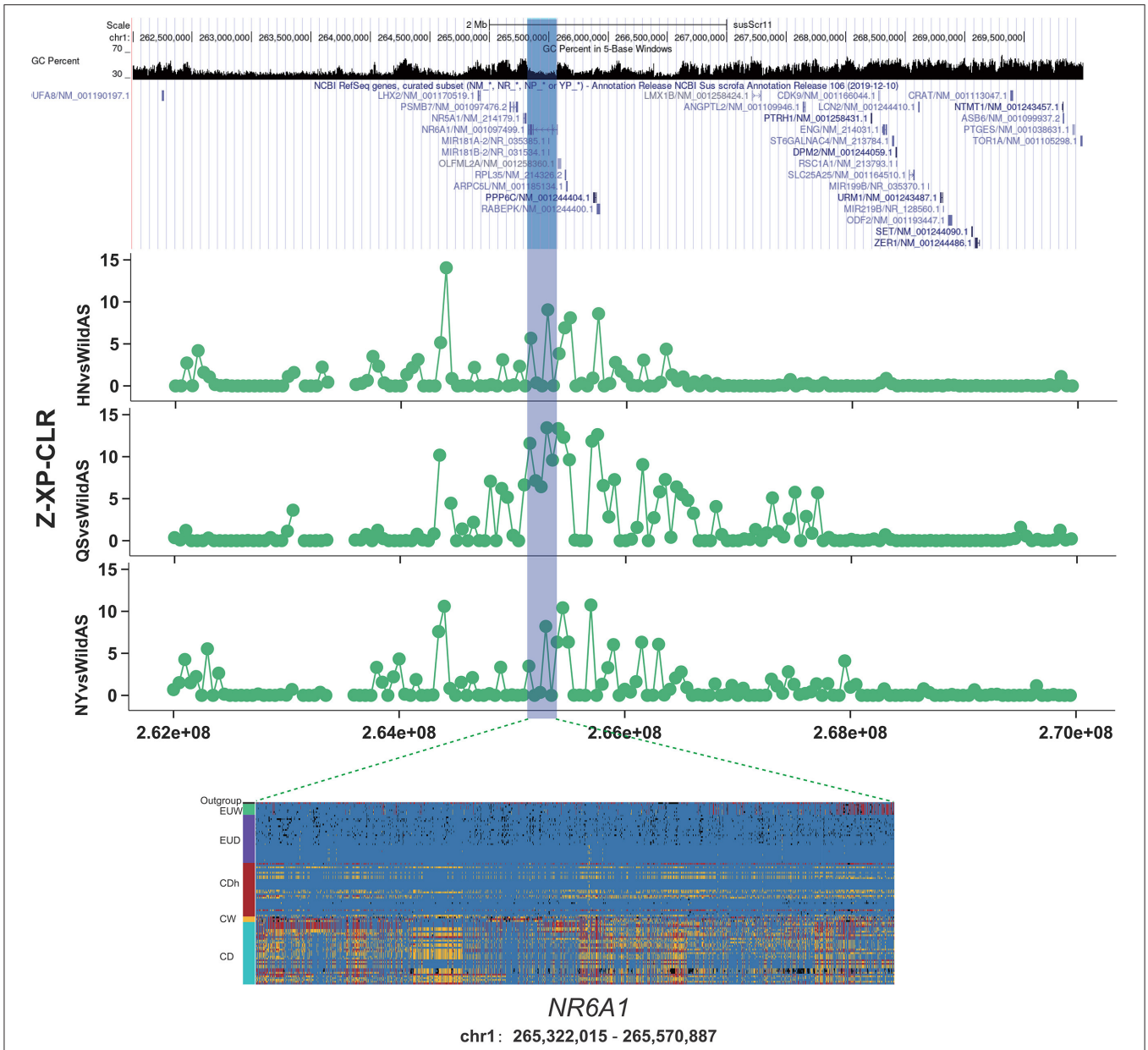

FIGURE 4 | Introgression at the *NR6A1* locus. XP-CLR showed a significant selection signature located in the *NR6A1* gene in Henan indigenous pigs compared to Chinese wild pigs. Results from genotyping a diverse panel of European and Chinese pigs for SNPs in the putative selective sweeps containing *NR6A1*.

Whole-Genome Resequencing and SNP Calling

The procedures for DNA extraction and library construction were described in our previous studies (Li et al., 2020; Shang et al., 2020). Paired end reads were generated using the BGISEQ-500 platform (BGI Genomics Co., Ltd.). Raw reads were filtered using the Trimmomatic software (seed mismatches: palindrome: simple clip −2:30:10) (Bolger et al., 2014). Filtered reads were aligned to a reference genome (Sscrofa v11.1) using the Burrows–Wheeler alignment (BWA) tool with default parameters (Li and Durbin, 2009). SAM files were merged and sorted using SAMtools (Li et al., 2009). Subsequently, duplicate reads were marked using the MarkDuplicates function of Picard Tools (http://broadinstitute.github.io/picard/). SNP calling for each individual was implemented using the HaplotypeCaller program of the GATK4 software and further joint calling using the GenomicsDBImport and GenotypeGVCFs programs of GATK4 (v4.1.2.0) (McKenna et al., 2010). Finally, we filtered our SNP panel using the Variant Filtration program of GATK4 with the parameters "*QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0.*"

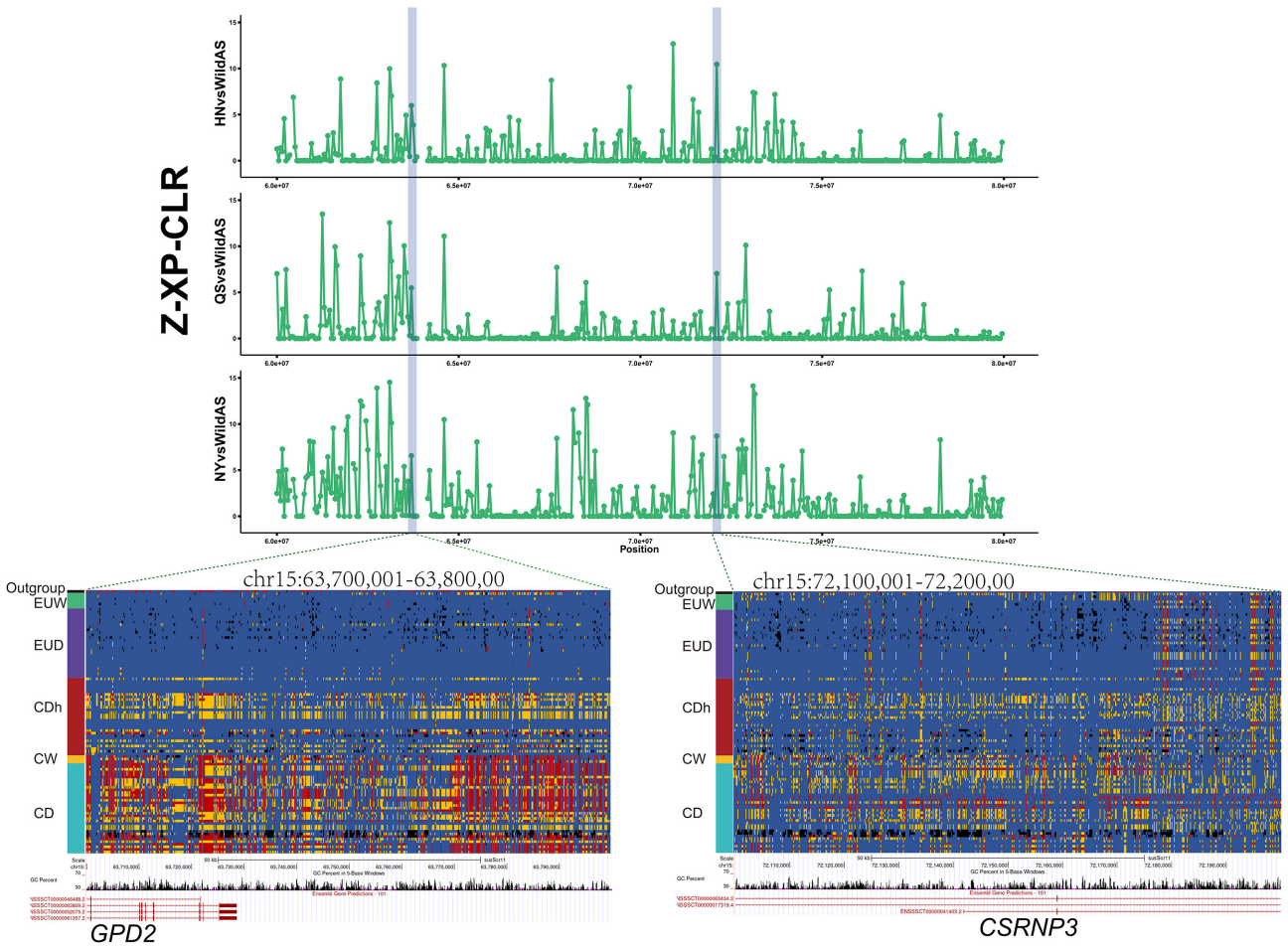

FIGURE 5 | Introgression at the *GPD2* and *CSRNP3* loci. XP-CLR showed significant selection signatures located in the *GPD2* and *CSRNP3* genes in Henan indigenous pigs compared to Chinese wild pigs. Results from genotyping a diverse panel of European and Chinese pigs for SNPs in the putative selective sweeps containing *GPD2* or *CSRNP3*.

Population Structure Analysis

Biallelic SNPs located on autosomes were identified and retained using the VCFtools software (0.1.16) (http://vcftools.github.io/) (Danecek et al., 2011). Subsequently, variants with a call rate >90% and MAF >0.01 were identified and retained (http://www.cog-genomics.org/plink2/) using PLINK (v1.9) (Chang et al., 2015). Principal Component Analysis (PCA) was also performed using PLINK. Scatter distributions for each individual were plotted using the first principal component. Population assignment analysis was conducted using the Admixture software (Kalyaanamoorthy et al., 2017), and the best number of clusters (K) was determined by the method from Evanno et al. (2005). A phylogenetic tree was constructed using the IQ-TREE software (v 1.6.12) with 1,000 bootstrap replicates based on the GTR model (Kalyaanamoorthy et al., 2017) and visualized using the iTOL online web server with a root at the outgroup (Letunic and Bork, 2019). The Java warty pig (*Sus verrucosus*) was defined as the outgroup for these analyses.

Whole-Genome Analysis of Genomic Introgression

To investigate the admixture and splitting of domesticated pigs, the separation and admixture were inferred using the TreeMix software (v. 1.13) (Pickrell and Pritchard, 2012). The *threepop* and *fourpop* programs of TreeMix were used to identify the introgression from European pigs to Chinese pigs. The *f4* (outgroup, EUD; CW, and CD) statistics were calculated to measure the shared drift between European and CD (Chinese domestic pigs) based on allelic frequency (Patterson et al., 2012). For the *f4* statistics, positive values indicated gene flow between European pigs and CD pigs under the hypothesis that there was no admixture of the outgroup into any group. The Java warty pig (*Sus verrucosus*) was defined as the outgroup in this study. The *outgroup-f3* (outgroup; CW, EUD) statistics measured the branch lengths from the outgroup to the common ancestor of Chinese domestic pigs and European pigs, and higher values indicated greater genetic similarity. Another *f3* statistic, *f3* (CD; CW, ED), was computed to identify the admixture of CD with CW and

EUD pigs. The pattern's *D-statistic* was also computed using the Dsuite software (Patterson et al., 2012). The *f_d* statistics for each window were also computed to investigate the introgression region with sliding windows containing 10,000 SNP and 1,000 SNP steps. A Z-transformation of *f_d* statistics was treated using the "scale" function in R, of which the top 5% were defined as significant introgression regions. Genes located within the introgression regions were retrieved using the Biomart program of the ENSEMBLE web server.

Signatures of Selection

Selection signatures across the whole genome were scanned in a comparison between Henan indigenous pigs and Chinese wild pigs. The XP-CLR (cross-population composite likelihood ratio) was calculated using a 100-kb sliding window with a step size of 50 kb (https://github.com/hardingnj/xpclr). Usually, a genetic distance of one cM is empirically equal to one Mb of physical distance. The regions with top 5% normalized xp-clr values were classified as significantly selected between each comparison. Genes located within the selective regions were retrieved using the Biomart program of the ENSEMBLE web server.

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://bigd.big.ac.cn, CRA004121.

AUTHOR CONTRIBUTIONS

KW, XinL, and XH participated in the design of the study and obtained project funding. KW, XinL, XH, LZ, DD, RQ, and XiuL collected the samples. KW carried out the analysis, interpretation of data, and initiated and drafted and revised the manuscript. XinL and XH revised the manuscript. All of the authors have read and approved the final manuscript.

FUNDING

This work was supported by the Pig Industry Technology System Innovation Team Project of Henan Province (S2012-06-G03), the National Natural Science Foundation of China (32002142), the Grand Science and Technology Special Project in Tibet (2021), and Henan Key Research & Development Program (Grant Nos. 192102110067, 212102110003, and 202102110242).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.705803/full#supplementary-material

Supplementary Figure 1 | Z-transformation of f4 (outgroup, X; CW, CD) and f3 (outgroup; CW, CD). X indicates one of the European pig breeds.

Supplementary Figure 2 | Z-XP-CLR values calculated for each combination between Henan indigenous pigs (CDh) and Chinese wild pigs (CW). The red dashed lines represent the 5% cutoff values used to define selective signatures.

Supplementary Table 1 | Sample information used in this study.

Supplementary Table 2 | Introgression genomic regions in Henan indigenous pigs from Duroc pigs.

Supplementary Table 3 | Genes located in the common regions of introgression from Duroc to Henan indigenous pigs.

Supplementary Table 4 | Common significantly selective regions of Henan indigenous pigs.

Supplementary Table 5 | Genes located in the common selective regions of Henan indigenous pigs.

REFERENCES

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

Bosse, M., Megens, H. J., Frantz, L. A., Madsen, O., Larson, G., Paudel, Y., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5:4392. doi: 10.1038/ncomms5392

Burgos, C., Latorre, P., Altarriba, J., Carrodeguas, J. A., Varona, L., and Lopez-Buesa, P. (2015). Allelic frequencies of NR6A1 and VRTN, two genes that affect vertebrae number in diverse pig breeds: a study of the effects of the VRTN insertion on phenotypic traits of a Duroc × Landrace–Large White cross. *Meat. Sci.* 100, 150–155. doi: 10.1016/j.meatsci.2014.09.143

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

Chen, H., Huang, M., Yang, B., Wu, Z., Deng, Z., Hou, Y., et al. (2020). Introgression of Eastern Chinese and Southern Chinese haplotypes contributes to the improvement of fertility and immunity in European modern pigs. *Gigascience* 9:giaa014. doi: 10.1093/gigascience/giaa014

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

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Walsh, S. W., Evans, R. D., et al. (2020). Genomic regions associated with skeletal type traits in beef and dairy cattle are common to regions associated with carcass traits, feed intake and calving difficulty. *Front. Genet.* 11:20. doi: 10.3389/fgene.2020.00020

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Stromstedt, L., et al. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* 4:e1000010. doi: 10.1371/journal.pgen.1000010

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Fallahshahroudi, A., Sorato, E., Altimiras, J., and Jensen, P. (2019). The domestic BCO2 allele buffers low-carotenoid diets in chickens: possible fitness increase through species hybridization. *Genetics* 212, 1445–1452. doi: 10.1534/genetics.119.302258

Fang, M., and Andersson, L. (2006). Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proc. Biol. Sci.* 273, 1803–1810. doi: 10.1098/rspb.2006.3514

Gerbitz, K. D., Gempel, K., and Brdiczka, D. (1996). Mitochondria and diabetes. genetic, biochemical, and clinical implications of the cellular energy circuit. *Diabetes* 45, 113–126. doi: 10.2337/diabetes.45.2.113

Giuffra, E., Kijas, J. M., Amarger, V., Carlborg, O., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785–1791. doi: 10.1093/genetics/154.4.1785

Goedbloed, D. J., Megens, H. J., Van Hooft, P., Herrero-Medrano, J. M., Lutz, W., Alexandri, P., et al. (2013). Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol. Ecol.* 22, 856–866. doi: 10.1111/j.1365-294X.2012.05670.x

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622

Jiang, Y. Z., Zhu, L., Tang, G. Q., Li, M. Z., Jiang, A. A., Cen, W. M., et al. (2012). Carcass and meat quality traits of four commercial pig crossbreeds in China. *Genet. Mol. Res.* 11, 4447–4455. doi: 10.4238/2012.September.19.6

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). Model finder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Sci.* 307, 1618–1621. doi: 10.1126/science.1106927

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

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, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAM tools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, X., Ye, J., Han, X., Qiao, R., Li, X., Lv, G., et al. (2020). Whole-genome sequencing identifies potential candidate genes for reproductive traits in pigs. *Genomics* 112, 199–206. doi: 10.1016/j.ygeno.2019.01.014

Lkhagvadorj, S., Qu, L., Cai, W., Couture, O. P., Barb, C. R., Hausman, G. J., et al. (2010). Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 298, R494–R507. doi: 10.1152/ajpregu.00632.2009

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Messad, F., Louveau, I., Koffi, B., Gilbert, H., and Gondret, F. (2019). Investigation of muscle transcriptomes using gradient boosting machine learning identifies molecular predictors of feed efficiency in growing pigs. *BMC Genomics* 20:659. doi: 10.1186/s12864-019-6010-9

Mikawa, S., Morozumi, T., Shimanuki, S., Hayashi, T., Uenishi, H., Domukai, M., et al. (2007). Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Res.* 17, 586–593. doi: 10.1101/gr.6085507

Morejohn, G. V. (1968). Study of plumage of the four species of the genus gallus. *The Condor* 70, 56–65. doi: 10.2307/1366508

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi: 10.1534/genetics.112.145037

Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967

Qiao, R., Li, X., Han, X., Wang, K., Lv, G., Ren, G., et al. (2019). Population structure and genetic diversity of four Henan pig populations. *Anim. Genet.* 50, 262–265. doi: 10.1111/age.12775

Ribani, A., Utzeri, V. J., Geraci, C., Tinarelli, S., Djan, M., Velickovic, N., et al. (2019). Signatures of de-domestication in autochthonous pig breeds and of domestication in wild boar populations from MC1R and NR6A1 allele distribution. *Anim. Genet.* 50, 166–171. doi: 10.1111/age.12771

Rubin, C. J., Megens, H. J., Martinez Barrio, A., 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

Shang, P., Li, W., Tan, Z., Zhang, J., Dong, S., Wang, K., et al. (2020). Population genetic analysis of ten geographically isolated tibetan pig populations. *Animals* 10:1297. doi: 10.3390/ani10081297

Vincent, A., Louveau, I., Gondret, F., Trefeu, C., Gilbert, H., and Lefaucheur, L. (2015). Divergent selection for residual feed intake affects the transcriptomic and proteomic profiles of pig skeletal muscle. *J. Anim. Sci.* 93, 2745–2758. doi: 10.2527/jas.2015-8928

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* 11:275. doi: 10.3390/genes11030275

Yang, G., Ren, J., Zhang, Z., and Huang, L. (2009). Genetic evidence for the introgression of Western NR6A1 haplotype into Chinese Licha breed associated with increased vertebral number. *Anim. Genet.* 40, 247–250. doi: 10.1111/j.1365-2052.2008.01820.x

Yang, S., Li, X., Li, K., Fan, B., and Tang, Z. (2014). A genome-wide scan for signatures of selection in Chinese indigenous and commercial pig breeds. *BMC Genet.* 15:7. doi: 10.1186/1471-2156-15-7

Zhang, S., Zhang, K., Peng, X., Zhan, H., Lu, J., Xie, S., et al. (2020). Selective sweep analysis reveals extensive parallel selection traits between large white and Duroc pigs. *Evol. Appl.* 13, 2807–2820. doi: 10.1111/eva.13085

Zhang, X., Li, C., Li, X., Liu, Z., Ni, W., Cao, Y., et al. (2019). Association analysis of polymorphism in the NR6A1 gene with the lumbar vertebrae number traits in sheep. *Genes Genomics* 41, 1165–1171. doi: 10.1007/s13258-019-00843-5

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fgene.2021.710770

Genetic Diversity of MHC B-F/B-L Region in 21 Chicken Populations

*Yiming Yuan[1], Huanmin Zhang[2], Guoqiang Yi[3], Zhen You[1], Chunfang Zhao[1], Haixu Yuan[1], Kejun Wang[4], Junying Li[1], Ning Yang[1] and Ling Lian[1]**

[1] National Engineering Laboratory for Animal Breeding and MOA Key Laboratory of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, [2] United States Department of Agriculture, Agricultural Research Service, Avian Disease and Oncology Laboratory, East Lansing, MI, United States, [3] Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [4] College of Animal Science and Technology, Henan Agricultural University, Zhengzhou, China

The chicken major histocompatibility complex (MHC) on chromosome 16 is the most polymorphic region across the whole genome, and also an ideal model for genetic diversity investigation. The MHC B-F/B-L region is 92 kb in length with high GC content consisting of 18 genes and one pseudogene (Blec4), which plays important roles in immune response. To evaluate polymorphism of the Chinese indigenous chickens as well as to analyze the effect of selection to genetic diversity, we used WaferGen platform to identify sequence variants of the B-F/B-L region in 21 chicken populations, including the Red Jungle Fowl (RJF), Cornish (CS), White Leghorns (WLs), 16 Chinese domestic breeds, and two well-known inbred lines 6_3 and 7_2. A total of 3,319 single nucleotide polymorphism (SNPs) and 181 INDELs in the B-F/B-L region were identified among 21 populations, of which 2,057 SNPs (62%) and 159 INDELs (88%) were novel. Most of the variants were within the intron and the flanking regions. The average variation density was 36 SNPs and 2 INDELs per kb, indicating dramatical high diversity of this region. Furthermore, *BF2* was identified as the hypervariable genes with 67 SNPs per kb. Chinese domestic populations showed higher diversity than the WLs and CS. The indigenous breeds, Nandan Yao (NY), Xishuangbanna Game (XG), Gushi (GS), and Xiayan (XY) chickens, were the top four with the highest density of SNPs and INDELs. The highly inbred lines 6_3 and 7_2 have the lowest diversity, which might be resulted from a long-term intense selection for decades. Collectively, we refined the genetic map of chicken MHC B-F/B-L region, and illustrated genetic diversity of 21 chicken populations. Abundant genetic variants were identified, which not only strikingly expanded the current Ensembl SNP database, but also provided comprehensive data for researchers to further investigate association between variants in MHC and immune traits.

Keywords: chicken, MHC, target enrichment sequencing, B-F/B-L region, genetic diversity

OPEN ACCESS

Edited by:
Sankar Subramanian, University of the Sunshine Coast, Australia

Reviewed by:
Long Zhang, China West Normal University, China
Shi-Yi Chen, Sichuan Agricultural University, China

***Correspondence:**
Ling Lian
lianlinglara@126.com

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *17 May 2021*
Accepted: *26 July 2021*
Published: *13 August 2021*

Citation:
Yuan Y, Zhang H, Yi G, You Z, Zhao C, Yuan H, Wang K, Li J, Yang N and Lian L (2021) Genetic Diversity of MHC B-F/B-L Region in 21 Chicken Populations. Front. Genet. 12:710770. doi: 10.3389/fgene.2021.710770

INTRODUCTION

China has the most abundant resources of domestic chickens, which possess highly genetic polymorphisms. As a highly polymorphic region across the genome, major histocompatibility complex (MHC) is a group of tightly linked genes, encoding major histocompatibility antigens, which discriminate between self and non-self cells and regulate immune responses. MHC is closely

related to disease resistance and production performance of animals (Bacon, 1987). Chicken MHC-B and MHC-Y regions reside on the micro-chromosome 16, which are located on the same side of nucleolar organizing region (NOR) (Miller and Taylor, 2016). MHC-B and MHC-Y are inherited independently because they are separated by a GC rich region (Miller et al., 2004). Until now, 46 genes within a 2,41,833 bp in MHC-B were identified (Shiina et al., 2007), including three highly polymorphic gene groups: *BF*, *BL*, and *BG*, which encode class I, class II, and class IV glycoproteins on cell surface, respectively (Lamont et al., 1987). The encoded proteins play important roles in rapid allograft rejection, immune response and determining susceptibility/resistance to pathogen infections (Pazderka et al., 1975; Plachy et al., 1992; Kaufman, 2018).

The BF/BL region (Guillemot et al., 1988), a 92-kb core region with a relatively high GC content (Shiina et al., 2007), has been described as a minimal essential MHC region since it contains two classical class I (B-F), two classical class II B (B-LB) genes, and some other genes involved in antigen processing and presentation (Bacon, 1987). This region contains 18 genes and 1 pseudogene (*Blec4*) (Lamont et al., 1987; Kaufman et al., 1995), of which, two MHC class I genes *BF1* and *BF2* encode α chain of BF antigen (Ewald and Livant, 2004). *BLB1* and *BLB2* encode β chains of classic MHC class II molecules (Lamont, 1989; Zekarias et al., 2002). MHC is a polymorphic region that has numerous SNPs and INDELs (Wong et al., 2004), and its variations can represent the whole genome-level variation to evaluate genomic diversity and population history (Yuhki and Obrien, 1990). The microsatellite locus *LEI0258*, a tetranucleotide repeat system located within the BF/BL region, has been used to analyze population genetic diversity, structure, distinctiveness, and relationships between populations (McConnell et al., 1999; Fulton et al., 2006; Izadi et al., 2011; Chang et al., 2012). Huang et al. (2016) used *LEI0258* to examine the genetic diversity and evolutionary history of south China domestic chickens and revealed that south China domestic chickens mainly originated from the Red Jungle Fowl (RJF). Han et al. (2013) found high genetic diversity of the MHC region in Chinese indigenous chickens by assessing genetic information on *LEI0258*. In addition to LEI0258, there are also many microsatellite loci such as MCW0370 and MCW0371. The use of these high polymorphism markers is of great significance for the study of chicken MHC diversity (Esmailnejad et al., 2017; Manjula et al., 2021).

However, to the best of our knowledge, there is no single-base-resolution polymorphism description of MHC BF/BL region in Chinese domestic chickens. In this study, we comprehensively evaluated the polymorphisms of BF/BL region among 16 Chinese domestic populations, as well as White Leghorns (WLs), Cornish (CS), RJFs, and two high inbred lines (lines 6_3 and 7_2) (Waters, 1945; Stone, 1975; Bacon et al., 2000) using the WaferGen platform, which is a more cost-effective method for identifying variations in targeted regions (De Wilde et al., 2014).

MATERIALS AND METHODS

Ethics Statement

All experimental procedures and used animals were approved by the Animal Care and Use Committee of China Agricultural University.

Sample Selection

A total of 195 chickens from 21 populations were used in this study. Sixteen were Chinese domestic chicken populations (**Table 1**), including Dagu (DG, n = 9), Wenchang (WC, n = 10), Wuding (WD, n = 10), Xiayan (XY, n = 10), Xishuangbanna Game (XG, n = 10), Luxi Game (LX, n = 7), Gushi (GS, n = 10), Wenshang Barred chicken (WS, n = 10), Nandan Yao (NY, n = 10), Yunyang Black-bone chicken (YY, n = 10), Tibetan (TB, n = 10), Beijing You (BY, n = 10), Chahua (CH, n = 10), Dongxiang Blue-eggshell chicken (DX, n = 10), Shouguang (SG, n = 10), and Silkie (SK, n = 10). Two introduced populations were CS(n = 10) and WL(n = 10). The blood samples of first ten local breeds were collected previously by our colleagues from local chicken breed preservation farm. The blood samples of latter six local breeds as well as CS and WL were from Experimental Farm of Poultry Resource in China Agricultural University. The individuals were randomly selected for the assay. RJF (n = 3) was also included in this study, and DNA was kindly provided by Dr. Xiaoxiang Hu. Two highly inbred WL lines 6_3 (n = 8) and 7_2 (n = 8) were that were cultivated by the Avian Disease and Oncology Laboratory (ADOL) of the United States Department of Agriculture. After decades of breeding, the inbreeding coefficient of the two lines reached 99%. Serological tests showed that the two lines have the same MHC *B* haplotype (*B2*). Although they have the same haplotype, there were significant differences in Marek's disease (MD) resistance: line 6_3 is resistant to MD tumors but susceptible to both Marek's disease virus (MDV) and avian leukosis virus (ALV). Conversely, the line 7_2 is resistant to ALV infection but susceptible to both MDV and MD tumors (Waters, 1945; Stone, 1975; Bacon et al., 2000). DNA of line 7_2 and 6_3 were extracted from spleens.

Target Enrichment Sequencing

Genome DNA were extracted from blood or tissue samples using DNA extraction kit (TIANGEN). The purity, concentration, and volume of the extracted DNA were tested. WaferGen Smartchip platform was used to perform highly parallelized PCR for the target region and the procedure was followed that by De Wilde et al. (2014). There were total of 431 amplicons were designed using tiling settings, taking into account known SNP positions and with a target annealing temperature of 60°C to cover target region (approximately 92 kb B-F/B-L region on chromosome 16). Primer3 (Version 2.3.7) global settings for primer design, length setting of amplicons were as follows: 250–300 bp and 301–350 bp, primer length: 18–30 nt, primer pair_max_diff_Tm: 3°C, and primer TM: 58–62°C (60°C as optimal). Primers were listed in **Supplementary Table 1**. Specific barcodes were added to each sample. Parallelized PCR were cycled in the SmartChip Cycler

(WaferGen Biosystems). Library was constructed by PCR pool. The concentration of PCR pool was measured using the dsDNA assay kit on the Qubit fluorometer (Invitrogen) and fragment analysis occurred on BioAnalyzer 2100 (Agilent). The library was sequenced by MiSeq platform.

Reads Mapping and Variant Calling

Chicken reference genome assembly (Galgal6 version) was download from NCBI.[1] Trimmomatic (Trimmomatic 0.39) was used to obtain clean reads, and the parameter of quality control was as follows: SLIDINGWINDOW:5:20 LEADING:3 TRAILING:3 MINLEN:36 (Bolger et al., 2014). SAMtools (Samtools 1.9) was used to create an index of the reference sequence (Li et al., 2009). Burrows–Wheeler Alignment tool (BWA-0.7.17) was used to mapping reads to the reference genome with the command "mem -t 10 -M" (Li and Durbin, 2010). The alignment data stream was piped to SAMtools and converted to a BAM file, which was then sorted according to the reference genome. The MarkDuplicates module in Genome Analysis Toolkit (GATK-4.1.3.0) was used to remove duplicated reads (McKenna et al., 2010). GATK HaplotypeCaller module and GATK GenotypeGVCFs module were used to convert BAM files to GVCF files and then switched to VCF files. The GATK VariantFiltration module was used to hard-filter SNP and INDEL. For SNP, we excluded potential false variants following these criteria: QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0. For INDEL, we took out variants with the following parameters: QUAL < 30.0 || QD < 2.0 || FS > 200.0 || SOR > 10.0 || MQRankSum < −12.5 || ReadPosRankSum < −20.0. The variants with multiple alleles were also excluded. SNPs was identified using the criteria of filtering minor allele frequency (MAF)<0.01 and missing genotype>10% in the 21 chicken populations. Subsequently, the Variants Effect Predictor (VEP) on Ensembl was used to annotate variants.

[1]https://www.ncbi.nlm.nih.gov/

TABLE 1 | The distribution of domestic chicken breeds in China.

Breeds	Origins
Beijing You	Beijing
Chahua	Yunnan
Dagu	Liaoning
Dongxiang Blue-eggshell	Jiangxi
Gushi	Henan
Luxi Game	Shandong
Nandan Yao	Guangxi
Shouguang	Shandong
Silkie	Jiangxi
Tibetan	Tibet
Wenchang	Hainan
Wuding	Guangxi
Wenshang Barred	Shandong
Xishuangbanna Game	Yunnan
Xiayan	Guangxi
Yunyang Black-bone	Hubei

Comparison Between Chickens Showing MD Resistance and Susceptibility

Comparisons were conducted between the line 6_3 and line 7_2. Allele frequency and frequency difference of each SNP were calculated with VCFtools (versions 4.0) – freq and –diff-site (Danecek et al., 2011). In each of the lines, loci with genotypic information was retained from more than four individuals. The difference of allele frequency among two populations was subjected to chi-squared testing.

RESULTS

Sequencing and Mapping Summary

On average, ∼1,14,000 (80%) clean reads were obtained from ∼1,42,000 raw reads for each individual. Eighty-eight percentage of reads were aligned to the chicken genome assembly (galGal6) (**Table 2**), and the captured specificity reached 99%. The coverage

TABLE 2 | Summary statistics of a targeted sequencing and mapping region of chicken major histocompatibility complex (MHC).

Population name[a]	Clean reads	Mapping rate[b] (%)	Capture specificity[c] (%)	Mean depth of target region (X)
Line 6_3	1,14,206	86.85	99.21	234.4
Line 7_2	1,14,488	86.56	99.18	232.9
BY	1,14,629	86.87	99.23	243.3
CH	1,17,262	89.02	99.39	255.0
CS	1,15,211	87.23	99.24	244.2
DG	1,15,510	91.05	99.39	255.8
DX	1,14,403	88.78	99.28	245.1
GS	1,10,158	88.79	99.34	238.3
LX	1,14,702	90.94	99.43	252.1
NY	1,13,370	90.11	99.23	246.4
RJF	1,12,549	83.34	99.24	224.3
SG	1,15,861	89.56	99.24	252.3
SK	1,13,112	86.76	99.18	238.8
TB	1,15,405	87.42	99.30	247.7
WC	1,15,773	85.65	99.14	239.5
WD	1,15,933	87.71	99.32	248.1
WL	1,15,457	88.30	99.10	248.1
WS	1,12,316	87.30	98.94	237.9
XB	1,15,583	89.19	99.37	249.6
XY	1,15,110	91.97	99.41	259.0
YY	1,02,682	81.04	98.96	204.4
Average	1,13,987	87.83	99.24	242.7

[a]*Chicken: BY, Beijing You; DG, Dagu; WC, Wenchang; WD, Wuding; XY, Xiayan; XB, Xishuangbanna Game; TB, Tibetan; CH, Chahua; LX, Luxi Game; GS, Gushi, WS, Wenshang Barred chicken; DX, Dongxiang Blue-eggshell; NY, Nandan Yao; SG, Shouguang; SK, Silkie; YY, Yunyang Black-bone; CS, Cornish; WL, White Leghorn. Line63, Marek's disease resistant populations; Line72, Marek's disease susceptible populations; RJF, Red Jungle Fowl.*

[b]*The reads aligned to genome accounted for the clean reads.*

[c]*The reads aligned to target region accounted for the reads aligned to genome.*

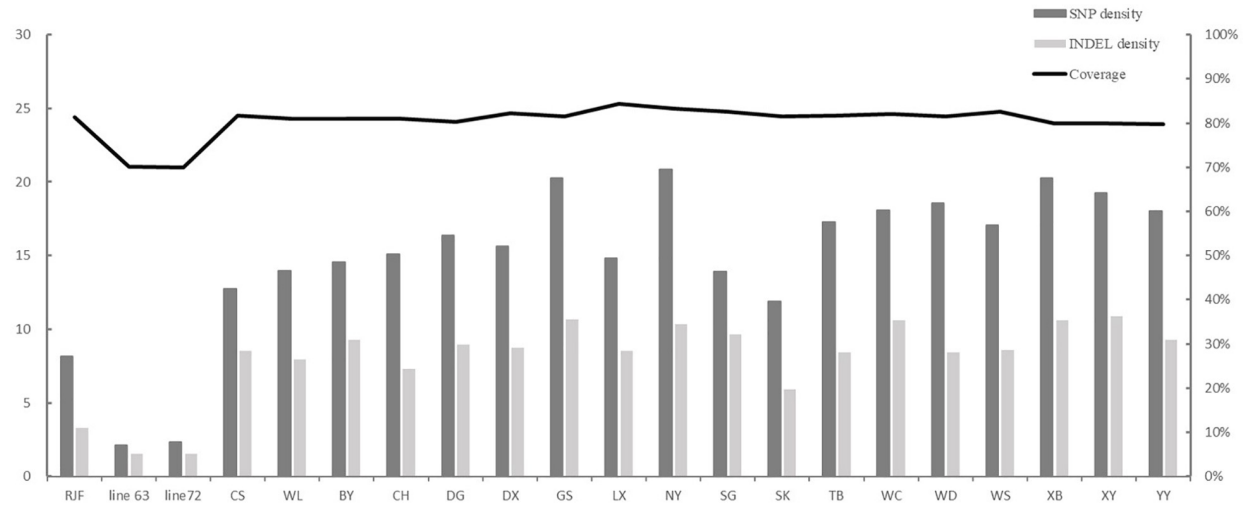

FIGURE 1 | SNP and INDEL density in each population. The variation density means that the count of SNPs per kb and the count of INDELs per 10 kb.

of target regions ranged from 70 to 84%, with an average of 80%. However, the target region coverage of the line 6_3 and line 7_2 was the lowest, about 70% (**Figure 1**). The average depth of the 92 kb target regions reached over 240×, among them, the coverage of 55 kb was more than 30×, and the coverage of 64 kb was more than 10×.

The SNP database at Ensembl (release 102) had 4,804 known variations in this region, including 4,572 SNPs and 232 INDELs. In this study, a total of 3,319 SNPs and 181 INDELs were identified in the 21 populations (**Table 3**). Among them, there were 1,262 known SNPs (38%) and 22 known INDELs (12%); and the rest 2,057 SNPs (62%) and 159 INDELs (88%) were discovered for the first time. The Chinese domestic populations GS, NY, XB, and XY had the most abundant SNPs and INDELs than others, and the highly inbred line 6_3 and 7_2 showed the fewest SNPs and INDELs among the populations. There were no breed-specific SNPs or INDELs identified in this study, and most of them were mutually shared between some populations. We also calculated the allele frequency to analyze the difference in the BF/BL region between the MD resistant line 6_3 and the MD-susceptible line 7_2. There were only 12 loci with allele frequency differences greater than 0.3, and the largest difference was 0.5, seven of them differed at significant levels ($P < 0.05$, chi-squared testing). There was no SNP locus that completely separates the two lines (**Table 4**).

Since the coverage of the different populations was different, we calculated variation density as follows: the numbers of SNPs and INDELs were divided by the coverage of each population. The population diversity was assessed as the number of SNP per kb and the number of INDEL per 10 kb. The top five populations with the highest density of SNPs were NY, XB, GS, XY, and WD, with 20.87, 20.29, 20.26, 19.28, and 18.57 per kb, respectively. The lowest density of SNP was found in RJF, inbred line 6_3 and line 7_2, with 2.13, 2.35, and 8.16, respectively. Similarly, NY, GS, XB, XY, and WC five populations showed the highest density of INDELs, and the three populations with the lowest density of INDELs were RJF, the line 6_3, and line 7_2. The SNP and INDEL densities in the two introduced populations, CS and WL, were lower than those of most of the Chinese indigenous populations (**Figure 1**).

Characteristic and Distribution of SNPs in Genes

The whole BF/BL region showed a very high genetic diversity, with an average of about 36 SNPs per kb. Among a total of 18 genes, the most diverse genes were *BF2* and *TAP2*, with more than 50 SNPs per kb. About 86% of the SNPs identified in *BF2* were newly discovered. The coding region variation was primarily concentrated in the second and third exon of *BF2*. *BF1* and *C4* followed with more than 40 SNPs per kb. *CENPA*, *BLEC2*, and two *BLB* genes had the lowest SNP rates of about 20 SNPs per kb (**Figure 2**). Among the 21 populations, RJF, line 6_3 and line 7_2 had the lowest diversity, while NY, GS, and XG were observed with the highest diversity in both BF genes (**Figure 3**). The diversity of CS and WL were lower than most of the Chinese domestic breeds. In addition, we also counted the number of variants in other genes (**Supplementary Table 2**). We then annotated all detected SNPs using VEP and found that most SNPs (54%) were located on the intron, 21% in coding region, about 7 and 4% in the downstream and upstream flanking regions, respectively, 3% in splice region and 5% were situated at UTR regions (**Figure 4A**). In the coding region, more than half of the SNP were missense variants; 43% were synonymous variants (**Figure 4B**).

Characteristic and Distribution of INDELs in Genes

The ratio of deletion (64%) was significantly higher than that of insertion (36%). Both the largest deletion and insertion detected in this study were 60 and 25 bp, respectively, and both the smallest deletion and insertion were 1 bp. A large proportion (89%) of INDELs were less than 10 bp. The majority was the single base-pair INDEL and accounted for 58% of all detected INDELs.

TABLE 3 | SNPs and INDELs detected in 21 chicken populations.

Breed	SNP count		INDEL count				Maximum length(bp)	
	Total	Novel (ratio)	Total	Insertion	Deletion	Novel (ratio)	Insertion	Deletion
RJF	618	249(40.3%)	25	7	18	19(76.0%)	7	31
Inbred population	254	175(68.9%)	14	2	12	12(85.7%)	2	60
Line 6_3	139	85(61.2%)	10	0	10	9(90.0%)	0	60
Line 7_2	153	98(64.1%)	10	2	8	9(90.0%)	2	60
Introduced breed	1,538	671(43.6%)	92	25	67	74(80.4%)	15	60
CS	971	430(44.3%)	65	20	45	51(78.5%)	11	60
WL	1,056	364(34.5%)	60	16	44	48(80.0%)	15	60
Domestic breed	2,862	1,671(58.4%)	157	42	115	137(87.3%)	25	60
BY	1,098	480(43.7%)	70	20	50	57(81.4%)	8	60
CH	1,138	487(42.8%)	55	13	42	43(78.2%)	11	60
DG	1,224	499(40.8%)	67	20	47	59(88.1%)	15	44
DX	1,197	484(40.4%)	67	22	45	55(82.1%)	25	60
GS	1,536	704(45.8%)	81	25	56	69(85.2%)	15	60
LX	1,163	486(41.8%)	67	17	50	53(79.1%)	7	60
NY	1,619	724(44.7%)	80	23	57	64(80.0%)	25	60
SG	1,072	458(42.7%)	74	24	50	64(86.5%)	8	44
SK	902	379(42.0%)	45	13	32	36(80.0%)	7	60
TB	1,314	555(42.2%)	64	23	41	50(78.1%)	15	34
WC	1,382	592(42.8%)	81	25	56	65(80.2%)	15	60
WD	1,409	595(42.2%)	64	18	46	54(84.4%)	11	60
WS	1,311	541(41.3%)	66	16	50	50(75.8%)	11	60
XB	1,510	658(43.6%)	79	22	57	65(82.3%)	15	60
XY	1,435	643(44.8%)	81	24	57	66(81.5%)	15	44
YY	1,340	585(43.7%)	69	19	50	52(75.4%)	11	44

TABLE 4 | Allele frequency differences between the line 6_3 and line 7_2.

Position	Gene	Line 63			Line 72			Difference	*P*-value
		Allele count	Ref	Alt	Allele count	Ref	Alt		
2533677	BTN1	10	G:0.8	A:0.2	8	G:0.5	A:0.5	0.3	0.1596
2537585	BTN1	12	C:1	T:0	10	C:0.6	T:0.4	0.4	0.0154
2547609	BTN2	8	A:0.5	G:0.5	10	A:1	G:0	0.5	0.0112
2547628	BTN2	8	T:0.5	C:0.5	10	T:1	C:0	0.5	0.0112
2547651	BTN2	8	T:0.5	G:0.5	10	T:1	G:0	0.5	0.0112
2586535	DMA	12	G:0.33	A:0.67	8	G:0	A:1	0.33	0.0679
2591649	DMB2	8	A:0.25	G:0.75	10	A:0.6	G:0.4	0.35	0.1376
2593501	BF1	8	A:0.5	AGC:0.5	10	A:0.2	AGC:0.8	0.3	0.1797
2597541	TAP1	10	T:0.3	A:0.7	12	T:0	A:1	0.3	0.0412
2599123	TAP1	10	T:0	G:1	12	T:0.33	G:0.67	0.33	0.0435
2604977	Downstream of TAP2	8	G:0	T:1	10	G:0.3	T:0.7	0.3	0.0897
2610020	Upstream of C4	10	G:0.6	A:0.4	12	G:1	A:0	0.4	0.0154

In addition, INDELs were spread out of a 826 bp segment, accounting for 1% of the target region, and the region affected by deletions was larger than that by insertions (**Figure 5**).

To explore the distribution of INDELs across the 18 genes in the BF/BL region, insertion and deletion of different genes were counted, and the INDEL rate was calculated as the number of INDEL per kb for each gene. We observed that the average of INDEL rate on the target region was 2 per kb. The *BLEC2* and *TAP2* showed the lowest INDEL rate. The two genes, *BF1* and *BF2*, coding the components of MHC class I molecules, showed the highest INDEL rate. Both showed more than 5 INDELs per kb. Our data showed that the *BF1* and *BF2* did not have large insertions and deletions, and the length of the INDELs in these genes was within 15 bp (**Table 5**). In order to have a deeper insight on the distribution of INDELs in the gene regions, we annotated all the detected INDELs using VEP. We found that 62% INDELs were located on introns, 14% in coding exon, 15% in downstream and upstream flanking region, 5% in the UTR regions, and 2% in

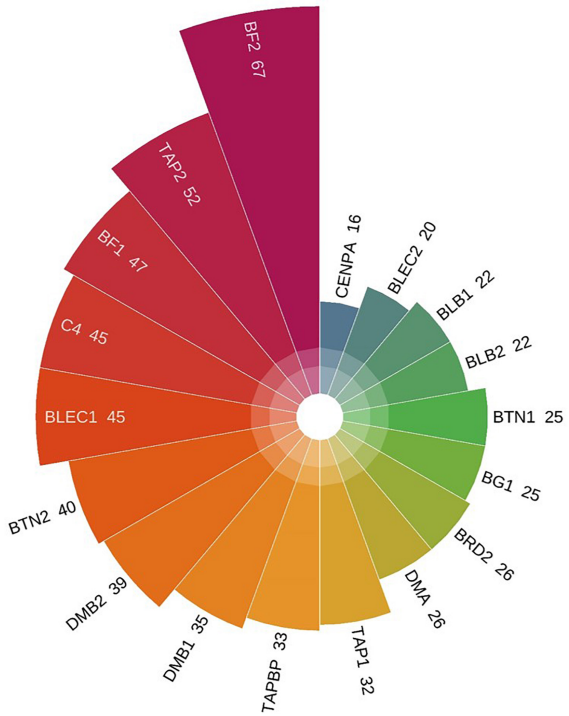

FIGURE 2 | The number of SNPs per kb in 18 genes. The SNP rates from low to high follow the clockwise direction.

splice region (**Figure 6A**). In the coding region, more than three-quarters of the identified variants were frameshift variant, 8% were inframe deletion variant, 7% were inframe insertion variant, and the remaining 3% were protein altering variant (**Figure 6B**).

DISCUSSION

Next generation sequencing (NGS) has been widely used for identification of genome-wide genetic variation (Davey et al., 2011; Nielsen et al., 2011). However, chromosome 16 has not been elucidated well by NGS analysis because the sequencing coverage of chr16 was relatively poor (only about 50%), much lower than that of the other chromosomes (Yan et al., 2014). Additionally, NGS is not economically suitable for large-scale samples due to the high cost, therefore many cost-efficient methods have been developed to analyze subsets of genome-scale data, such as target enrichment sequencing of specific genes, which can significantly reduce the costs (Dapprich et al., 2016). WaferGen, a target region sequencing platform, is a PCR based platform that can generate as many as 5,184 amplification reactions on one chip (De Wilde et al., 2014). In this study, we conducted target enrichment sequencing to evaluate the polymorphism in the BF/BL region of 21 diverse chicken populations. Our capture efficiency was 99%, which was higher than a previous study in turkey (capture efficiency 53%) using a custom MHC SureSelect capture array (Reed et al., 2016). The average depth of our data in the target region reached about 200×, and the coverage of target region with more than 30× were 55 kb on average. The mean depth of the target region in this study was three times deeper than similar reported research in human (Jones and Good, 2016), which indicated high accuracy and reliability of the variations identified in this study.

A total of 3,500 variations were detected in the BF/BL region of the 21 chicken populations in this study. Among them, 2,057 SNPs and 159 INDELs were novel, which drastically expand the current Ensembl SNP database. The number of novel INDELs detected in this study was quite large, accounted for 5% of the total variations, but most of them privately appeared in some populations, which suggested that using additional populations with large genetic background differences may detect more genetic variation of the genome. We identified an average of 36 SNPs and 2 INDELs per kb in the BF/BL region, which is higher than that reported by Brandstrom and Ellegren (2007). International Chicken Polymorphism Map Consortium reported an average mutation density of 5 SNP/kb in the chicken genome (Wong et al., 2004), and that is much lower than what we found in the region, which also indicated high polymorphism of MHC BF/BL region. More than one-third of the variations have a minimum allele frequency above 0.05 in the populations, which meets the requirements of molecular markers for genetic analysis and may be implemented in breeding programs for genetic improvement (Mills et al., 2006, 2011; Vali et al., 2008; Maw et al., 2013).

Among these 18 genes, BF gene, and TAP gene showed high diversity, which is consistent with their antigen presenting roles. The BF genes encode typical MHC class I glycoproteins. The exon 2 and exon 3 of BF genes encoding α1 and α2 domains of MHC-I are related to the resistance of many avian diseases (Hunt and Fulton, 1998; Kaufman et al., 1999; Rogers et al., 2003). TAP molecules are a part of the MHC class I antigen-processing pathway (Pamer and Cresswell, 1998). TAP heterodimer formed by TAP1 and TAP2 transports the antigen peptides into endoplasmic reticulum (ER; Deverson et al., 1990; Trowsdale et al., 1990; Spies et al., 2008). In human, one TAP heterodimer, four tapasin molecules, and four class I-β2m molecules form a peptide-loading complex that assists efficient loading of peptide onto class I molecules (Ortmann et al., 1997). Once the peptide in the lumen of the ER associates with class I molecules, the complex of MHC class I, β2m and the peptide will move to the cell surface, and the peptide will be presented to CD8 + T cells (Pamer and Cresswell, 1998). As key components of the MHC-I antigen presenting pathway, super high polymorphism of BF gene and TAP gene potentially contribute to highly effective antigen presentation and broad-spectrum host resistance to pathogens. Therefore, the high diversity that we found in most of the Chinese local populations implies that they potentially possess a broad-spectrum of disease resistance capability.

Among the domestic populations, NY chicken, GS chicken, XG fowl, and XY chicken were found with the highest polymorphism. These breeds were not yet subjected too strong artificial selection. NY chicken, a local breed in Guangxi province of China, showed the most abundant diversity, but the information about this chicken is limited. Guangxi local poultry resource book described that it was once a pheasant population in

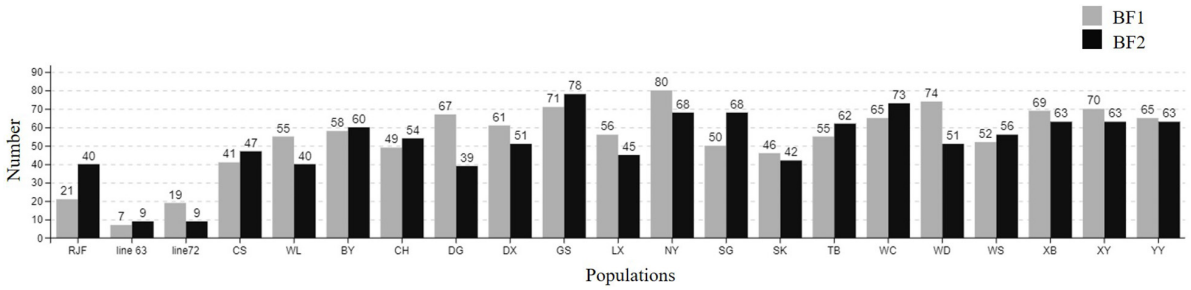

FIGURE 3 | The number of SNPs in *BF1* and *BF2* among different chicken populations.

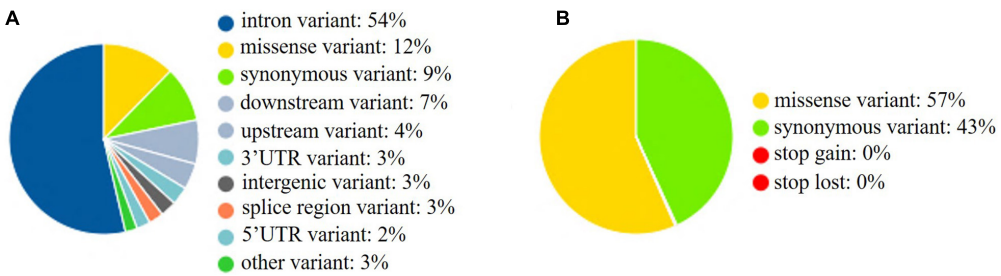

FIGURE 4 | SNPs in functional regions. **(A)** SNPs in functional regions. **(B)** SNPs in coding sequences. upstream: 500 bp apart from the transcription start site; downstream: 500 bp apart from the transcription end site.

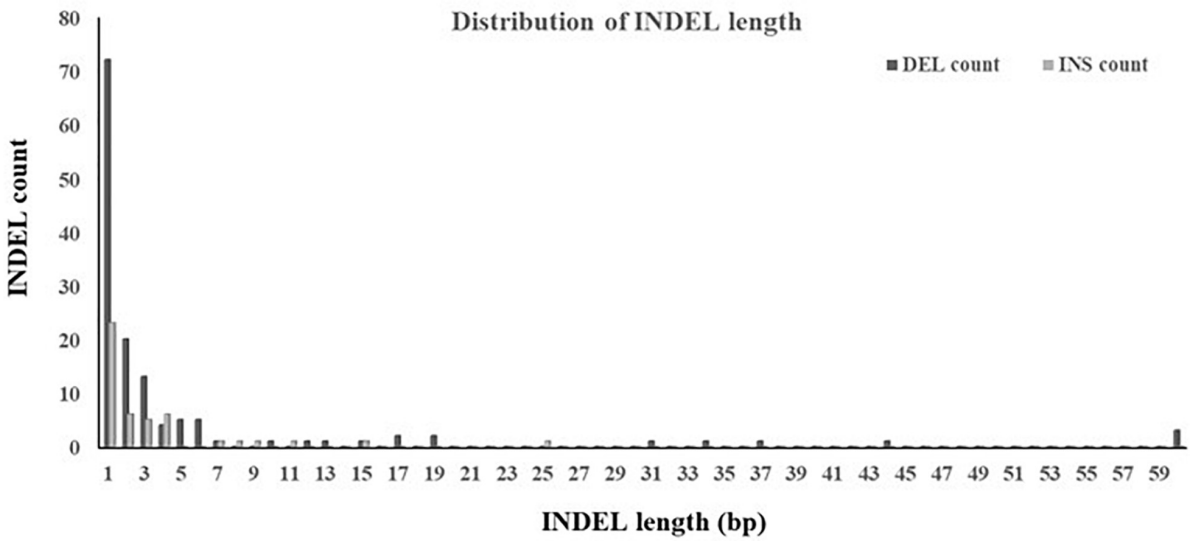

FIGURE 5 | Distribution of INDEL length. INDELs with multiple alleles were not included.

Guangxi, China. The habitat of the NY chicken was primarily in the forests, and it was domesticated after capture by the natives. Previous research suggested that RJFs showed extraordinarily high diversity (Hoa et al., 2016), but it was observed with fewer variations in our study. These might result from that only three RJFs were used in this study. High polymorphism of Chinese domestic chickens was also observed by Qu et al. (2006). They found the average heterozygosity of 78 Chinese indigenous chicken breeds was 0.622 by analyzing 27 microsatellite markers. The diversity of WL and CS in our research is lower than that of most Chinese domestic chicken populations, which might be resulted from historical selection for performance characteristics. Izadi et al. (2011) reported that intensively selected commercial populations seem to have low genetic variability in the MHC regions compared with that of the non-commercial flocks, which are less intensively selected. Manjula et al. (2020) reported that

TABLE 5 | Distribution of INDELs across 18 genes.

Gene	Length	INDEL count			Maximum length (bp)		INDEL rate(kb^{-1})	Novel (ratio)
		Num	Insertion	Deletion	Insertion	Deletion		
BTN1	14,220	28	8	20	7	60	1.9691	22(78.6%)
BTN2	6,884	14	2	12	2	60	2.0337	12(85.7%)
BG1	5,195	10	2	8	11	3	1.9249	9(90.0%)
BLEC2	2,493	1	0	1	0	3	0.4011	1(100%)
BLEC1	4,150	9	3	6	15	6	2.1687	8(88.9%)
BLB1	1,868	2	2	0	4	0	1.0707	2(100%)
TAPBP	3,532	7	2	5	2	5	1.9819	6(85.7%)
BLB2	1,537	1	0	1	0	1	0.6506	1(100%)
BRD2	8,854	8	3	5	1	6	0.9035	6(75.0%)
DMA	2,210	2	0	2	0	60	0.9050	2(100%)
DMB1	2,462	4	1	3	1	37	1.6247	4(100%)
DMB2	3,024	12	2	10	25	15	3.9683	9(75.0%)
BF1	2,803	15	4	11	4	12	5.3514	15(100%)
TAP1	5,195	4	1	3	4	2	0.7700	4(100%)
TAP2	3,563	2	0	2	0	2	0.5613	1(50.0%)
BF2	2,204	13	6	7	4	3	5.8984	13(100%)
C4	14,355	27	3	24	3	60	1.8809	24(88.9%)
CENPA	1,610	4	2	2	1	2	2.4845	4(100%)

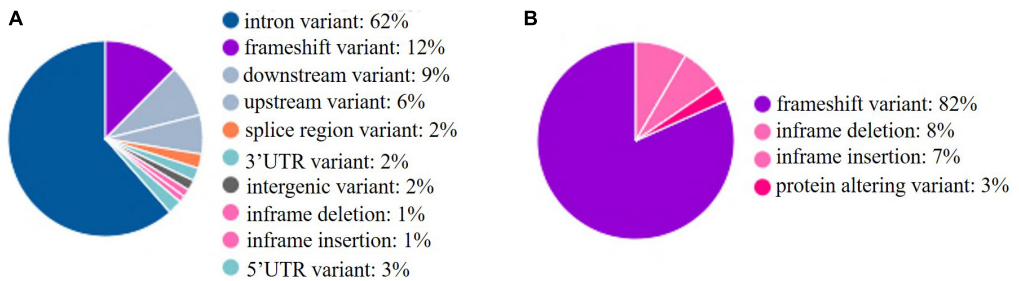

FIGURE 6 | INDELs in functional regions. **(A)** INDELs in functional regions. **(B)** INDELs in coding sequences. upstream: 500 bp apart from the transcription start site; downstream: 500 bp apart from the transcription end site.

compared with commercial breeds, Korea native breeds had high genetic diversity in the MHC-B region, and they showed immune capabilities and genetic potential for resistance to many different pathogens. Hillel et al. (2003) used 22 microsatellite markers to analyze the polymorphisms of 52 chicken breeds from different countries and found that the average heterozygosity of local breeds was 0.50, but that of the commercial chicken breeds was only 0.29, which indicated that long-term intensity selection reduces population diversity. In this study, we found the polymorphisms of line 6_3 and line 7_2 were among the lowest, which must be resulted from high-pressure selection and is in good agreement with the other studies (Waters, 1945; Stone, 1975; Bacon et al., 2000).

CONCLUSION

We identified abundant genetic variations in the chicken MHC-B-F/B-L region, which not only strikingly expanded the current Ensembl SNP database, but also provided comprehensive data for researchers to further investigate association between variants in MHC and immune traits. The Chinese domestic breeds showed high diversity, which suggest their broad-spectrum resistance to pathogens. Long-term intensity selection significantly reduced diversity of MHC B-F/B-L region of the highly inbred line 6_3 and 7_2 chickens.

DATA AVAILABILITY STATEMENT

All raw sequence data had been deposited in NCBI Sequence Read Achieve (SRA) under the Bioproject number PRJNA728310. The experiment numbers for the 195 chickens are SAMN19075170–SAMN19075364.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Review Committee for Laboratory Animal Welfare and Animal

Experiment of China Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

YY performed the experiments, analyzed the data, and wrote the manuscript. HZ supported the samples and provided comments on the manuscript. GY provided comments on the data analysis and revised the manuscript. KW provided comments on the data analysis. ZY, CZ, HY, and JL prepared and detected the samples. NY supported the samples and provided comments on the experiment. LL conceived and designed the experiment, provided comments on the data analysis, manuscript writing, and revision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (U1901206 and 31301957), the Young Scientist Supporting Project, Program for Changjiang Scholars and Innovative Research Team in University (IRT_15R62), the National Germplasm Center of Domestic Animal Resources, China Agriculture Research Systems (CARS-40), and the Beijing Key Laboratory for Animal Genetic Improvement.

ACKNOWLEDGMENTS

We thank Xiaoxiang Hu for providing DNA of Red Jungle Fowl.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.710770/full#supplementary-material

Supplementary Table 1 | Primer design for chicken MHC BF/BL region using WaferGen Smartchip Platform.

Supplementary Table 2 | The number of SNPs and INDELs in 18 genes.

REFERENCES

Bacon, L. D. (1987). Influence of the major histocompatibility complex on disease resistance and productivity. *Poult. Sci.* 66, 802–811. doi: 10.3382/ps.0660802

Bacon, L. D., Hunt, H. D., and Cheng, H. H. (2000). A review of the development of chicken lines to resolve genes determining resistance to diseases. *Poult. Sci.* 79, 1082–1093. doi: 10.1093/ps/79.8.1082

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

Brandstrom, M., and Ellegren, H. (2007). The genomic landscape of short insertion and deletion polymorphisms in the chicken (*Gallus gallus*) genome: a high frequency of deletions in tandem duplicates. *Genetics* 176, 1691–1701. doi: 10.1534/genetics.107.070805

Chang, C. S., Chen, C. F., Berthouly-Salazar, C., Chazara, O., Lee, Y. P., Chang, C. M., et al. (2012). A global analysis of molecular markers and phenotypic traits in local chicken breeds in Taiwan. *Anim. Genet.* 43, 172–182. doi: 10.1111/j.1365-2052.2011.02226.x

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

Dapprich, J., Ferriola, D., Mackiewicz, K., Clark, P. M., Rappaport, E., D'Arcy, M., et al. (2016). The next generation of target capture technologies – large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics* 17:486. doi: 10.1186/s12864-016-2836-6

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

De Wilde, B., Lefever, S., Dong, W., Dunne, J., Husain, S., Derveaux, S., et al. (2014). Target enrichment using parallel nanoliter quantitative PCR amplification. *BMC Genomics* 15:184. doi: 10.1186/1471-2164-15-184

Deverson, E. V., Gow, I. R., Coadwell, W. J., Monaco, J. J., Butcher, G. W., and Howard, J. C. (1990). Mhc Class-Ii region encoding proteins related to the multidrug resistance family of transmembrane transporters. *Nature* 348, 738–741. doi: 10.1038/348738a0

Esmailnejad, A., Brujeni, G. N., and Badavam, M. (2017). LEI0258 microsatellite variability and its association with humoral and cell mediated immune responses in broiler chickens. *Mol. Immunol.* 90, 22–26. doi: 10.1016/j.molimm.2017.06.027

Ewald, S. J., and Livant, E. J. (2004). Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. *Poult. Sci.* 83, 600–605. doi: 10.1093/ps/83.4.600

Fulton, J. E., Juul-Madsen, H. R., Ashwell, C. M., McCarron, A. M., Arthur, J. A., O'Sullivan, N. P., et al. (2006). Molecular genotype identification of the Gallus gallus major histocompatibility complex. *Immunogenetics* 58, 407–421. doi: 10.1007/s00251-006-0119-0

Guillemot, F., Billault, A., Pourquie, O., Behar, G., Chausse, A. M., Zoorob, R., et al. (1988). A molecular map of the chicken major histocompatibility complex – the Class-Ii beta-genes are closely linked to the class-i genes and the nucleolar organizer. *EMBO J.* 7, 2775–2785. doi: 10.1002/j.1460-2075.1988.tb03132.x

Han, B., Lian, L., Qu, L. J., Zheng, J. X., and Yang, N. (2013). Abundant polymorphisms at the microsatellite locus LEI0258 in indigenous chickens. *Poult. Sci.* 92, 3113–3119. doi: 10.3382/ps.2013-03416

Hillel, J., Groenen, M. A. M., Tixier-Boichard, M., Korol, A. B., David, L., Kirzhner, V. M., et al. (2003). Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genet. Sel. Evol.* 35, 533–557. doi: 10.1186/1297-9686-35-6-533

Hoa, N. P., Fulton, J. E., and Berres, M. E. (2016). Genetic variation of major histocompatibility complex (MHC) in wild red Junglefowl (*Gallus gallus*). *Poult. Sci.* 95, 400–411. doi: 10.3382/ps/pev364

Huang, X., Li, L., Zhang, J., He, D., Zhang, X., Chen, J., et al. (2016). Evaluation of diversity and evolution of the microsatellite LEI0258 in chicken MHC-B region from South China. *Acta Vet. Zootech. Sin.* 47, 2175–2183.

Hunt, H. D., and Fulton, J. E. (1998). Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken. *Immunogenetics* 47, 456–467. doi: 10.1007/s002510050383

Izadi, F., Ritland, C., and Cheng, K. M. (2011). Genetic diversity of the major histocompatibility complex region in commercial and noncommercial chicken flocks using the LEI0258 microsatellite marker. *Poult. Sci.* 90, 2711–2717. doi: 10.3382/ps.2011-01721

Jones, M. R., and Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* 25, 185–202. doi: 10.1111/mec.13304

Kaufman, J. (2018). Generalists and specialists: a new view of how MHC Class I molecules fight infectious pathogens. *Trends Immunol.* 39, 367–379. doi: 10.1016/j.it.2018.01.001

Kaufman, J., Milne, S., Gobel, T. W. F., Walker, B. A., Jacob, J. P., Auffray, C., et al. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401, 923–925. doi: 10.1038/44856

Kaufman, J., Volk, H., and Wallny, H. J. (1995). A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol. Rev.* 143, 63–88. doi: 10.1111/j.1600-065x.1995.tb00670.x

Lamont, S. J. (1989). The chicken major histocompatibility complex in disease resistance and poultry breeding. *J. Dairy Sci.* 72, 1328–1333. doi: 10.3168/jds.S0022-0302(89)79240-7

Lamont, S. J., Hou, Y. H., Young, B. M., and Nordskog, A. W. (1987). Research note: differences in major histocompatibility complex gene frequencies associated with feed efficiency and laying performance. *Poult. Sci.* 66, 1064–1066. doi: 10.3382/ps.0661064

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

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

Manjula, P., Bed'Hom, B., Hoque, M. R., Cho, S., Seo, D., Chazara, O., et al. (2020). Genetic diversity of MHC-B in 12 chicken populations in Korea revealed by single-nucleotide polymorphisms. *Immunogenetics* 72, 367–379. doi: 10.1007/s00251-020-01176-4

Manjula, P., Kim, M., Cho, S., Seo, D., and Lee, J. H. (2021). High levels of genetic variation in MHC-linked microsatellite markers from native chicken breeds. *Genes* 12:240. doi: 10.3390/genes12020240

Maw, A. A., Shimogiri, T., Yamamoto, K., Kawabe, K., Hamada, K., Kawamoto, Y., et al. (2013). The genetic diversity of eight chicken populations assessed by 102 indels markers. *J. Poult. Sci.* 50, 99–103. doi: 10.2141/jpsa.0120088

McConnell, S. K., Dawson, D. A., Wardle, A., and Burke, T. (1999). The isolation and mapping of 19 tetranucleotide microsatellite markers in the chicken. *Anim. Genet.* 30, 183–189. doi: 10.1046/j.1365-2052.1999.00454.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, 1297–1303. doi: 10.1101/gr.107524.110

Miller, M. M., Bacon, L. D., Hala, K., Hunt, H. D., Ewald, S. J., Kaufman, J., et al. (2004). 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics* 56, 261–279. doi: 10.1007/s00251-004-0682-1

Miller, M. M., and Taylor, R. L. (2016). Brief review of the chicken major histocompatibility complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poult. Sci.* 95, 375–392. doi: 10.3382/ps/pev379

Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., et al. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* 16, 1182–1190. doi: 10.1101/gr.4565806

Mills, R. E., Pittard, W. S., Mullaney, J. M., Farooq, U., Creasy, T. H., Mahurkar, A. A., et al. (2011). Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* 21, 830–839. doi: 10.1101/gr.115907.110

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986

Ortmann, B., Copeman, J., Lehner, P. J., Sadasivan, B., Herberg, J. A., Grandea, A. G., et al. (1997). A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* 277, 1306–1309. doi: 10.1126/science.277.5330.1306

Pamer, E., and Cresswell, P. (1998). Mechanisms of MHC class I - Restricted antigen processing. *Annu. Rev. Immunol.* 16, 323–358. doi: 10.1146/annurev.immunol.16.1.323

Pazderka, F., Longenecker, B. M., Law, G. R. J., and Ruth, R. F. (1975). The major histocompatibility complex of the chicken. *Immunogenetics* 2, 101–130. doi: 10.1007/BF01572280

Plachy, J., Pink, J. R., and Hala, K. (1992). Biology of the chicken MHC (B complex). *Crit. Rev. Immunol.* 12, 47–79.

Qu, L. J., Li, X. Y., Xu, G. F., Chen, K. W., Yang, H. J., Zhang, L. C., et al. (2006). Evaluation of genetic diversity in Chinese indigenous chicken breeds using microsatellite markers. *Sci. China Ser. C-Life Sci.* 49, 332–341. doi: 10.1007/s11427-006-2001-6

Reed, K. M., Mendoza, K. M., and Settlage, R. E. (2016). Targeted capture enrichment and sequencing identifies extensive nucleotide variation in the turkey MHC-B. *Immunogenetics* 68, 219–229. doi: 10.1007/s00251-015-0893-7

Rogers, S., Shaw, I., Ross, N., Nair, V., Rothwell, L., Kaufman, J., et al. (2003). Analysis of part of the chicken Rfp-Y region reveals two novel lectin genes, the first complete genomic sequence of a class I alpha-chain gene, a truncated class II beta-chain gene, and a large CR1 repeat. *Immunogenetics* 55, 100–108. doi: 10.1007/s00251-003-0553-1

Shiina, T., Briles, W. E., Goto, R. M., Hosomichi, K., Yanagiya, K., Shimizu, S., et al. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *J. Immunol.* 178, 7162–7172. doi: 10.4049/jimmunol.178.11.7162

Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., et al. (2008). A gene in the human major histocompatibility complex class II region controlling the class I antigen presentation pathway (Reprinted from Nature, vol 348, pg 744-747, 1990). *J. Immunol.* 180, 2737–2740.

Stone, H. A. (1975). *Use of Highly Inbred Chickens in Research. USDA Agricultural Research Service Technical Bulletin No. 1514.* Washington, DC: USDA.

Trowsdale, J., Hanson, I., Mockridge, I., Beck, S., Townsend, A., and Kelly, A. (1990). Sequences encoded in the Class-Ii Region of the Mhc Related to the Abc superfamily of transporters. *Nature* 348, 741–744. doi: 10.1038/348741a0

Vali, U., Brandstrom, M., Johansson, M., and Ellegren, H. (2008). Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genet.* 9:8. doi: 10.1186/1471-2156-9-8

Waters, N. F. (1945). Breeding for resistance and susceptibility to avian lymphomatosis. *Poult. Sci.* 24, 259–269. doi: 10.3382/ps.0240259

Wong, G. K. S., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z. J., et al. (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432, 717–722. doi: 10.1038/nature03156

Yan, Y. Y., Yi, G. Q., Sun, C. J., Qu, L. J., and Yang, N. (2014). Genome-wide characterization of insertion and deletion variation in chicken using next generation sequencing. *PLoS One* 9:e104652. doi: 10.1371/journal.pone.0104652

Yuhki, N., and Obrien, S. J. (1990). DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. *Proc. Natl. Acad. Sci. U.S.A.* 87, 836–840. doi: 10.1073/pnas.87.2.836

Zekarias, B., Ter Huurne, A. A. H. M., Landman, W. J. M., Rebel, J. M. J., Pol, J. M. A., and Gruys, E. (2002). Immunological basis of differences in disease resistance in the chicken. *Vet. Res.* 33, 109–125. doi: 10.1051/vetres:2002001

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

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fvets.2021.725367

Genomic Analyses Revealed the Genetic Difference and Potential Selection Genes of Growth Traits in Two Duroc Lines

Desen Li[1†], *Min Huang*[1†], *Zhanwei Zhuang*[1], *Rongrong Ding*[1], *Ting Gu*[1], *Linjun Hong*[1], *Enqin Zheng*[1], *Zicong Li*[1], *Gengyuan Cai*[1], *Zhenfang Wu*[1,2*] *and Jie Yang*[1*]

[1] College of Animal Science and National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangzhou, China, [2] Lingnan Guangdong Laboratory of Modern Agriculture, Guangzhou, China

Duroc pigs are famous for their high growth rate, feed conversion efficiency, and lean meat percentage. Given that they have been subjected to artificial selection and breeding in multiple countries, various lines with obvious differences in production performance have formed. In this study, we genotyped 3,770 American Duroc (AD) pigs and 2,098 Canadian Duroc (CD) pigs using the GeneSeek Porcine SNP50 Beadchip to dissect the genetic differences and potential selection genes of growth traits in these two Duroc pig lines. Population structure detection showed that there were significant genetic differences between the two Duroc pig lines. Hence, we performed F_{ST} and cross-population extended haplotype homozygosity (XP-EHH) analyses between the two lines. As a result, we identified 38 annotated genes that were significantly enriched in the gland development pathway in the AD line, and 61 annotated genes that were significantly enriched in the immune-related pathway in the CD line. For three growth traits including backfat thickness (BFT), loin muscle depth (LMD), and loin muscle area (LMA), we then performed selection signature detection at 5 and 10% levels within the line and identified different selected regions and a series of candidate genes that are involved in lipid metabolism and skeletal muscle development or repair, such as *IRX3*, *EBF2*, *WNT10B*, *TLR2*, *PITX3*, and *SGCD*. The differences in selected regions and genes between the two lines may be the cause of the differences in growth traits. Our study suggests significant genetic differences between the AD and CD lines, which provide a theoretical basis for selecting different Duroc lines as sires for different needs.

Keywords: Duroc, genetic difference, growth trait, F_{ST}, XP-EHH

OPEN ACCESS

Edited by:
Hai Xiang,
Foshan University, China

Reviewed by:
Donghyun Shin,
Jeonbuk National University, South Korea
Sayed Haidar Abbas Raza,
Northwest A and F University, China

****Correspondence:***
Zhenfang Wu
wzfemail@163.com
Jie Yang
jieyang2012@hotmail.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 Veterinary Science

Received: *15 June 2021*
Accepted: *03 August 2021*
Published: *07 September 2021*

Citation:
Li D, Huang M, Zhuang Z, Ding R, Gu T, Hong L, Zheng E, Li Z, Cai G, Wu Z and Yang J (2021) Genomic Analyses Revealed the Genetic Difference and Potential Selection Genes of Growth Traits in Two Duroc Lines. Front. Vet. Sci. 8:725367. doi: 10.3389/fvets.2021.725367

INTRODUCTION

Lean pigs have been selected and bred in various countries for a long time and have formed distinctive lines, such as English Large White and French Large White. Since the 1980s, China introduced Duroc pigs from America (American line), Canada (Canadian line), Denmark (Danish line), China Taiwan (Taiwan line), and Japan (Japanese line). Among these Duroc lines, the Taiwan line has the characteristics of beautiful body shape, rough feeding resistance, and strong disease resistance, but the growth rate is slow during the late fattening period. The American line has a higher growth rate, stress resistance, and lean meat

percentage than the Taiwan line. The Canadian line is well-known for its high average daily gain, rich intramuscular fat (IMF) content, and excellent meat quality (1).

Quan et al. assessed the carcass traits of American Duroc × (Landrace × Yorkshire) three-way cross hybrid (ADLY) pigs and Taiwan Duroc × (Landrace × Yorkshire) three-way cross hybrid (TDLY) pigs and found that the lean meat percentage of ADLY pigs was better than that of TDLY pigs (ADLY: 57.39 vs. TDLY: 55.27%) ($p < 0.01$), while the live mass (ADLY: 104.06 vs. TDLY: 110.02 kg), carcass mass (ADLY: 88.31 vs. TDLY: 94.14 kg), loin muscle depth (LMD) (ADLY: 51.16 vs. TDLY: 54.61 cm), and carcass italic length (ADLY: 84.16 vs. TDLY: 86.01 cm) of TDLY were better than those of ADLY ($p < 0.01$) (2). Zhuang et al. conducted genome-wide association studies (GWAS) for loin muscle area (LMA) and LMD in American and Canadian Duroc pigs and identified 75 significantly associated SNPs, of which a 283-kb region on chromosome 7 was a pleiotropic quantitative trait loci (QTL) that affected both traits. Among these 75 SNPs, the ALGA0040260 marker was the key SNP for the QTL and explained 1.77 and 2.48% of the phenotypic variance in LMA and LMD, respectively (3). In addition, Zhuang et al. also performed GWAS for teat number in American and Canadian Duroc pigs and detected a QTL on chromosome 7 with marker rs692640845 explaining 8.68% of the phenotypic variance in the Canadian Duroc (4). Although these studies conducted GWAS for different lines of Duroc pigs and identified different candidate genes and QTLs, they did not uncover the genetic differences between different Duroc lines.

For different production needs, different lines of pigs have formed their own characteristics through artificial selection. Animals are usually selected for certain traits, and the internal mechanism is the selection of genes. The selection signature detection can reveal potential selection genes, which is of great significance for understanding the evolution of species and identifying genes for economic traits. Ma et al. performed the cross-population extended haplotype homozygosity (XP-EHH) and F_{ST} to detect trait-specific selection signatures by making backfat thickness (BFT) gradient differential population pairs in Yorkshire pigs, and identified that a number of genes were associated with fat metabolism, such as *OSBPL8*, *ASAH2*, *GBE1*, and *ABL1* (5). Kim et al. used the Duroc pigs that were sampled from the sixth generation of a selection experiment for IMF to divide the high and low IMF groups and to conduct selection signature detection, and a total of 16 consensus regions were obtained using the three methods [including F_{ST}, the integrated haplotype score (IHS), and the standardized score of the ratio of extended haplotype homozygosity (Rsb)] (6). The above studies show the feasibility of dividing the phenotypic gradient groups within the population for selective signature detection, but there is no relevant research on the use of this method in different lines of Duroc.

Therefore, in this study, we used Porcine SNP50 Beadchip to genotype 3,770 America Duroc (AD) and 2,098 Canadian Duroc (CD) pigs and carried out selection signature detection between the two lines. Besides, for the same line, we also performed selection signature detection through dividing extreme phenotypic groups according to the estimated breeding values (EBV) ranking of BFT, LMD, and LMA, and then have identified the selected regions and genes in different lines to reveal the potential genetic mechanisms that caused the differences in growth traits.

MATERIALS AND METHODS

Ethics Statement

All animals used in this study met the guidelines for the care and use of experimental animals established by the Ministry of Agriculture of China. The whole of this study was approved by the ethics committee of South China Agriculture University (SCAU, Guangzhou, China), and written informed consent was obtained prior to data collection from Wens Foodstuff Group Co., Ltd. (Guangdong, China). There was no use of human participants, data, or tissues.

Sample collection, SNP Genotyping, and Phenotype Detection

A total of 3,770 American Duroc pigs and 2,098 Canadian Duroc pigs were genotyped using the GeneSeek Porcine SNP50 Beadchip in this study. All pigs in the two populations sustained uniform feeding conditions, fine fodder, and consistent management during the fattening period from 30 to 100 kg live weight to minimize the impact of non-genetic factors. The details of sample collection, DNA extraction, SNP genotyping, and recording of LMA and LMD were described by Zhuang et al. (3). In addition, BFT was obtained by measuring the thickness of backfat between the 10th and 11th rib of the pigs at the weight of 100 ± 5 kg using an Aloka 500 V SSD B ultrasound (Corometrics Medical Systems, USA). The three phenotypes were corrected by 100 kg body weight. Quality filtering of genotypes was performed using PLINK v1.9 (7) with the criteria of minor allele frequencies (MAF) >0.01, individual call rate >95%, and SNP call rate >95%. After removing non-autosomal and unmapped SNPs, a total of 39,567 SNPs remained and used in subsequent analyses.

Population Structure and Estimation of Inbreeding Coefficient

Genetic distance among individuals was calculated via an identity-by-state (IBS) similarity matrix by PLINK v1.9. A neighbor-joining relationship tree (NJ-tree) based on the genetic distance was constructed using PHYLP v3.69 (8) and was visualized using Figtree v1.4 (9). We randomly selected 100 individuals from each of the two lines 10 times and used PLINK v1.9 to estimate linkage disequilibrium decay (LD decay) distance. When $r^2 = 0.3$, the physical distance was used to identify the range of the annotated genes (10). Runs of homozygosity (ROH) of the two lines were performed using the consecutiveRUNS.run function of the R detectRUNS package (11). The inbreeding coefficient based on ROH (F_{ROH}) was calculated for each individual using the following formula (12):

$$F_{ROH} = \frac{\sum_i L_{ROH}}{L_{AUTO}}$$

where L_{ROH} is the length of ROH of individual i, and L_{AUTO} is the autosomal genome length covered by the SNPs in this study.

Estimated Breeding Value Calculation

Additive effect (breeding value), dominance effect, and epistatic effect can affect quantitative traits, among which additive effect can be stably inherited by offspring, which means the EBV can be calculated from phenotypes and parentage. In this study, the EBV of BFT, LMD, and LMA were calculated based on the restricted maximum likelihood (REML) method via the dumai model of DMU software (13). The calculation model is as follows:

$$\boldsymbol{y} = \boldsymbol{u} + \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}$$

where y is the vector of phenotypic values, b is the fixed effects vector, including sex, farm, year, season, and parity (**Supplementary Table 1**), a is the vector of individual random additive effect, e is the vector of random residuals, and X and Z are the structural matrices of a and b, respectively. For BFT, a single-trait animal model was used to obtain the EBV. A multitrait animal model was used in LMD and LMA because of their strong correlation with muscle. As a result, 50% of EBV of the two traits was taken to obtain the total breeding value (TBV) for the two traits.

Selection Signature Detection Between the American Duroc Line and the Canadian Duroc Line

To detect the degree of genetic differentiation between the AD and CD lines, F_{ST} between the two lines was performed using VCFtools (14). The XP-EHH was operated using the "–xpehh" function of Selscan (15) based on the haplotypes constructed by Beagle (16). The top 1% values were taken as the significant thresholds of F_{ST} and XP-EHH, respectively, and the overlapping SNPs of two methods exceeding the thresholds were considered as the selected markers.

Selection Signature Detection for Genetic Differential in Different Gradients of Growth Traits Within the American Duroc Line and the Canadian Duroc Line

According to the EBV ranking of BFT, LMD, and LMA in the AD and CD lines, the individuals with the top 5% EBV were selected for a group and the bottom 5% of which were selected for the other group, defined as the 5% level. A similar strategy was defined at the 10% level. Then the XP-EHH and the F_{ST} were used to detect the selection signature in the two levels for each trait within the two lines. A value of 0.05 was used as the significance threshold for F_{ST}, and the top 5% was used as the significance threshold for XP-EHH.

Candidate Gene and Functional Annotation

When $r^2 = 0.3$, the LD decay distances of the two lines were used as the upstream and downstream ranges of the selected markers to determine the selected regions. We then searched the annotated genes from the selected regions and conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis by Metascape (https://metascape.org/) to identify the candidate genes whose functions are associated with growth traits. In addition, the annotated genes were compared with the pig QTL database (https://www.animalgenome.org/) to identify the genes within the QTL regions for growth traits.

RESULTS

Population Structure

Our previous study performed principal component analysis (PCA) between these two Duroc pig populations (3), and the results showed that PC1 divided the AD and CD lines into two obvious groups, and PC2 showed that there was stratification phenomenon in the AD line as well. Then, in this study, the NJ tree showed that the two lines formed two apparently independent branches (**Supplementary Figure 1A**), and there were multiple lineages within the AD line. PCA and NJ-tree analysis revealed that there was significant genetic differentiation between the AD and CD lines, which implied that there were obvious genetic differences between the two Duroc lines. To avoid the impact of population stratification in the AD line, PCA was conducted for each lineage of the AD line and the CD line according to the NJ tree of the AD line (**Supplementary Figure 2**). Finally, a total of 1,969 AD line pigs with relatively concentrated clustering and more differentiated with the CD line were selected for subsequent analysis. The results of LD decay analysis showed that the decay rate of the CD line was slower than that of the AD line. When $r^2 = 0.3$, the average LD decay distances in the AD and CD lines were approximately 150 and 202 kb, respectively (**Supplementary Figure 1B**). Although the same Porcine SNP50 Beadchip data were utilized for LD decay analysis in the study by Zhuang et al. (4), our study had different population size, analysis methods, and threshold line criteria, and therefore, the LD decay analysis was reperformed in this study rather than directly cited. Besides, the average F_{ROH} of the AD line was lower than that of the CD line (**Supplementary Figure 1C**). The results of LD decay and F_{ROH} indicated a higher degree of inbreeding in the CD line.

EBV Calculation

We corrected and summarized the statistics for BFT, LMD, and LMA in the AD and CD lines, respectively, and all of these traits were normally distributed as shown in **Supplementary Figure 3A**. We found that there were significant differences between the two lines in these phenotypes, of which LMD and LMA in the AD line were greater than those in the CD line, while BFT was lower than the latter (**Supplementary Figure 3B**). Next, the EBV of BFT, LMD, and LMA were calculated separately in the two lines (**Supplementary Tables 2**, **3**).

Selection Signature Detection Between the American Duroc Line and the Canadian Duroc Line

Population structure analysis revealed a large degree of genetic differentiation between the AD and CD lines. Hence, we performed selection signature detection between 1,969 and 2,098

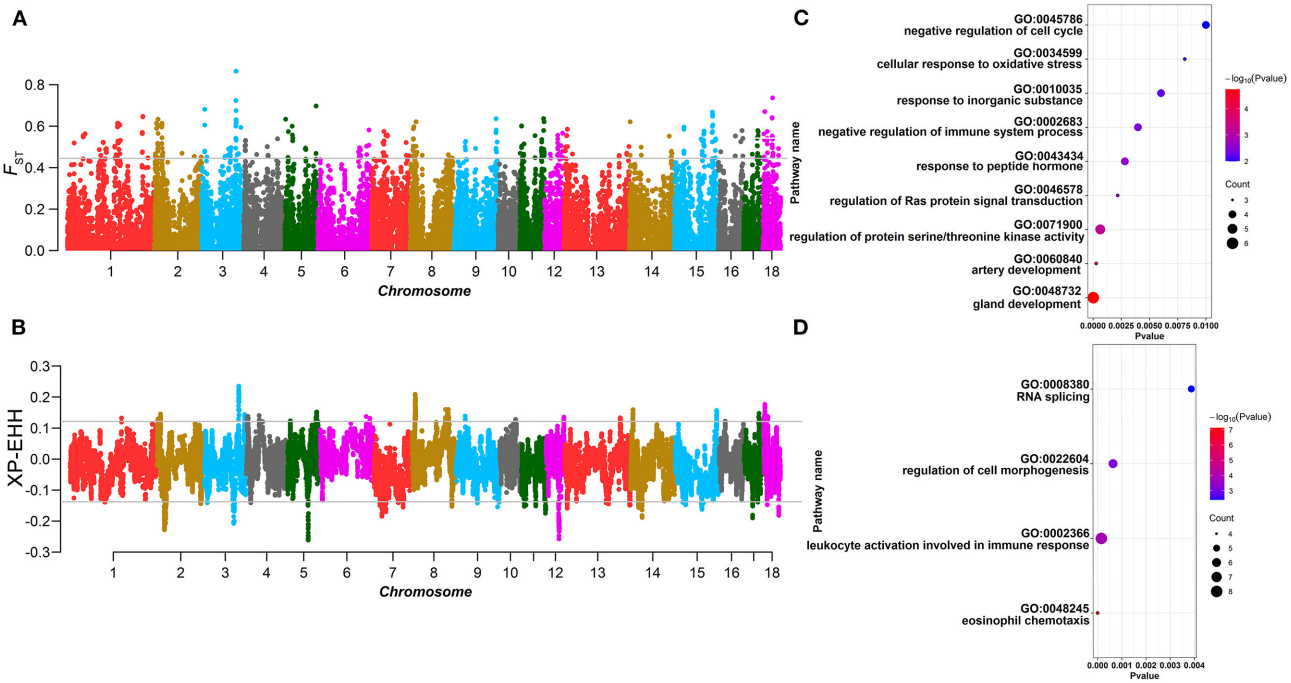

FIGURE 1 | Selection signature detection between the American Duroc (AD) line and the Canadian Duroc (CD) line. **(A)** Manhattan plot of F_{ST}. The gray line denotes threshold line ($F_{STtop1\%} = 0.44$). **(B)** Manhattan plot of XP-EHH. The gray lines denote threshold lines ($XP\text{-}EHH_{AD} = 0.12$, $XP\text{-}EHH_{CD} = -0.14$). **(C,D)** are the bubble chart of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of the selected genes identified from the AD and CD lines, respectively.

individuals from the AD and CD lines, respectively. In the XP-EHH analysis, the AD line was used as the test population and the CD line as the reference population, which means that the positive values represent recent selection in the AD line, and conversely, the negative values represent that of the CD line (**Figure 1B**). The top 1% was used as the significant threshold for F_{ST} (**Figure 1A**) and XP-EHH ($F_{STtop1\%} = 0.44$, $XP\text{-}EHH_{AD} = 0.12$, and $XP\text{-}EHH_{CD} = -0.14$), and the overlapping significant SNPs in the two statistics were defined as the selected markers. A total of 28 selected markers and 38 annotated genes were identified in the AD line from the 300-kb selected region (150 kb upstream and downstream of the selected SNPs), while a total of 30 selected markers and 61 annotated genes were found in the CD line from the 404-kb selected region (202 kb upstream and downstream of the selected SNPs). GO enrichment and KEGG pathway analysis of the annotated genes showed that the genes were significantly enriched in the gland development pathway (GO: 0048732) in the AD line (**Figure 1C**). For the CD line, the genes were significantly enriched in immune-related pathways (GO: 0048245, GO: 0002366) (**Figure 1D**; **Supplementary Table 4**).

Selection Signature of Backfat Thickness Within the American Duroc line and the Canadian Duroc Line

The 5% Level

For the AD line, according to the EBV ranking of BFT, 98 individuals with the top and the bottom EBV values were selected from 1,969 individuals as two groups to carry out F_{ST} and XP-EHH. In XP-EHH, the top 5% group was used as the test population, and the bottom 5% group was used as the reference population. The overlapping SNPs that exceeded the significant threshold in both methods ($F_{ST} > 0.05$, $XP\text{-}EHH_{AD} = 0.08$) were regarded as the selected markers in the AD line. A total of 46 selected markers and 35 annotated genes were identified. GO and KEGG enrichment analysis of the annotated genes (**Supplementary Figure 4**) revealed that genes such as *IRX3* and *EBF2* are involved in lipid metabolism. Similarly, 105 individuals with the top and bottom EBV values were selected from the CD line, respectively, as the two groups to detect the selection signature. A total of 131 selected markers exceeded the thresholds in both F_{ST} and XP-EHH ($XP\text{-}EHH_{CD} = 0.05$), and 167 annotated genes were identified. We found that *SAMD4A*, *DLGAP5*, *CTSF*, etc. (**Figures 2A–D**; **Table 1**) are involved in lipid deposition-related pathways according to GO and KEGG enrichment analysis (**Supplementary Figure 5**).

The 10% Level

For the AD line, 196 individuals were selected for the top 10% and bottom 10% groups, respectively, and were used to perform selection signature analysis. The results showed that there were differences from the 5% level. Several significantly differentiated areas at the 5% level disappeared at the 10% level, which means that with an increase in the number of selected individuals for the same trait, the selection intensity within the same Duroc line decreased. Five selected markers were identified in the AD line,

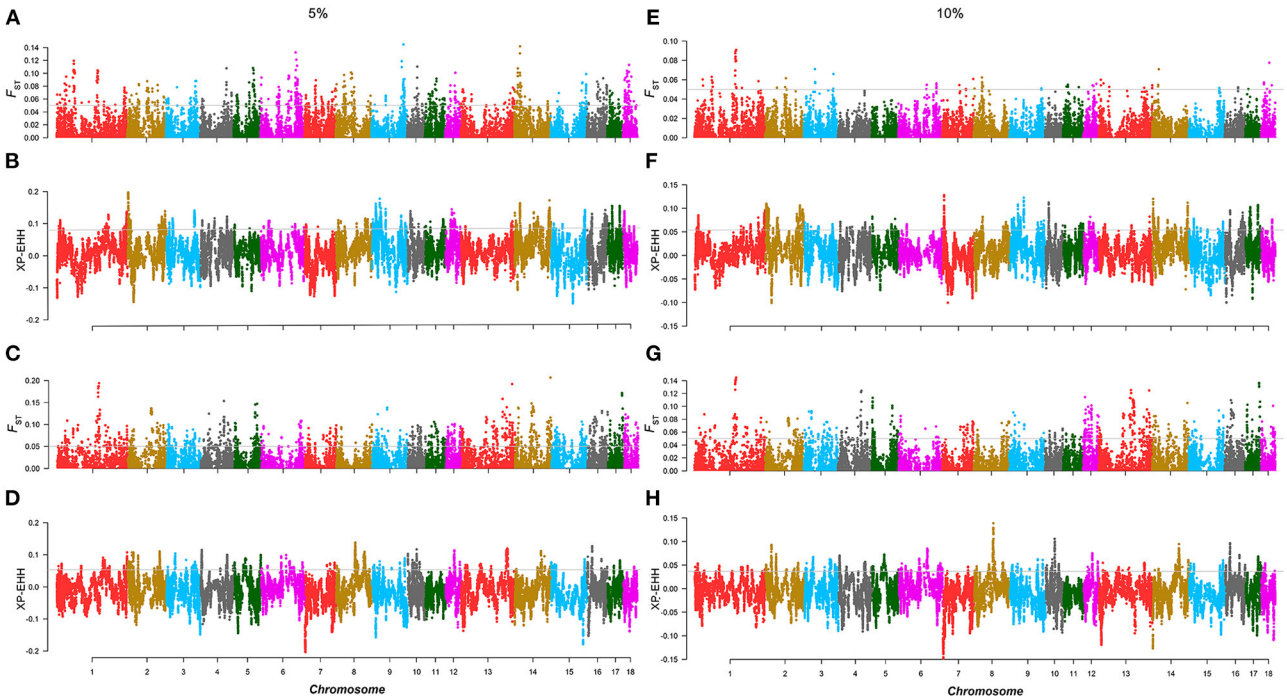

FIGURE 2 | Selection signature detection for backfat thickness (BFT) in the 5 and 10% levels. **(A,C,E,G)** are the Manhattan plots of F_{ST} of the AD and CD lines in the 5 and 10% level. The gray lines denote threshold lines ($F_{ST} = 0.05$). **(B)**, **(D,F,H)** are the Manhattan plots of XP-EHH of AD and CD lines in the 5 and 10% levels. The gray lines denote threshold lines ($XP\text{-}EHH_{5\%AD} = 0.080$, $XP\text{-}EHH_{5\%CD} = 0.053$, $XP\text{-}EHH_{10\%AD} = 0.054$, $XP\text{-}EHH_{10\%CD} = 0.036$).

and seven annotated genes were identified in the selected regions, but there were no genes related to BFT or lipid traits. For the CD line, 210 individuals in the top 10% and bottom 10% groups were selected for analysis, respectively, and a total of 38 selected markers were found, including 72 annotated genes, among which *SERPINE1*, *PROX2*, *GLP2R*, etc., were related to lipid metabolism (**Figures 2E–H**; **Table 1**). In addition, in the AD line, two overlapping selected markers and one gene *ATP8A1* were identified from both the 5 and 10% levels. *ATP8A1* participates in catalyzing the hydrolysis of ATP coupled to the transport of amino phospholipids from the outer to the inner leaflet of various membranes and ensures the maintenance of the asymmetric distribution of phospholipids (17). In the CD line, there were 18 overlapping selected markers and 19 genes from the two levels of which *PROX2* was related to the BFT trait and located in the obesity index QTL region (18) (**Supplementary Table 5**).

Selection Signature of Loin Muscle Depth and Loin Muscle Area Within the American Duroc Line and the Canadian Duroc Line

The 5% Level

Similar to the above strategy, according to the rank of the TBV estimated from LMD and LMA, two 5% level groups were used to detect selection signatures in the two lines, respectively. For the AD line, we found 71 selected markers, and 130 genes were identified. GO and KEGG enrichment analyses were performed for these genes, among which *WNT10B*, *TLR2*, *PITX3*, and *SGCD* were involved in muscle tissue regulation and muscle development (**Supplementary Figure 6**). Similarly, the same method was performed in the CD line, and the results showed that 103 selected markers and 97 genes were identified. GO and KEGG enrichment analysis (**Supplementary Figure 7**) showed that *TMOD3*, *NEGR1*, and *PITX2* were associated with the trait (**Figures 3A–D**; **Table 2**).

The 10% level

Similar to the results of the 10% level analysis in BFT, the regions of significant differentiation in the 5% level analysis of LMD and LMA disappeared in that of the 10% level. For the AD line, a total of 22 selected markers were found, and 30 genes were identified, but none of these genes were related to muscle traits. As for the CD line, 58 selected markers and 23 annotated genes were identified, of which only *SLC44A5* was related to muscle development (**Figures 3E–H**). In addition, in the AD line, 6 overlapping selected markers and eight genes were identified from both the 5 and 10% levels, while a total of 39 overlapping selected markers and eight genes were identified in the CD line (**Supplementary Table 6**).

DISCUSSION

In this study, we conducted genomic analyses of 3,770 American Duroc pigs and 2,098 Canadian Duroc pigs to dissect the genetic differences and potential selection genes of growth traits in the

TABLE 1 | Candidate genes for BFT in the 5% level analysis.

Level	Line	Chr	Pos (bp)	Candidate gene	QTL
5%	AD	6	31,046,098–31,048,822	*IRX3*	-
		14	9,816,993–10,025,165	*EBF2*	-
	CD	1	183,915,339–184,140,122	*SAMD4A*	Intramuscular fat content
		1	184,499,914–184,550,020	*DLGAP5*	Intramuscular fat content
		2	5,855,111–5,860,253	*CTSF*	-
		2	6,070,897–6,074,146	*CD248*	-
		7	97,614,707–97,624,273	*VRTN*	Intramuscular fat content et al.
		7	97,730,514–97,740,331	*NPC2*	Obesity Index
		7	98,065,323–98,074,822	*PROX2*	Obesity Index
		12	33,265,529–33,299,630	*SCPEP1*	-
10%	CD	3	8,863,695–8,871,564	*SERPINE1*	Abdominal fat weight Backfat weight et al.
		7	98,065,323–98,074,822	*PROX2*	Obesity Index
		12	54,680,057–54,738,103	*GLP2R*	Intramuscular fat content
		12	55,190,429–55,278,539	*MYH4*	Intramuscular fat content
		12	55,347,087–55,375,353	*MYH3*	-
		12	55,438,196–55,454,590	*TMEM220*	Intramuscular fat content

Chr, chromosome; Pos (bp), gene position in Ensembl; QTL, quantitative trait loci.

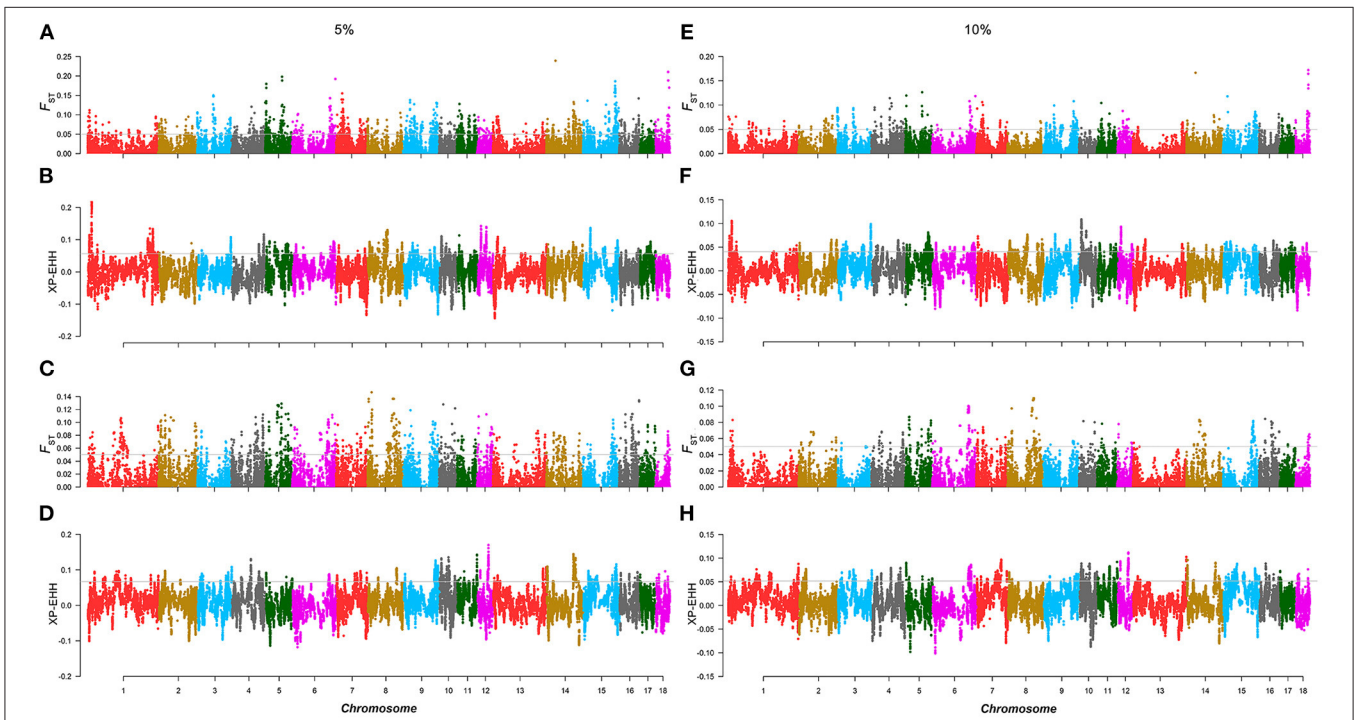

FIGURE 3 | Selection signature detection for LMD and LMA in the 5 and 10% level. **(A,C,E,G)** are the Manhattans plots of F_{ST} of AD and CD lines in the 5 and 10% level. The gray lines denote threshold lines ($F_{ST} = 0.05$). **(B,D,F,H)** are the Manhattan plots of XP-EHH of AD and CD lines in the 5 and 10% level. The gray lines denote threshold lines ($XP\text{-}EHH_{5\%AD} = 0.057$, $XPEHH_{5\%CD} = 0.067$, $XP\text{-}EHH_{10\%AD} = 0.040$, $XP\text{-}EHH_{10\%CD} = 0.052$).

two Duroc pig populations. Our results showed that the BFT in the CD line was higher than that of the AD line, while the LMD and LMA were lower than those of the AD line, which was consistent with the description of Wang et al. (1). Here, we think that the genetic differences between the two lines account for the differences in these traits. Therefore, we performed PCA (3), NJ-tree analysis, and LD decay analysis in the two lines, and the results showed that the AD and CD lines were clearly divided into two separate groups, with multiple branches within the AD line. The results of LD decay and F_{ROH} indicated that the CD

TABLE 2 | Candidate genes for loin muscle depth (LMD) and loin muscle area (LMA) in the 5% level analysis.

Level	Line	Chr	Pos (bp)	Candidate gene	QTL
5%	AD	5	14,997,381–15,013,919	*WNT10B*	Average backfat thickness et al.
		8	75,411,991–75,446,850	*TLR2*	-
		14	113,230,965–113,241,360	*PITX3*	-
		16	66,452,416–66,887,924	*SGCD*	-
	CD	1	119,862,244–119,948,883	*TMOD3*	-
		6	140,780,452–141,647,813	*NEGR1*	-
		8	111,698,664–111,723,298	*PITX2*	Drip loss et al.

Chr, chromosome; Pos (bp), gene position in Ensembl; QTL, quantitative trait loci.

line exhibited higher inbreeding. According to the differences in traits, we hypothesized that the genetic differences between the two lines may be caused by a combination of natural selection and artificial selection based on different breeding criteria, which resulted in phenotypic differentiation. Therefore, we performed selection signature analysis to detect the different selected genes in the two lines. The results showed that a series of genes selected in the AD line were enriched in the gland development pathway, while the genes were mainly involved in immune-related pathways in the CD line.

Many effects can affect quantitative traits, among which additive effect can be stably inherited by offspring. The EBV can be used for early selection, and even before the individual is born, the breeding value of the offspring can be predicted according to the performance of the two parents determined by the breeding plan. The AD and CD lines are commercial populations. In response to different breeding needs, selective breeding based on EBV ranking has been widely used in business (19). The individuals with better performance in traits in the same line can be retained; otherwise, they will be eliminated. Therefore, we adopted EBV and divided extreme individuals according to the ranking of EBV for analysis in this study, which is in line with actual production needs and patterns.

To reveal the genetic mechanism of growth traits in different lines, we divided different gradient levels to perform selection signature analysis. For the *BFT*, *IRX3*, and *EBF2* related to fat metabolism were identified in the AD line at the 5% level. *IRX3* is a functional long-range target of obesity-associated variants within *FTO*, *IRX3*-deficient mice reduces body weight by reducing fat mass and increasing basal metabolic rate and browning of white adipose tissue (20), and *EBF2* promotes the recruitment of beige adipocytes in white adipose tissue and protects animals against obesity (21). Seven genes including *SAMD4A*, *DLGAP5*, *VRTN*, *NPC2*, *PROX2*, *CD248*, and *SCPEP1* were identified in the CD line. Among these genes, *SAMD4A*, *DLGAP5*, and *VRTN* located in the IMF content QTL region (22–24), and *NPC2*, *PROX2* located in the obesity index QTL region (18). A missense mutation in the *CTSF* was significantly associated with average day gain, lean meat percentage, BFT, and feed conversion efficiency according to the study by Russo et al. (25). *CD248* is a sensitive marker of adipocyte function, increased expression of which leads to disturbances in glucose metabolism and ectopic deposition of lipids (26). *SCPEP1* regulates body fat content and is correlated with IMF deposition in pigs (27). However, at the 10% level, no trait-related genes were identified in the AD line. In contrast, *GLP2R*, *MYH4*, *TMEM220*, *SERPINE1*, and *MYH3* were identified in the CD line. *GLP2R*, *MYH4*, and *TMEM220* are located within the IMF content QTL region (23, 28), and *SERPINE1* is located in the QTL regions of abdominal fat weight, backfat weight, and subcutaneous fat, respectively (29). Besides, *MYH3* is a causal gene for the ratio of muscle fiber type, IMF content, and fat formation in pigs and mice (30).

We then performed selection signature analysis for LMD and LMA based on their TBV. At the 5% level, we identified *WNT10B*, *TLR2*, *PITX3*, and *SGCD* related to skeletal muscle development and repair in the AD line. *WNT10B* is involved in the *Wnt* signaling pathway and associated with skeletal muscle developmental regulation and regeneration (31). *TLR2* controls skeletal muscle repair mechanisms following different forms of injury (32). *PITX3* is widely expressed in skeletal muscles and promotes myogenic differentiation of muscle satellite cells (33). *SGCD* is a muscular dystrophy protein-related glycoprotein and abundantly expressed in skeletal and cardiac muscles (34). In the CD line, *TMOD3*, *NEGR1*, and *PITX2* were identified. *TMOD3* is involved in the regulation of actin and skeletal muscle contractions (35). *NEGR1* mediates neural cell communication and synapse formation, and deletion of this gene leads to increased adiposity and decreased muscle quality in mice (36). *PITX2* is involved in the regulation of skeletal muscle tissue development and animal organ morphogenesis (37). However, at the 10% level, we did not find trait-related genes in the AD line, but among the 30 genes obtained from the CD line, *SLC44A5* was found and fell within the QTL region of muscle fiber diameter (38).

In conclusion, population genetic analysis based on large samples showed that there was significant genetic differentiation between Duroc pigs with different genetic backgrounds in this study, which was also reflected in traits of different lines, such as the CD line with higher BFT, and the AD line with higher LMA and LMD. Selection signature detection between the AD and CD lines showed that there were different selective regions in the two lines. For the same line, we carried out selection signature detection at different levels based on EBV of BFT, LMD, and LMA phenotypes, and a series of genes associated with the three traits were identified, further illustrating the complexity of the genetic mechanism of quantitative traits. This study reveals the genetic differences between different lines of Duroc pigs after

strong artificial selection and provides a reference for selecting different lines of Duroc pigs as sires for different needs.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://doi.org/10.6084/m9.figshare.8019551.v1.

ETHICS STATEMENT

The animal study was reviewed and approved by South China Agriculture University.

AUTHOR CONTRIBUTIONS

ZW and JY proposed the idea for the study. DL performed the bioinformatics analysis and wrote the paper. MH directed the analyses and revised the article. ZZ, RD, TG, LH, EZ, ZL, and GC collected the samples and recorded the phenotypes. ZW and GC contributed the materials. All authors reviewed and approved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 31972540), the Natural Science Foundation of Guangdong Province (Grant Nos. 2018B030313011 and 2020A1515111103), the Local Innovative and Research Teams Project of Guangdong Province (Grant No. 2019BT02N630), and the Guangdong Province Rural Revitalization Strategy Special Project (Grant No. 200-2018-XMZC-0001-107-0145).

ACKNOWLEDGMENTS

The authors would like to thank all the staff at the pig core breeding farms of Wens Foodstuff Group Co., Ltd. (Guangdong, China) for the help in sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.725367/full#supplementary-material

REFERENCES

1. Wang L, Wang A, Wang L, Li K, Yang G, He R, et al. *In Animal Genetic Resources in China:Pigs*. China National Commission of Animal Genetic Resources, editor. China Agricultural Press (2011).p. 2–16.
2. Quan J, Yang M, Ding R, Yang L, Liu J, Liu J, et al. Genetic comparison of carcass and meat quality traits of different strain Duroc three-way cross hybrid pigs. *J South China Agric Univ.* (2016) 6:46–51. doi: 10.7671/j.issn.1001-411X.2016.06.007
3. Zhuang Z, Li S, Ding R, Yang M, Zheng E, Yang H, et al. Meta-analysis of genome-wide association studies for loin muscle area and loin muscle depth in two Duroc pig populations. *PLoS ONE.* (2019) 14:e0218263. doi: 10.1371/journal.pone.0218263
4. Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, et al. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. *BMC Genom.* (2020) 21:344. doi: 10.1186/s12864-020-6742-6
5. Ma H, Zhang S, Zhang K, Zhan H, Peng X, Xie S, et al. Identifying selection signatures for backfat thickness in yorkshire pigs highlights new regions affecting fat metabolism. *Genes-Basel.* (2019) 10:254. doi: 10.3390/genes10040254
6. Kim E, Ros-Freixedes R, Pena RN, Baas TJ, Estany J, Rothschild MF. Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations. *J Anim Sci.* (2015) 93:3292–302. doi: 10.2527/jas.2015-8879
7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:rising to the challenge of larger and richer datasets. *Gigascience.* (2015) 4:7. doi: 10.1186/s13742-015-0047-8
8. Felsenstein J. PHYLIP -phylogenetic inference package (Version 3.2). *Cladistics.* (1989) 5:164–6.
9. Rambaut A. *FigTree v1.4.3 2006–2016.* (2006). Available online at: http://tree.bio.ed.ac.uk/software (accessed 2016).
10. Ai H, Huang L, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE.* (2013) 8:e56001. doi: 10.1371/journal.pone.0056001
11. Biscarini F, Cozzi P, Gaspa G, Marras G. *detectRUNS:Detect runs of homozygosity and runs of heterozygosity in diploid genomes*. R package version 0.9.6 (2019). Available online at: https://CRAN.R-project.org/package=detectRUNS (accessed 2019).
12. McQuillan R, Leutenegger A, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet.* (2008) 83:359–72. doi: 10.1016/j.ajhg.2008.08.007
13. Madsen P, Srensen P, Su G, Damgaard L, Labouriau R. DMU - a package for analyzing multivariate mixed models. In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, Minas Gerais, Brazil (2006).p. 27–11.
14. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330
15. Szpiech ZA, Hernandez RD. selscan:an efficient multithreaded program to perform ehh-based scans for positive selection. *Mol Biol Evol.* (2014) 31:2824–7. doi: 10.1093/molbev/msu211
16. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* (2009) 2:210–23. doi: 10.1016/j.ajhg.2009.01.005
17. Hiraizumi M, Yamashita K, Nishizawa T, Nureki O. Cryo-EM structures capture the transport cycle of the P4-ATPase flippase. *Science.* (2019) 365:1149–55. doi: 10.1126/science.aay3353
18. Kogelman LJA, Pant SD, Fredholm M, Kadarmideen HN. Systems genetics of obesity in an F2 pig model by genome-wide association, genetic network, and pathway analyses. *Front Genet.* (2014) 5:214. doi: 10.3389/fgene.2014.00214
19. Newcom DW, Baas TJ, Stalder KJ, Schwab CR. Comparison of three models to estimate breeding values for percentage of loin intramuscular fat in Duroc swine1. *J Anim Sci.* (2005) 83:750–6. doi: 10.2527/2005.834750x
20. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature.* (2014) 507:371–5. doi: 10.1038/nature13138
21. Stine RR, Shapira SN, Lim H, Ishibashi J, Harms M, Won K, et al. EBF2 promotes the recruitment of beige adipocytes in white adipose tissue. *Mol Metab.* (2016) 5:57–65. doi: 10.1016/j.molmet.2015.11.001

22. Duarte DAS, Fortes MRS, Duarte MDS, Guimarães SEF, Verardo LL, Veroneze R, et al. Genome-wide association studies, meta-analyses and derived gene network for meat quality and carcass traits in pigs. *Anim Prod Sci.* (2018) 58:1100. doi: 10.1071/AN16018
23. Ma J, Yang J, Zhou L, Zhang Z, Ma H, Xie X, et al. Genome-wide association study of meat quality traits in a White DurocxErhualian F2 intercross and Chinese Sutai pigs. *PLoS ONE.* (2013) 8:e64047. doi: 10.1371/journal.pone.0064047
24. Hirose K, Mikawa S, Okumura N, Noguchi G, Fukawa K, Kanaya N, et al. Association of swine vertnin (VRTN) gene with production traits in Duroc pigs improved using a closed nucleus breeding system. *Anim Sci J.* (2013) 84:213–21. doi: 10.1111/j.1740-0929.2012.01066.x
25. Russo V, Fontanesi L, Scotti E, Beretti F, Davoli R, Nanni Costa L, et al. Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs1. *J Anim Sci.* (2008) 86:3300–14. doi: 10.2527/jas.2008-0920
26. Petrus P, Fernandez TL, Kwon MM, Huang JL, Lei V, Safikhan NS, et al. Specific loss of adipocyte CD248 improves metabolic health via reduced white adipose tissue hypoxia, fibrosis and inflammation. *Ebiomedicine.* (2019) 44:489–501. doi: 10.1016/j.ebiom.2019.05.057
27. Davoli R, Luise D, Mingazzini V, Zambonelli P, Braglia S, Serra A, et al. Genome-wide study on intramuscular fat in Italian Large White pig breed using the PorcineSNP60 BeadChip. *J Anim Breed Genet.* (2016) 133:277–82. doi: 10.1111/jbg.12189
28. Luo W, Cheng D, Chen S, Wang L, Li Y, Ma X, et al. Genome-wide association analysis of meat quality traits in a porcine large white × minzhu intercross population. *Int J Biol Sci.* (2012) 8:580–95. doi: 10.7150/ijbs.3614
29. Weisz F, Bartenschlager H, Knoll A, Mileham A, Deeb N, Geldermann H, et al. Association analyses of porcine SERPINE1 reveal sex-specific effects on muscling, growth, fat accretion and meat quality. *Anim Genet.* (2012) 43:614–9. doi: 10.1111/j.1365-2052.2011.02295.x
30. Cho I, Park H, Ahn JS, Han S, Lee J, Lim H, et al. A functional regulatory variant of MYH3 influences muscle fiber-type composition and intramuscular fat content in pigs. *PLoS Genet.* (2019) 15:e1008279. doi: 10.1371/journal.pgen.1008279
31. Ross SE, Hemati N, Longo KA, Bennett CN, Lucas PC, Erickson RL, et al. Inhibition of adipogenesis by Wnt signaling. *Science.* (2000) 5481:950–3. doi: 10.1126/science.289.5481.950
32. Mojumdar K, Giordano C, Lemaire C, Liang F, Divangahi M, Qureshi ST, et al. Divergent impact of Toll-like receptor 2 deficiency on repair mechanisms in healthy muscle versus Duchenne muscular dystrophy. *J Pathol.* (2016) 239:10–22. doi: 10.1002/path.4689
33. Knopp P, Figeac N, Fortier M, Moyle L, Zammit PS. Pitx genes are redeployed in adult myogenesis where they can act to promote myogenic differentiation in muscle satellite cells. *Dev Biol.* (2013) 377:293–304. doi: 10.1016/j.ydbio.2013.02.011
34. Sabharwal R, Chapleau MW. Autonomic, locomotor and cardiac abnormalities in a mouse model of muscular dystrophy:targeting the renin-angiotensin system. *Exp Physiol.* (2014) 99:627–31. doi: 10.1113/expphysiol.2013.074336
35. Gokhin DS, Lewis RA, McKeown CR, Nowak RB, Kim NE, Littlefield RS, et al. Tropomodulin isoforms regulate thin filament pointed-end capping and skeletal muscle physiology. *J Cell Biol.* (2010) 189:95–109. doi: 10.1083/jcb.201001125
36. Joo Y, Kim H, Lee S, Lee S. Neuronal growth regulator 1-deficient mice show increased adiposity and decreased muscle mass. *Int J Obesity.* (2019) 43:1769–82. doi: 10.1038/s41366-019-0376-2
37. Shih HP, Gross MK, Kioussi C. Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. *Proc Natl Acad Sci USA.* (2007) 104:5907–12. doi: 10.1073/pnas.0701122104
38. Zhang L, Guo Y, Wang L, Liu X, Yan H, Gao H, et al. Genomic variants associated with the number and diameter of muscle fibers in pigs as revealed by a genome-wide association study. *Animal.* (2020) 14:475–81. doi: 10.1017/S1751731119002374

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

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

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 16 September 2021
doi: 10.3389/fvets.2021.701850

Domestication and Feed Restriction Programming Organ Index, Dopamine, and Hippocampal Transcriptome Profile in Chickens

Siyu Chen[1,2], *Chao Yan*[2,3], *Jinlong Xiao*[2,3], *Wen Liu*[3], *Zhiwei Li*[3], *Hao Liu*[3], *Jian Liu*[2], *Xiben Zhang*[2], *Maojun Ou*[2], *Zelin Chen*[2], *Weibo Li*[2] *and Xingbo Zhao*[1,2,3*]

[1] *Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, Key Laboratory of Animal Molecular Design and Precise Breeding of Guangdong Higher Education Institutes, School of Life Science and Engineering, Foshan University, Foshan, China,* [2] *Guizhou Nayong Professor Workstation, China Agricultural University, Bijie, China,* [3] *College of Animal Science and Technology, China Agricultural University, Beijing, China*

OPEN ACCESS

Edited by:
Jie Mei,
Huazhong Agricultural University, China

Reviewed by:
Wenjie Guo,
Huazhong Agricultural University, China
Zhongwei Wang,
Institute of Hydrobiology, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Xingbo Zhao
zhxbcau@126.com

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

Received: *28 April 2021*
Accepted: *19 July 2021*
Published: *16 September 2021*

Citation:
Chen S, Yan C, Xiao J, Liu W, Li Z, Liu H, Liu J, Zhang X, Ou M, Chen Z, Li W and Zhao X (2021) Domestication and Feed Restriction Programming Organ Index, Dopamine, and Hippocampal Transcriptome Profile in Chickens. Front. Vet. Sci. 8:701850. doi: 10.3389/fvets.2021.701850

The domestication process exerts different phenotypic plasticity between slow- and fast-growing breeds of chicken. Feed restriction has a critical role in production performance, physiological plasticity, and stress response. Our study aimed to explore how feed restriction programed the organ index, dopamine, and hippocampal transcriptome profile between slow- and fast-growing chickens, which were fed either *ad libitum* (SA and FA), or feed restricted to 70% of *ad libitum* (SR and FR), for 30 days. Results showed that feed restriction influenced the brain organ index ($P < 0.05$), but not the organ index of the heart, liver, and spleen. The slow-growing breed tested had a higher brain organ index than the fast-growing breed ($P < 0.05$). Under feed restriction conditions, both the slow- and fast-growing breeds had significantly elevated dopamine concentrations ($P < 0.05$) compared to those fed *ad libitum*. In the GO term, upregulated genes in the FA group were enriched in the mitochondria, respiratory chain, and energy metabolism compared to the SA group ($P < 0.05$). Membranes and ribosomes were enriched in the cellular component between the SR and FR groups ($P < 0.05$). In the KEGG functional pathways, upregulated DEGs in the FR group were enriched in the cardiovascular disease category and neurodegenerative disease category compared to the FA group ($P < 0.05$). Downregulated DEGs in the FA group were enriched in the oxidative phosphorylation and neurodegenerative disease categories (Parkinson's disease and Huntington's disease) compared with the SA group ($P < 0.05$). Upregulated DEGs in the FR group were enriched in the cardiovascular disease category, neurodegenerative disease category, and energy metabolism than the SR group ($P < 0.05$). In conclusion, feed restriction had profound effects on the brain organ index and plasma dopamine in the slow- and fast-growing chickens. Feed restriction may result in issues relating to cardiovascular and neurodegenerative diseases in the fast-growing breed tested, but not in the slow-growing breed.

Keywords: domestication, feed restriction, dopamine, hippocampus, chicken

INTRODUCTION

Modern chickens, the descendants of the Red Junglefowl, have undergone basic changes, including to behavior and reproduction (1–3), as well as in brain morphology, gene expression, and DNA methylation as compared to their ancestors (4). Of which, the fast-growing breeds under intensive domestication were directly selected for meat in order to meet market demands in the past decades (5). Fast-growing chickens have achieved the intended beneficial effects to meet the needs of humans, which results in animal welfare problems and compounding of undesirable traits in response to intensive selection. For example, over-feeding and rapid growth cause cardiovascular disease, skeletal burden, and metabolic stress (5, 6), as well as immune function and parent stock management challenges in broiler chickens (7). On the other hand, most Chinese native chicken breeds are dual-purpose breeds that grow slowly under less intensive domestication processes, with divergent phenotypes from fast-growing modern breeds (8). Existing evidence has proven that slow- and fast-growing breeds have undergone phenotypic changes relating to leg muscle gene expression (9), production performance (10), as well as stress resistance (11). Notably, feed restriction programming—the restriction of nutrient intake by limiting the growth rate—was widely applied in the poultry industry. Animals under a long-term period of starvation suffer chronic stress.

Accordingly, the timing, duration, and intensity of feed restriction comprehensively influence the growth, physiological phenotypes, behavior (12, 13), and stress response of birds (14). For example, feed restriction to 90% from 5- to 11-days-old resulted in higher body weight and superior capability for meat production than feed restriction to 70% from 5- to 18-days-old, compared to feed free intake group. Likewise, early-stage (from 5 days old), high intensity (70%), and long duration (14 days) of feed restriction affected production performance and the plasma hormone in broilers (15). A previous study indicated that feed restriction can influence the liver organ index in broilers (16), but the effect is vague. For instance, 1 h of feeding and 3, 5, or 7 h feed restriction show a low to zero impact on the relative weights of broilers from the 8^{th} to 28^{th} day of age when compared to a control group (17). Feed restriction during the first 12 weeks of life decreased the density of new neurons involved in neurogenesis in the hippocampal formation but did not affect the hippocampal volume and the total number of neurons (18). The hippocampus is related to functions of emotion and reaction, and is responsible for spatial learning ability and memory in birds (19). In chickens, it is additionally responsible for neuroplasticity (20) and altering dopaminergic components of the hippocampus (21). On the other hand, dopamine, a major catecholamine neurotransmitter in the central nervous system of mammals and birds, was known to be affected by learning ability, food intake behaviors, and feed restriction (22–24), and can modulate the hippocampal synaptic plasticity (25). That is, therefore, we think that dopamine may act in a role connecting behavior and neuron activity of hippocampal transcriptome in response to feed restriction. Figuring out the biological characterization of the slow- and fast-growing breeds on the development of organs, hormone secretion, and the hippocampus is essential, and would provide a better understanding of the process of domestication and artificial selection in domesticated chickens.

The study aimed to explore the changes in organ index, plasma dopamine, and transcriptome profile of the hippocampus between slow-growing dual-purpose chickens and fast-growing broiler strains in response to feed restriction. This study will shed light on to the different breeds' biological traits, with possible uses in improving breeding strategy and feed management.

MATERIALS AND METHODS

Animals and Treatments

The experimental protocols were approved by the China Agricultural University Laboratory Animal Welfare and Animal Experimental Ethical Inspection Committee (approval number: CAU20180619-5). One hundred healthy 1-day-old female Weining dual-purpose chicks were provided by Yuansheng Animal Husbandry Co., Ltd., China This is a heritage, slow-growing (11 g/d growth rate) breed. One hundred 1-day-old female Jinlinghua broiler chicks were provided by Nanning Jinlinghua Agriculture and Animal Husbandry Group Co., Ltd., China. This is a modern, fast-growing (27 g/d growth rate) breed. All birds were reared in a brooding barn using two enclosures ($0.50 \times 0.50 \times 0.30$ m; one for each breed), respectively. The temperature was kept above 32°C from post-hatching to 16 days old. Thereafter, the temperature was gradually decreased to room temperature. At the age of 27 to 29 days, each bird was moved to a single cage ($0.19 \times 0.30 \times 0.40$ m) constructed on all sides with wire mesh. Birds were numbered with two 16 mm diameter foot rings on each leg and adapted to new environmental conditions for feed restriction program preparedness. Each cage had a feeder, drinker, perch, and droppings board. All chickens were randomly allocated to *ad libitum* or control feeding regimes to achieve a balanced sample size in each combination of breed and feeding regime. The treatments were therefore slow-growing Weining chickens fed *ad libitum* (SA, $n = 50$) or under feed restriction (SR, $n = 50$), as well as fast-growing Jinlinghua chickens fed *ad libitum* (FA, $n = 50$) or under feed restriction (FR, $n = 50$). Slow- and fast-growing chickens in the feed-restricted group were restricted from the age of 30 days to 60 days [the feed of the feed-restricted group was 70% of that of the control (15)]. The *ad libitum* measures were referred to in our previous study using the same breed and cage rearing (26). The amount of feed and the leftover feed of each chick was recorded daily.

Organ Index

At 61 days, 10 randomized birds in each group were humanely slaughtered. The weights of the body, heart, liver, spleen, and brain were immediately collected and weighted by electrical scale (quantitative analysis at 0.01 g level). Organ index was calculated by formula as follows:

$$Organ\ index = \frac{organ\ weight}{body\ weight} \times 100\%$$

TABLE 1 | Effects of feed restriction on organ index between slow- and fast-growing chickens.

Effects	Factors	Group	Heart	Liver	Spleen	Brain
Main effects	Breed	Slow-growing	0.0060 ± 0.0005	0.0155 ± 0.0021	0.0025 ± 0.0009	0.0032 ± 0.0006[a]
		Fast-growing	0.0064 ± 0.0012	0.0142 ± 0.0017	0.0024 ± 0.0010	0.0014 ± 0.0002[b]
	Treatment	*Ad libitum*	0.0062 ± 0.0009	0.0150 ± 0.0021	0.0025 ± 0.0010	0.0026 ± 0.0012[A]
		Feed restriction	0.0062 ± 0.0009	0.0148 ± 0.0019	0.0025 ± 0.0009	0.0020 ± 0.0008[B]
Interaction effect	Breed × Treatment *P*-value		0.189	0.136	0.667	0.334
Simple effects	Breed	SA	0.0059 ± 0.0005	0.0160 ± 0.0020	0.0024 ± 0.0007	0.0037 ± 0.0003
		FA	0.0065 ± 0.0010	0.0139 ± 0.0017	0.0025 ± 0.0013	0.0015 ± 0.0002
		SR	0.0061 ± 0.0005	0.0150 ± 0.0021	0.0026 ± 0.0010	0.0028 ± 0.0003
		FR	0.0062 ± 0.0013	0.0145 ± 0.0017	0.0024 ± 0.0007	0.0012 ± 0.0002
	Treatment	SA	0.0059 ± 0.0005	0.0160 ± 0.0020	0.0024 ± 0.0007	0.0037 ± 0.0003
		SR	0.0061 ± 0.0005	0.0150 ± 0.0021	0.0026 ± 0.0010	0.0028 ± 0.0003
		FA	0.0065 ± 0.0010	0.0139 ± 0.0017	0.0025 ± 0.0013	0.0015 ± 0.0002
		FR	0.0062 ± 0.0013	0.0145 ± 0.0017	0.0024 ± 0.0007	0.0012 ± 0.0002

[a,b]represents the significant difference between slow- and fast-growing breeds in the same treatment. [A,B]represents the significant difference between ad libitum vs. feed restriction in the same breed. SA, slow-growing dual-purpose chickens ad libitum; SR, slow-growing dual-purpose chickens feed restriction; FA, fast-growing broilers ad libitum; FR, fast-growing broilers feed restriction.

TABLE 2 | Effects of feed restriction on plasma dopamine concentration between slow- and fast-growing chickens.

Effects	Factors	Group	Concentrations
Main effect	Breed (pooled)	Slow-growing breed	663.7 ± 96.1
		Fast-growing breed	677.0 ± 204.5
	Treatment (pooled)	*ad libitum*	566.5 ± 111.8[B]
		Feed restriction	774.3 ± 125.0[A]
Interaction effect	Breed × Treatment		
	P-value		0.028
Simple effect	Breed	SA	605.8 ± 76.8
		FA	527.2 ± 131.8
		SR	721.7 ± 78.9
		FR	826.8 ± 144.7
	Treatment	SA	605.8 ± 76.8[B]
		SR	721.7 ± 78.9[A]
		FA	527.2 ± 131.8[B]
		FR	826.8 ± 144.7[A]

[A,B]represents the significant difference between ad libitum vs. feed restriction in the same breed. SA, slow-growing dual-purpose chickens ad libitum; SR, slow-growing dual-purpose chickens feed restriction; FA, fast-growing broilers ad libitum; FR, fast-growing broilers feed restriction.

Plasma Dopamine

At 61 days, plasma samplings were immediately collected from the abovementioned 10 birds in each group and placed in an anticoagulation tube, 4000 g centrifuge for 5 min at 4°C, then stored in 1.5 ml tubes at −20°C to prepare them for the subsequent dopamine detection. The concentrations of dopamine were then detected by an enzyme-linked immune sorbent assay kit (FU-Q411, China).

Transcriptome Profile Analysis

At 61 days, the hippocampus of eight birds was immediately collected from the abovementioned 10 birds in each group and stored in the dry ice, and then at −80°C until further processing. Total RNA was extracted from the tissue using TRIzol® Reagent (Invitrogen) according to the manufacturer's instructions, and genomic DNA was removed using DNase I (TaKara). Then RNA quality was determined by a 2100 Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). Only high-quality RNA samples were used to construct a sequencing library.

RNA-seq transcriptome library was prepared following TruSeq™ RNA sample preparation Kit from Illumina (San Diego, CA) using 1 μg of total RNA. Shortly, messenger RNA was isolated according to poly-A selection method by oligo (dT) beads and then fragmented by a fragmentation buffer. Next, double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) with random hexamer primers (Illumina). Then the synthesized cDNA was subjected to end-repair, phosphorylation, and "A" base addition according to Illumina's library construction protocol. Libraries were size selected for cDNA target fragments of 200-300 bp on 2% Low Range Ultra Agarose, followed by PCR amplified using Phusion DNA polymerase (NEB) for 15 PCR cycles. After being quantified by TBS380, the paired-end RNA-seq sequencing library was sequenced with the Illumina HiSeq X ten/NovaSeq 6000 sequencer (2 × 150 bp read length).

The raw paired-end reads were trimmed and quality controlled by SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle), with default parameters. Then clean reads were separately aligned to the reference genome with the orientation mode using TopHat (http://tophat.cbcb.umd.edu/, version2.0.0) (27) software. The mapping criteria used in Bowtie were as follows: sequencing reads should be uniquely matched to the genome allowing up to two mismatches, without insertions or deletions. Then, the region of the gene was explored according to different site depths and the operon was obtained. In addition, the whole genome was

split into multiple 15k bp windows that each share an overlap of 5k bp. New transcribed regions were defined as more than two consecutive windows without the overlapped region of the gene, where at least two reads mapped per window were in the same orientation.

To identify DEGs (differentially expressed genes) between two different samples, the expression level of each transcript was calculated according to the fragments per kilobase of exon per million mapped reads (FRKM) method. RSEM (28) was used to quantify gene abundances. R statistical package software EdgeR (Empirical Analysis of Digital Gene Expression in R) (29) was utilized for differential expression analysis. In addition, functional-enrichment analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), was performed to identify which DEGs were significantly enriched in GO terms and metabolic pathways at Bonferroni-corrected P-value ≤ 0.05 compared with the whole-transcriptome background. GO functional enrichment and KEGG pathway analysis were carried out by Goatools (https://github.com/tanghaibao/Goatools) and KOBAS (http://kobas.cbi.pku.edu.cn/home.do) (30).

Statistical Analysis

All data were analyzed and tested throughout IBM SPSS Statistics 21. The organ index has not met the assumptions for parametric analysis, and therefore has been analyzed via a non-parametrical method. Two-way non-parametrical ANOVA analysis, specifically the Scheirer-Ray-Hare test, was used for statistical analysis, and then the Kruskal-Wallis test (K-W test) was used for pairwise analysis. The main effects were analyzed when they showed no significant or interaction effects, whereas the simple effects were analyzed when the main effects and interaction effects were both significant. Plasma dopamine concentration was checked for normality and homogeneity of variance and meets the assumptions for parametric analysis, and was analyzed using a two-way ANOVA. The Duncan test was used in the Postdoc testing. All values with $P < 0.05$ were regarded as statistically significant.

RESULTS

Organ Index

Organ indexes are shown in **Table 1**. The brain organ index was affected by both breed and feeding regime, but not by the interaction of breed and treatment. The brain organ index was higher in slow-growing dual-purpose chickens than in the fast-growing broiler breed, and higher in the *ad libitum* group than the feed restricted group ($P < 0.05$). The organ index of the heart, liver, and spleen showed no difference between slow- and fast-growing breeds, or between *ad libitum* and feed restriction treatments.

Plasma Dopamine

The plasma dopamine was affected by feed restriction and the interaction of breed and treatment ($P < 0.05$, **Table 2**), but not by breed alone. In the feed restricted group, the concentration was higher than that of the *ad libitum* group ($P < 0.05$, **Table 2**). The dopamine concentration in the SR group and FR group was higher than that in the SA group and the FA group, accordingly ($P < 0.05$, **Table 2**).

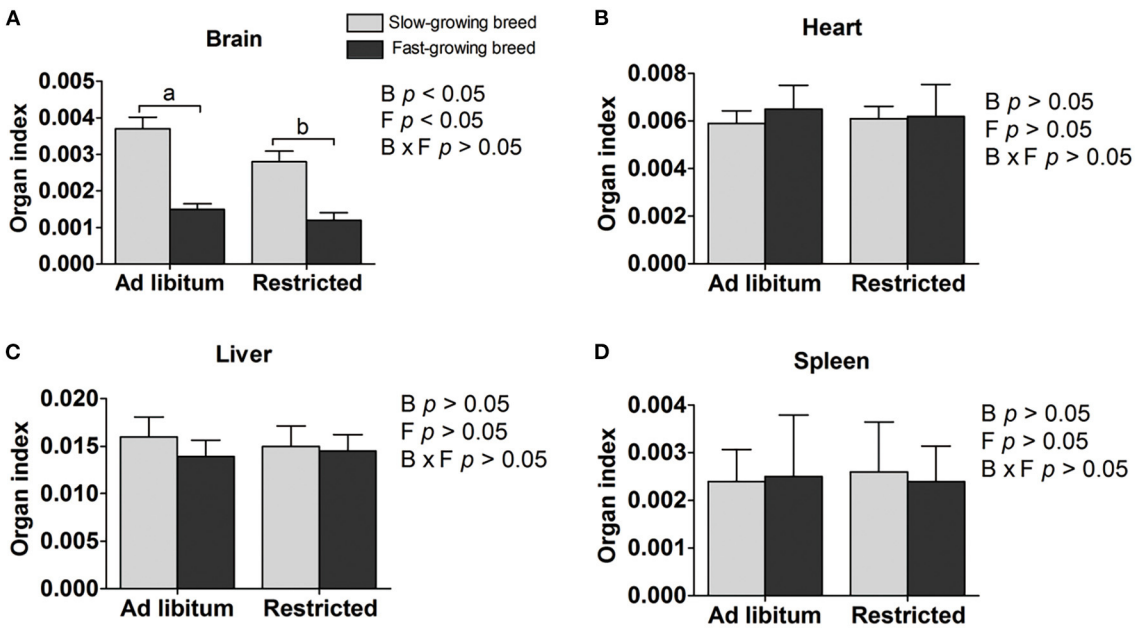

FIGURE 1 | Organ indexes of brain **(A)**, heart **(B)**, liver **(C)**, and spleen **(D)** between slow-growing and fast-growing breed birds. SA, slow-growing dual-purpose chickens *ad libitum*; SR, slow-growing dual-purpose chickens feed restriction; FA: fast-growing broilers *ad libitum*; FR, fast growing broilers feed restriction. a, b means the significant difference between ad libitum vs feed restriction in the same breed.

Hippocampal Transcriptome Profile Analysis

The sequenced data (clean reads) after quality control were compared with the reference genome, and the data obtained from each group were averaged. It can be seen that the mapping rate between the sequenced data and the reference genome was 88.67%.

After the feed restriction treatment, the Venn diagram formed by the data processing of each group is shown below (**Figure 1**).

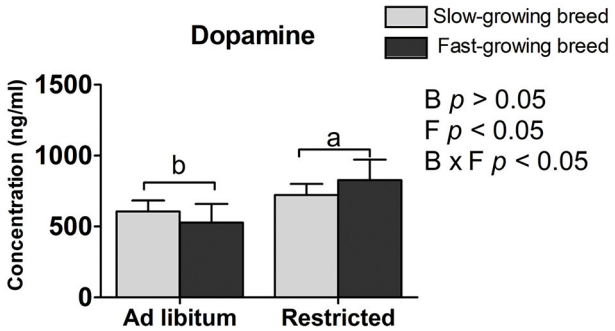

FIGURE 2 | Differently expressed genes between groups. The discrepancy was more significant with |log2 fold change| < 1 and adjusted *P*-value < 0.05. SA, slow-growing dual-purpose chickens *ad libitum*; SR, slow-growing dual-purpose chickens feed restriction; FA, fast-growing broilers *ad libitum*; FR, fast-growing broilers feed restriction. a, b means the significant difference between *ad libitum* vs feed restriction in the same breed.

The SA and SR groups had 11930 genes co-expressed, and 967 and 135 genes specifically expressed. The FA and FR groups had 13,383 co-expressed genes, and 172 and 768 genes specifically expressed. The SA and FA groups have 12,665 genes co-expressed, and 232 and 890 genes specifically expressed. The SR and FR groups had 11,828 genes co-expressed, and 238 and 2,324 genes specifically expressed.

Differently Expressed Genes

After the feed restriction treatment, the gene expression levels of each group were significantly adjusted up and down as shown below (|log2 fold change| < 1, and adjusted $P < 0.05$) (**Figure 2**). Compared with the SA group, the SR group was upregulated in 84 genes and downregulated in 62 genes. As compared with the FA group, the FR group was upregulated in 701 genes and downregulated in 521 genes. Compared with the SA group, the FA group was upregulated 475 genes and downregulated 964 genes. As compared with the SR group, the FR group was upregulated in 2,647 genes and downregulated in 2,389 genes.

Gene Ontology Enrichment Analysis

The GO enrichment analysis results of the significant up-down genes between the SA and FA groups are shown in **Figure 3**. The DEGs in the SA and FA groups were significantly enriched in pathways related to the mitochondrion and respiratory electron transport chains. With respect to the biological functions: respiratory electron transport chains, electron transport chains, ATP synthesis coupled electron transport, proton transport, hydrogen transport, and generation of precursor metabolites

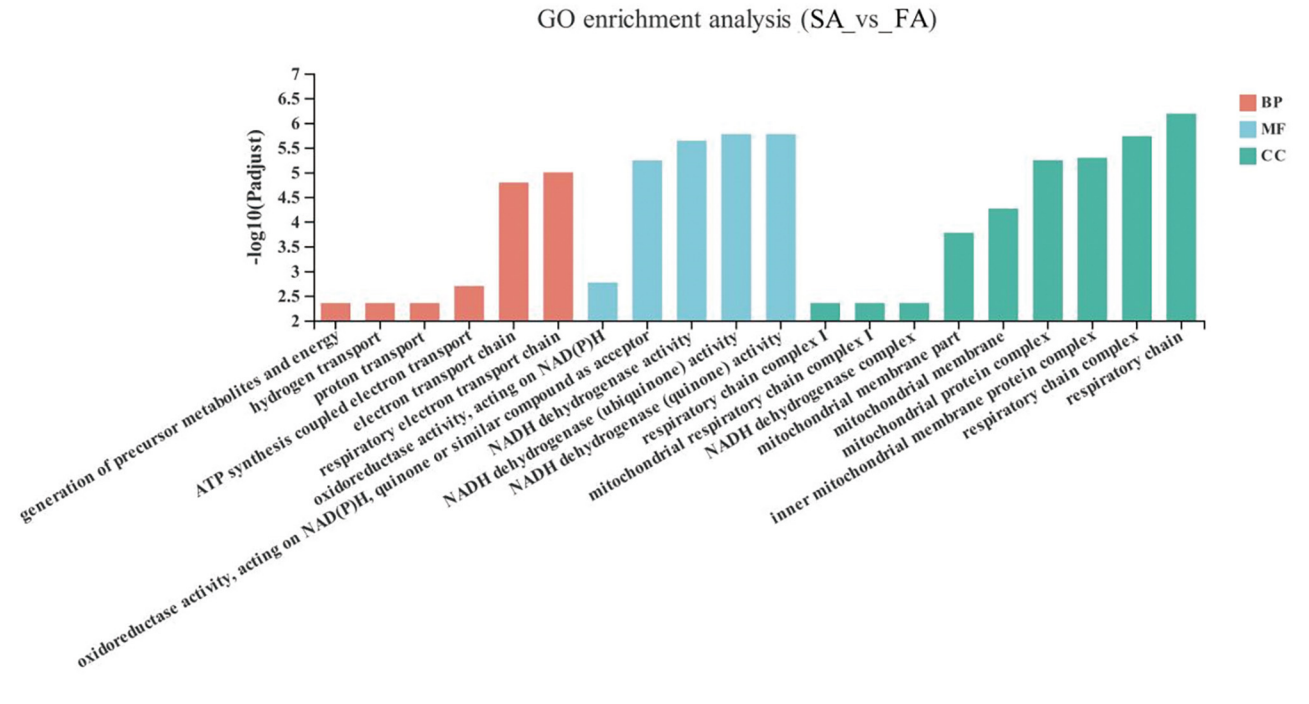

FIGURE 3 | GO terms of differentially expressed genes (DEGs) between SA and FA groups. SA, slow-growing dual-purpose chickens *ad libitum*; FA, fast-growing broilers *ad libitum*.

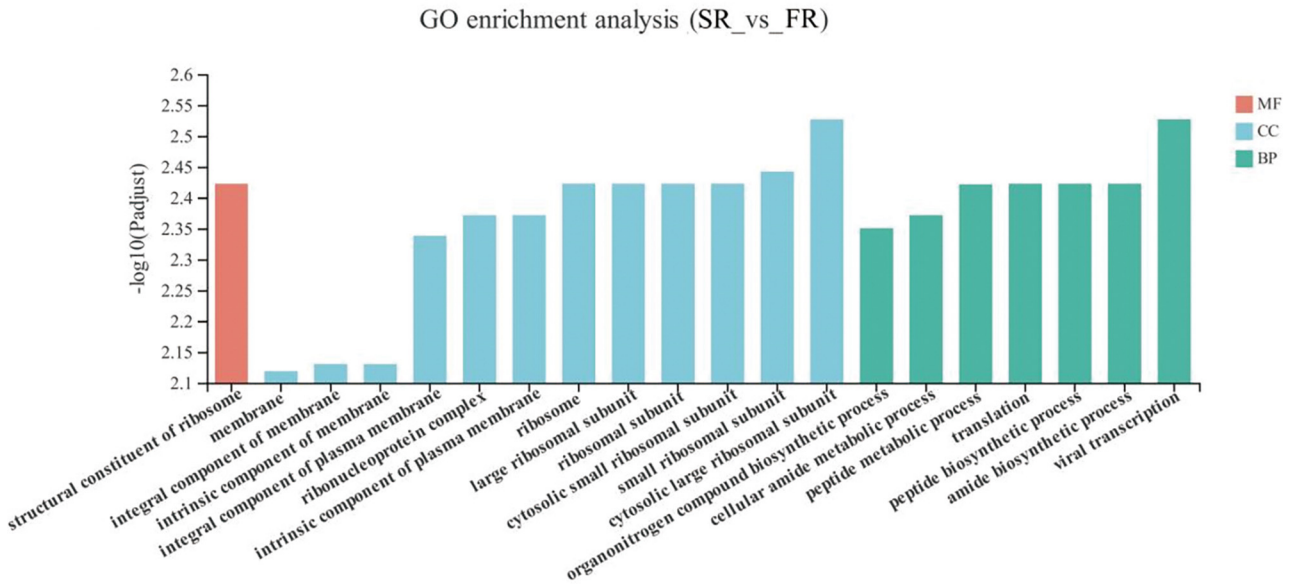

FIGURE 4 | GO terms of differentially expressed genes (DEGs) between SR and FR groups. SR, slow-growing dual-purpose chickens feed restriction; FR, fast-growing broilers feed restriction.

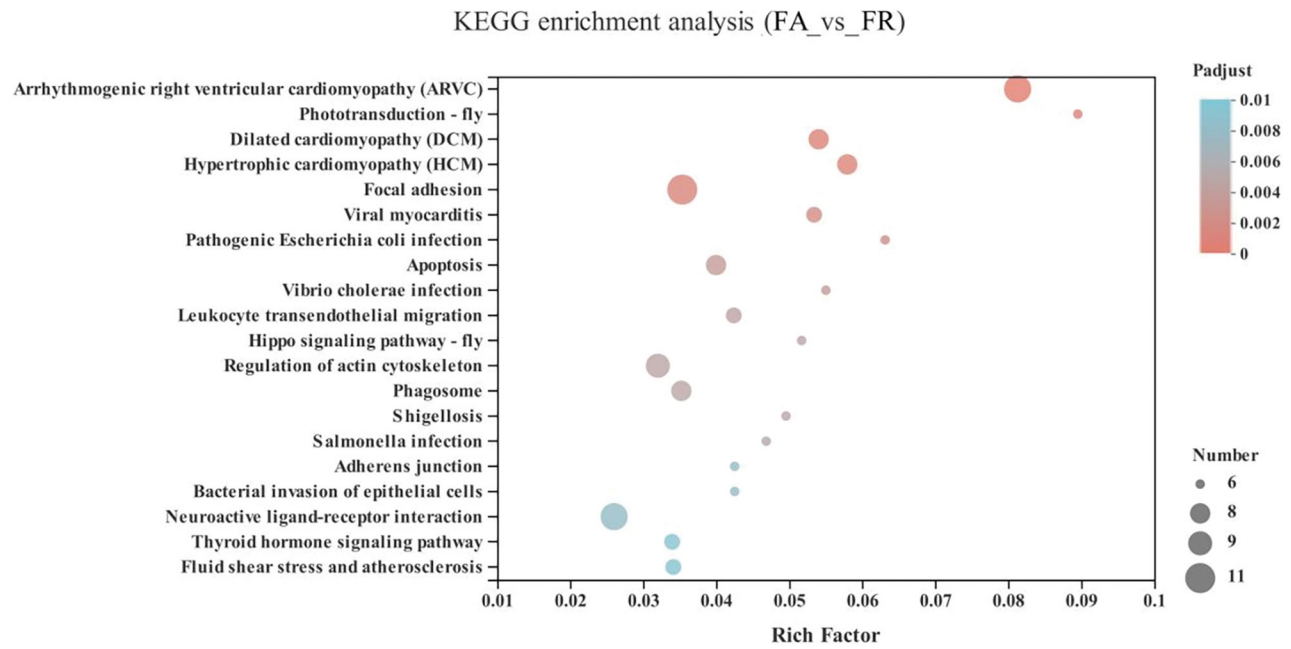

FIGURE 5 | KEGG pathway enrichment analysis of differentially expressed genes (DEGs) between SA and FR groups. SA, slow-growing dual-purpose chickens *ad libitum*; FR, fast-growing broilers feed restriction.

and energy were enriched between groups ($P < 0.05$). For the molecular function: NADH dehydrogenase (quinone) activity; NADH dehydrogenase (ubiquinone) activity; NADH dehydrogenase activity; oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor; and oxidoreductase activity, acting on NAD(P)H were found to be significantly enriched ($P < 0.05$). Regarding the cellular component: respiratory chain, respiratory chain complex, inner mitochondrial membrane protein complex, mitochondrial protein complex, mitochondrial membrane, mitochondrial

membrane part, NADH dehydrogenase complex, mitochondrial respiratory chain complex I, and respiratory chain complex I were differently enriched ($P < 0.05$).

The DEGs between the SR group and the FR group enriched in GO as displayed in **Figure 4**. With respect to biological function, DEGs were mainly enriched in pathways related to Amide biosynthetic process and cellular amide metabolic process ($P < 0.05$). The structural constituent of ribosome was enriched in the molecular function ($P < 0.05$). Membrane and ribosomes were enriched in the cellular component ($P < 0.05$).

Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis

Compared with the FA group, the significantly upregulated genes in the FR group were mainly enriched in the cardiovascular diseases category, including arrhythmogenic right ventricular cardiomyopathy (ARVC) (10 genes), dilated cardiomyopathy (DCM) (8 genes), hypertrophic cardiomyopathy (HCM) (8 genes), viral myocarditis (7 genes), fluid shear stress and atherosclerosis (7 genes), as well as neurodegenerative diseases, including neuroactive ligand-receptor interaction (*GALR1*, *HTR5A*, *GLP1R*, *VIPR1*, *NPB*, ENSGALG00000038742, ENSGALG00000039810, ENSGALG00000045754, ENSGALG00000040736, and ENSGALG00000040953) ($P < 0.05$) (**Figure 5** and **Supplementary Table 1**).

In the energy metabolism category of oxidative phosphorylation (*MT-ND3*, *MT-ND2*, *MT-CO1*, *ND6*, *MT-ND4L*, *MT-ATP8*, *MT-ND4*, *ND5*, *COX3*, *MT-CYB*, *ND1*, *ATP6V1G3*, *MT-CO2*, and *ATP6*), the neurodegenerative diseases category, including Parkinson disease (*MT-ND3*, *MT-ND2*, *MT-CO1*, *ND6*, *MT-ND4L*, *MT-ND4*, *ND5*, *COX3*, *MT-CYB*, *ND1*, *MT-CO2*, and *ATP6*) and Huntington disease, and the retrograde category of the nervous system related to endocannabinoid signaling, were enriched in the SA and FA groups ($P < 0.05$) (**Figure 6** and **Supplementary Table 1**).

Compared with the SR group, the FR group was enriched in pathways including arrhythmogenic right ventricular cardiomyopathy (ARVC) (46 genes), dilated cardiomyopathy (DCM) (49 genes), hypertrophic cardiomyopathy (HCM) (46 genes), viral myocarditis (44 genes), and fluid shear stress and atherosclerosis (57 genes), plus in neurodegenerative diseases category, including Parkinson disease (50 genes) and Alzheimer disease (51 genes), and lastly in the energy metabolism category of oxidative phosphorylation (43 genes) ($P < 0.05$) (**Figure 7** and **Supplementary Table 1**).

DISCUSSION

Organs are the basic "facilities" of animal life processes and the material basis of their physiological functions, which directly affect the speed of weight gain, the health status of the animal, and even the adaptation to their environment of chickens (31). Organ index is a biological characteristic index, and the index size can reflect the workload of the organ in the organism to a certain extent (16, 17). In our study, the brain organ index showed sd higher in the *ad libitum* groups than the feed-restricted groups; these results are similar to the findings of a review which concluded that stress can reduce the volume of the brain during times of adversity in humans (32). Previous studies indicate that the highest relative liver weight was observed in the birds fed *ad libitum* compared to the feed restricted group when withdrawing the feed for 7 hours from 10 days to 30 days of age in broilers (16). The relative weight of the heart, liver, spleen, gizzard, pancreas, and intestine remained unaffected by the feed-restricted treatments from 8 to 28 days of age (17). However, feed restriction to 70% for a duration of 30 days had a limited effect on the heart, liver, and spleen organ index between slow- and fast-growing breeds in our study. Furthermore, the brain organ index displayed a greater increase in slow-growing dual-purpose chickens than in fast-growing broiler breeds, which might be supported by previous findings that artificial selection has altered the internal morphology of various animals, such as causing an overall decrease in brain size in mink (33). Besides, artificial selection has selected for large body size and pectoral muscles, rapid growth, and increased relative gut length in fast-growing broilers (34). The difference in brain organ index may be attributed to the process of intensive breeding and selection programs, which has been selected for a reduction of brain size (2, 35). In Red junglefowl, the brain size is reduced in response to artificial selection, and it is a tradeoff for other physiological traits, such as reproduction and growth (36).

The dopamine concentration in the feed restricted group was higher than the *ad libitum* group. Our result may be supported by that previous finding that the intracerebroventricular injection of dopamine decreases food intake in chickens (37), and further confirms the association of dopamine with feed intake (22, 23). One reason may be that feed restriction stimulates a greater state of excitement in the birds when feeding owing to their greater hunger. The mechanism is similar to findings that dopamine hyperfunction provides an abnormal driving force in afflicted patients that causes mental excitement (38). To some extent, increased dopamine levels are associated with social anxiety disorder (39, 40). That is, too much dopamine and dopamine hyperfunction lead to irritability, which results in emotional dysfunction, nerve reflex rapid reaction, and extreme hyperactivity (41). In addition, dopamine was influenced by the interaction of breed and treatment. Similarly, the higher dopamine varieties have other functions affecting the process of hyperfunction and hyperactivity, which may be responsible for the physiological process between the SR and SA groups, as well as the FR and FA groups.

When it comes to gene expression of the hippocampus, none of DEGs were enriched in the GO term and KEGG pathways. A previous study indicated that a 12-week duration of feed restriction reduces hippocampal neurogenesis and causes potential chronic stress, but is not of consequence to health outcomes in birds (18). Feed restriction in our study lasting 4 weeks has shown significant impacts on the hippocampal transcriptome, and a greater breed effect in the fast-growing breed than slow-growing breed. The greater number of DEGs, GO terms and functional KEGG pathways may implicate the increased plasticity of gene expressions in the fast-growing breed

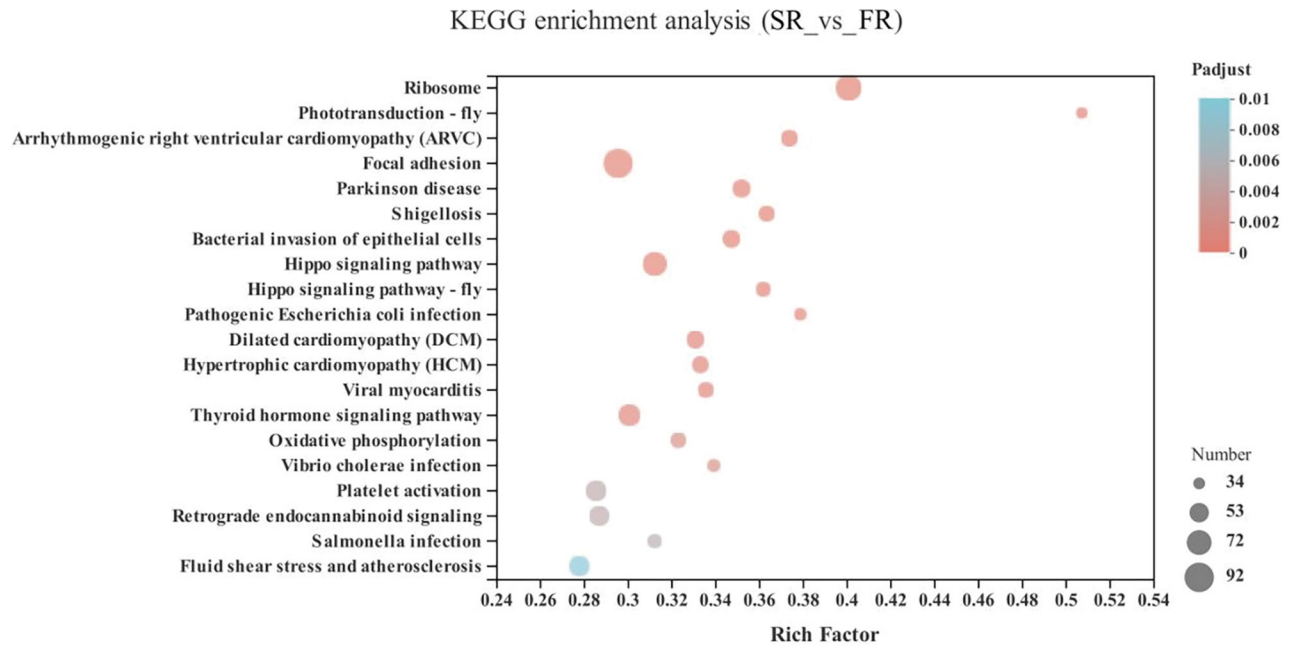

FIGURE 6 | KEGG pathway enrichment analysis of differentially expressed genes (DEGs) between SA and FA groups. SA, slow-growing dual-purpose chickens *ad libitum*; FA, fast-growing broilers *ad libitum*.

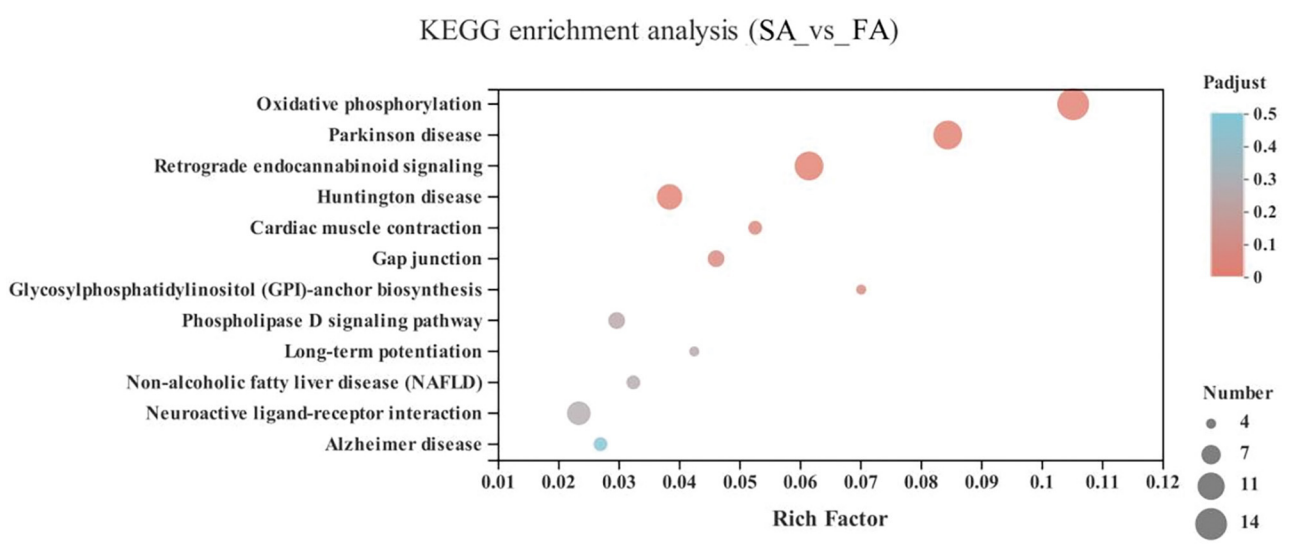

FIGURE 7 | KEGG pathway enrichment analysis of differentially expressed genes (DEGs) between SR and FR groups. SR, slow-growing dual-purpose chickens feed restriction; FR, fast-growing broilers feed restriction.

compared with the slow-growing breed in response to feed restriction (11). Prior studies showed that slow- and fast-growing breeds had a different transcriptome profile in the breast muscle (42) and leg muscle (9), which may mirror the transcriptome profile between breeds in our study. Thus, the gene expression pattern is greatly affected due to feed restriction, which has also demonstrated that the hippocampus is a sensitive area easily influenced by external stimuli (18).

Compared with the FA group, the significantly upregulated genes in the FR group were mainly enriched in the cardiovascular disease category, including arrhythmogenic right ventricular cardiomyopathy (ARVC), dilated cardiomyopathy (DCM),

hypertrophic cardiomyopathy (HCM), and viral myocarditis. Accordingly, fast-growing broilers are highly susceptible to stress-induced cardiac arrhythmia (43). Cardiovascular diseases may cause high mortality rates, recognized as sudden death syndrome (44). In addition, dopamine is a precursor to norepinephrine in noradrenergic nerves and in certain areas of the central nervous system involved in the cardiovascular system (45), which may support the higher concentration of dopamine in the FR group than the FA group in response to cardiovascular function. Also, DEGs in the FR group were enriched in the neuroactive ligand-receptor interaction pathway of the neurodegenerative disease category. A previous study indicated that feed restriction can decrease the density of neurogenesis in hippocampal formation in broilers (18). Previous results and our own seem to suggest that feed restriction is a detriment for neural-related development. Also, dopamine is related to hippocampal synaptic plasticity (25), cognitive functions (e.g., episodic memory, speed, fluency), and facilitating the responsivity of divergent neural networks (46). Thus, the higher dopamine concentration in the FR group than the FA group may be connected to the neural-related development in the hippocampus in fast-growing broilers.

The upregulated DEGs of FA were enriched in Parkinson's disease and Huntington's disease compared to the SA group. It is possible that their accelerated breeding speed gives the fast-growing broilers increased potential to develop neural and pharmacological problems (4). Besides, oxidative phosphorylation in the energy metabolism category in the FA group was higher than in the SA group. Oxidative phosphorylation is associated with the oxidoreductase chain in the mitochondria, which implicated that slow- and fast-growing breeds had different energy metabolism abilities and trade-offs in energy allocation (2, 31). Additionally, the GO terms, biological function, molecular function, and cellular component process were enriched in the mitochondria, respiratory chain, and energy metabolism. It is accepted that differences in metabolic rate, oxidative stress ability, and energy metabolism arise in response to the different purposes of artificial selection (11, 31).

As when we compared FR to FA, the cardiovascular diseases category, including arrhythmogenic right ventricular cardiomyopathy (ARVC), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), viral myocarditis, fluid shear stress, and atherosclerosis pathways, were found between the comparisons of FR to SR. Thus, it seems that feed restriction is likely linked to cardiovascular diseases, illustrating that feed restriction to 70% for 30 days is more detrimental in the fast-growing breed than the slow-growing breed. Similarly, more DEGs were enriched in the Parkinson's disease and Alzheimer's disease areas of the neurodegenerative diseases category. The hippocampus is a sensitive brain area, which is influenced by Alzheimer's disease, epilepsy, and depression (47). Parkinson's disease and Alzheimer's disease were critically implicated by dopamine (48). Unexpectedly, the dopamine did not differ between SR and FR groups, for which we speculate that a significant difference in dopamine levels may not necessarily lead to those diseases, and this requires further study in different breeds. In the GO terms, it seemed that DEGs were enriched in the membrane and ribosome-related processes; the reason for this was unclear and requires further investigation. All in all, we conclude that the fast-growing breed is more susceptible to the detrimental effects of feed restriction. Further study should be focused on the validation of the key genes involved and the role of these genes on related pathways. Besides, the mechanism of dopamine in connecting feeding behavior and genetic aspects requires further investigation, especially for the classification of the role of dopamine on feeding behavior and emotion.

CONCLUSION

Brain organ index was affected by feed restriction and breed. The Feed restricted group had greater dopamine hyperactivity than the *ad libitum* group in both slow- and fast-growing breeds. Differently expressed genes were enriched in the cardiovascular disease and neurodegenerative disease categories in the fast-growing breed, suggesting feed restriction to 70% for 30 days is a disadvantage for the fast-growing breed. Feed restriction had less effect on the hippocampal transcriptome profile in the slow-growing breed.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession numbers can be found below: https://www.ncbi.nlm.nih.gov/sra/, PRJNA745252.

ETHICS STATEMENT

The animal study was reviewed and approved by China Agricultural University Laboratory Animal Welfare and Animal Experimental Ethical Inspection Committee (approval number: CAU20180619-5).

AUTHOR CONTRIBUTIONS

XZhao obtained the funding. CY, SC, and XZhao designed this project. CY, SC, WLiu, JX, ZL, HL, JL, XZhang, MO, ZC, and WLi performed the experiment. CY, WLiu, and SC analyzed and interpreted the data. CY, SC, and XZhao drafted and revised the manuscript. All authors agreed upon publication.

FUNDING

This study received funding from the Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding (2019B030301010), the Key Laboratory of Animal Molecular Design and Precise Breeding of Guangdong Higher Education Institutes (2019KSYS011), the Joint Fund of Basic and Applied Basic Research Fund of Guangdong Province (Number: 2019A1515110598), and the Joint Projects of GuizhouNayong Professor Workstation (Number: 201705510410352). The funder was not involved in the study design, collection, analysis, interpretation of

data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

We express our sincere gratitude to the staff from Guizhou Nayong Yuanshengmuye Ltd., Bijie, 553300, Guizhou, China, for the support during the study and for providing the Weining chicks. We sincerely appreciate the help from Rong He and all persons in Qianlong organic farm for the supply of the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.701850/full#supplementary-material

REFERENCES

1. Schütz KE, Forkman BR, Jensen P. Domestication effects on foraging strategy, social behaviour and different fear responses: a comparison between the red junglefowl (Gallus gallus) and a modern layer strain. *Appl Anim Behav Sci.* (2001) 74:1–14. doi: 10.1016/S0168-1591(01)00156-3
2. Al-Nasser A, Al-Khalaifa H, Al-Saffar A, Khalil F, Albahouh M, Ragheb G, et al. Overview of chicken taxonomy and domestication. *Worlds Poultry Science Journal.* (2007) 63:285–300. doi: 10.1017/S004393390700147X
3. Lindqvist C, Lind J, Jensen P. Effects of domestication on food deprivation-induced behaviour in red junglefowl, Gallus gallus, and White Leghorn layers. *Anim Behav.* (2009) 77:893–9. doi: 10.1016/j.anbehav.2008.12.015
4. Bélteky J. *Chicken domestication: effects of tameness on brain gene expression and DNA methylation* (Dissertation). Linköping University, Linköping, Sweden (2016).
5. Schmidt CJ, Persia ME, Feierstein E, Kingham B, Saylor WW. Comparison of a modern broiler line and a heritage line unselected since the 1950s. *Poult Sci.* (2009) 88:2610–9. doi: 10.3382/ps.2009-00055
6. Mench JA. Broiler breeders: feed restriction and welfare. *Worlds Poult Sci J.* (2002) 58:23–9. doi: 10.1079/WPS20020004
7. Zuidhof MJ, Schneider BL, Carney VL, Korver DR, Robinson FE. Growth, efficiency, and yield of commercial broilers from 1957, 1978, and 2005. *Poult Sci.* (2014) 93:2970–82. doi: 10.3382/ps.2014-04291
8. Chen GH, Wang KH, Wang JY, Ding CNY. *Poultry Genetic Resources in China*. 1st edn. Shanghai: Shanghai Scientific and Technological Press (2004).
9. Wu P, Dai G, Chen F, Chen L, Zhang T, Xie K, et al. Transcriptome profile analysis of leg muscle tissues between slow- and fast-growing chickens. *PLoS ONE.* (2018) 13:e0206131. doi: 10.1371/journal.pone.0206131
10. Fanatico AC, Pillai PB, Hester PY, Falcone C, Mench JA, Owens CM, et al. Performance, livability, and carcass yield of slow- and fast-growing chicken genotypes fed low-nutrient or standard diets and raised indoors or with outdoor access. *Poult. Sci.* (2008) 87:1012–21. doi: 10.3382/ps.2006-00424
11. Mattioli S, Dal Bosco A, Ruggeri S, Martino M, Moscati L, Pesca C, et al. Adaptive response to exercise of fast-growing and slow-growing chicken strains: blood oxidative status and non-enzymatic antioxidant defense. *Poult Sci.* (2017) 96:4096–102. doi: 10.3382/ps/pex203
12. Rajman M, Jurani M, Lamosova D, Macajova M, Sedlackova M, Kost'al L, et al. The effects of feed restriction on plasma biochemistry in growing meat type chickens (Gallus gallus). *Comp Biochem Physiol A Mol Integr Physiol.* (2006) 145:363–71. doi: 10.1016/j.cbpa.2006.07.004
13. Yu MW, Robinson FE. The application of short-term feed restriction to broiler chicken production: a review. *J Appl Poul Res.* (1992) 1:147–53. doi: 10.1093/japr/1.1.147
14. Alexander S, Kitaysky EV, Evgenia V, Kitaiskaia, Wingfield JC, Piatt JF. Dietary restriction causes chronic elevation of corticosterone and enhances stress response in red-legged kittiwake chicks. *J Comp PhysiolB Biochem Syst Environ Physiol.* (2001) 171:701–9. doi: 10.1007/s003600100230
15. Li L, Zhao G, Ren Z, Duan L, Zheng H, Wang J, et al. Effects of early feed restriction programs on production performance and hormone level in plasma of broiler chickens. *Front Agric China.* (2011) 5:94–101. doi: 10.1007/s11703-010-1066-y
16. Zubair AK, Leeson S. Effect of Early Feed Restriction and Realimentation on Heat Production and Changes in Sizes of Digestive Organs of Male Broilers. *Poul Sci.* (1994) 73:529–38. doi: 10.3382/ps.0730529
17. Mahmood S, Mehmood S, Ahmad F, Masood A, Kausar R. Effects of feed restriction during starter phase on subsequent growth performance, dressing percentage, relative organ weights and immune response of broilers. *Pak Vet J.* (2007) 27:137–41.
18. Robertson BA, Rathbone L, Cirillo G, D'Eath RB, Bateson M, Boswell T, et al. Food restriction reduces neurogenesis in the avian hippocampal formation. *PLoS. ONE.* (2017) 12:e0189158. doi: 10.1371/journal.pone.0189158
19. Székely AD. The avian hippocampal formation: subdivisions and connectivity. *Behav Brain Res.* (1999) 98:225. doi: 10.1016/S0166-4328(98)00087-4
20. Freire R, Cheng HW. Experience-dependent changes in the hippocampus of domestic chicks: a model for spatial memory. *Eur J Neurosci.* (2004) 20:1065–8. doi: 10.1111/j.1460-9568.2004.03545.x
21. Patzke N, Ocklenburg S, van der Staay FJ, Gunturkun O, Manns M. Consequences of different housing conditions on brain morphology in laying hens. *J Chem Neuroanat.* (2009) 37:141–8. doi: 10.1016/j.jchemneu.2008.12.005
22. Zendehdel M, Hassanpour S. Central regulation of food intake in mammals and birds: a review. *Neurotransmitter.* (2014) 1:1–7. doi: 10.14800/nt.251
23. Zendehdel M, Moosadoost Y, Masoumi R, Rostami B, Shahir MH, Hassanpour S. Endogenous nitric oxide and dopamine regulates feeding behavior in neonatal layer-type chicken. *Ann Anim Sci.* (2017) 17:1029–42. doi: 10.1515/aoas-2016-0094
24. Kostál L, Vboh P, Savory CJ, Juráni M, Blazícek P. Influence of food restriction on dopamine receptor densities, catecholamine concentrations and dopamine turnover in chicken brain. *Neuroscience.* (1999) 94:323–8. doi: 10.1016/S0306-4522(99)00255-9
25. Li S, Cullen WK, Anwyl R, Rowan MJ. Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. *Nat Neurosci.* (2003) 526–31. doi: 10.1038/nn1049
26. Chen S, Yan C, Xiang H, Xiao J, Liu J, Zhang H, et al. Transcriptome changes underlie alterations in behavioral traits in different types of chicken. *J Anim Sci.* (2020) 98:6. doi: 10.1093/jas/skaa167
27. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* (2009) 25:1105–11. doi: 10.1093/bioinformatics/btp120
28. LiColin B, Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* (2011) 12:323. doi: 10.1186/1471-2105-12-323
29. Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139. doi: 10.1093/bioinformatics/btp616
30. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* (2011) 39:316–22. doi: 10.1093/nar/gkr483
31. Konarzewski M, Gavin A, Mcdevitt R, Wallis RI. Metabolic and organ mass responses to selection for high growth rates in the domestic chicken (Gallus domesticus). *Physiol Biochem Zool Pbz.* (2000) 73:237–48. doi: 10.1086/316729
32. Lupien SJ, Mcewen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci.* (2009) 10:434–45. doi: 10.1038/nrn2639
33. Kruska D. The effect of domestication on brain size and composition in the mink (Mustela vison). *J Zool.* (1996) 239:645–61. doi: 10.1111/j.1469-7998.1996.tb05468.x

34. Diamond JJ. Metabolic and digestive responses to artificial selection in chickens. *Evolution.* (1996) 50:1638–50. doi: 10.1111/j.1558-5646.1996.tb03936.x
35. Agnvall B, Bélteky J, Jensen P. Brain size is reduced by selection for tameness in Red Junglefowl– correlated effects in vital organs. *Sci Rep.* (2017) 7:3306. doi: 10.1038/s41598-017-03236-4
36. Isler K, van Schaik CP. Metabolic costs of brain size evolution. *Biol Lett.* (2006) 2:557–60. doi: 10.1098/rsbl.2006.0538
37. Zendehdel M, Hasani K, Babapour V, Mortezaei SS, Khoshbakht Y, Hassanpour S. Dopamine-induced hypophagia is mediated by D1 and 5HT-2c receptors in chicken. *Vet Res Commun.* (2014) 38:11–19. doi: 10.1007/s11259-013-9581-y
38. Hamner MB, Diamond BI. Plasma dopamine and norepinephrine correlations with psychomotor retardation, anxiety, and depression in non-psychotic depressed patients: a pilot study. *Psychiatry Res.* (1996) 64:209–11. doi: 10.1016/S0165-1781(96)02879-X
39. Zarrindast MR, Khakpai F. The modulatory role of dopamine in anxiety-like behavior. *Arch Iran Med.* (2015) 18:591–603.
40. Leblanc J, Ducharme MB. Plasma dopamine and noradrenaline variations in response to stress. *Physiol Behav.* (2007) 91:208–11. doi: 10.1016/j.physbeh.2007.02.011
41. Viggiano D, Vallone D, Ruocco LA, Sadile AG. Behavioural, pharmacological, morpho-functional molecular studies reveal a hyperfunctioning mesocortical dopamine system in an animal model of attention deficit and hyperactivity disorder. *Neurosci Biobehav Rev.* (2003) 27:683–9. doi: 10.1016/j.neubiorev.2003.08.011
42. Davis RVN, Lamont SJ, Rothschild MF, Persia ME, Ashwell CM, Schmidt CJ. Transcriptome analysis of post-hatch breast muscle in legacy and modern broiler chickens reveals enrichment of several regulators of myogenic growth. *PLoS. ONE.* (2015) 10:e0122525. doi: 10.1371/journal.pone.0122525
43. Olkowski AA, Wojnarowicz C, Nain S, Ling B, Alcorn JM, Laarveld B. A study on pathogenesis of sudden death syndrome in broiler chickens. *Res Vet Sci.* (2008) 85:140. doi: 10.1016/j.rvsc.2007.08.006
44. Olkowski AA, Nain S, Wojnarowicz C, Laarveld B, Alcorn J, Ling BB. Comparative study of myocardial high energy phosphate substrate content in slow and fast growing chicken and in chickens with heart failure and ascites. *Comp Biochem Physiol A Mol Integr Physiol.* (2007) 148:230–8. doi: 10.1016/j.cbpa.2007.04.015
45. Amenta F. Density and distribution of dopamine receptors in the cardiovascular system and in the kidney. *Auton Autacoid Pharmacol.* (1990) 10:s11–8. doi: 10.1111/j.1474-8673.1990.tb00222.x
46. Bäckman L, Nyberg L, Lindenberger U, Li SC, Farde L. The correlative triad among aging, dopamine and cognition: current status and future prospects. *Neurosci Biobehav Rev.* (2006) 30:791–807. doi: 10.1016/j.neubiorev.2006.06.005
47. Dhikav V, Anand KS. Hippocampus in health and disease: an overview. *Ann Indian Acad Neurol.* (2012) 15:239–46. doi: 10.4103/0972-2327.104323
48. Li J, Zhu M, Manning-Bog AB, Di Monte DA, Fink AL. Dopamine and L-dopa disaggregate amyloid fibrils: implications for Parkinson's and Alzheimer's disease. *FASEB J.* (2004) 18:962–4. doi: 10.1096/fj.03-0770fje

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fgene.2021.745284

Genetic Variations and Differential DNA Methylation to Face Contrasted Climates in Small Ruminants: An Analysis on Traditionally-Managed Sheep and Goats

*Laure Denoyelle[1,2], Pierre de Villemereuil[3], Frédéric Boyer[1], Meidhi Khelifi[1], Clément Gaffet[1], Florian Alberto[1], Badr Benjelloun[1,4] and François Pompanon[1]**

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Grenoble, France, [2]GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France, [3]Institut de Systématique, Évolution, Biodiversité (ISYEB), École Pratique des Hautes Études | PSL, MNHN, CNRS, SU, UA, Paris, France, [4]Institut National de la Recherche Agronomique Maroc (INRA-Maroc), Centre Régional de Beni Mellal, Beni Mellal, Morocco

OPEN ACCESS

Edited by:
Xingbo Zhao,
China Agricultural University, China

Reviewed by:
Yu Jiang,
Northwest A and F University, China
Juliane Friedrich,
University of Edinburgh,
United Kingdom

***Correspondence:**
François Pompanon
francois.pompanon@univ-grenoble-alpes.fr

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 July 2021*
Accepted: *02 September 2021*
Published: *28 September 2021*

Citation:
Denoyelle L, de Villemereuil P, Boyer F, Khelifi M, Gaffet C, Alberto F, Benjelloun B and Pompanon F (2021) Genetic Variations and Differential DNA Methylation to Face Contrasted Climates in Small Ruminants: An Analysis on Traditionally-Managed Sheep and Goats. Front. Genet. 12:745284. doi: 10.3389/fgene.2021.745284

The way in which living organisms mobilize a combination of long-term adaptive mechanisms and short-term phenotypic plasticity to face environmental variations is still largely unknown. In the context of climate change, understanding the genetic and epigenetic bases for adaptation and plasticity is a major stake for preserving genomic resources and the resilience capacity of livestock populations. We characterized both epigenetic and genetic variations by contrasting 22 sheep and 21 goats from both sides of a climate gradient, focusing on free-ranging populations from Morocco. We produced for each individual Whole-Genome Sequence at 12X coverage and MeDIP-Seq data, to identify regions under selection and those differentially methylated. For both species, the analysis of genetic differences (F_{ST}) along the genome between animals from localities with high vs. low temperature annual variations detected candidate genes under selection in relation to environmental perception (5 genes), immunity (4 genes), reproduction (8 genes) and production (11 genes). Moreover, we found for each species one differentially methylated gene, namely AGPTA4 in goat and SLIT3 in sheep, which were both related, among other functions, to milk production and muscle development. In both sheep and goats, the comparison between genomic regions impacted by genetic and epigenetic variations suggests that climatic variations impacted similar biological pathways but different genes.

Keywords: adaptation, methylated DNA, whole genome, Morocco, small ruminant, acclimation

INTRODUCTION

The evolution and distribution of species is driven by the variation of their environment. Optimization of the individual's phenotype to fit the environment may occur on the long-term, as populations adapt to the local values of environmental drivers (i.e., native individuals having on average a higher fitness than migrants, Savolainen et al., 2013). The access to whole genome sequences now allows to study the genomic bases of local adaptation by identifying genes and

genomic regions under selection and the environmental parameters responsible for their selection (e.g., honeybees in the Iberian climate Henriques et al., 2018; dogs with the elevation Gou et al., 2014; or goats in the Moroccan climate Benjelloun et al., 2015). In the meantime, the context of global changes lead to an increasing interest on the ability of populations to develop short-term responses to face e.g., climate variations, as well as the interaction between short- and long-term mechanisms. Indeed, populations might quickly react to an environmental change through migration towards more favorable environments, which is well documented in the global warming context (Chen et al., 2011; McDonald et al., 2012). In the short-term, phenotypic plasticity can also increase the ability of organisms to cope with environmental changes. The ability of a genotype to produce, under different conditions, different phenotypes that are best fitted to the environment involves non-genetic mechanisms by which favorable variations can be acquired and even sometimes transmitted. They consist in epigenetic variations, parental effects, ecological and cultural variations (Danchin et al., 2011). One of such mechanism is epigenetics, notably molecular modifications of chromatin without modification of the DNA sequence itself (including e.g., DNA methylation or histone modification) which affects gene expression (Gibney and Nolan, 2010). Until now, most of the studies pointing out the role of molecular epigenetic mechanisms were performed in stress-controlled conditions on plant or animal models. In this case, the effect of one varying parameter on the physiological or epigenetic responses of the organisms is assessed. For example, the glyphosate herbicide injury on *Arabidopsis thaliana* (Kim et al., 2017), or the maternal diet in mice (Cooney et al., 2002) were shown to affect the methylome.

Since the end of the 2000s, the role of these epigenetic processes in ecology and evolution is increasingly being studied (Bossdorf et al., 2008). In this context, epigenetic marks were analyzed in natural populations. For example, the search for epigenetic differentiation by contrasting natural populations of *Lilium bosniacum* living in different habitat conditions (Zoldoš et al., 2018), or in wild baboons (*Papio cynocephalus*) populations with different food resources (natural fodder in a savanna environment or human food scraps, Lea et al., 2016).

Like wild species, farm animals will have to face future climate change, and understanding the genetic and epigenetic mechanisms responding to environmental changes, especially for free-ranging populations, is a key issue for the conservation of Farm Animal Genomic Resources (Bruford et al., 2015). Our study focuses on small domestic ruminants, namely sheep (*Ovis aries*) and goats (*Capra hircus*), which are a very good model for studying local adaptation. Indeed they spread all over the world for about 10,000 years from the domestication center in the Middle East, through human migration and commercial trades (Zeder, 2008). Until recent centuries, they were traditionally managed in a sustainable way leading to populations well adapted to the large variety of environments worldwide (Taberlet et al., 2008). Nowadays, the effect of the environment on farm animals strongly depends on breeding conditions. Intensively farmed animals are rather preserved from environmental variation because they are generally kept indoors with access to quality supply, in contrast to extensive livestock which are raised outdoors with little additional water and feed. Since the rise of the breed concept about 200 years ago, the selection for morphological and productivity traits improved the productivity, but resulted in the loss of genetic variability in many breeds, compromising their ability to adapt to future environmental changes (Taberlet et al., 2008). Thus, in the context of climate change gathering "epigenomic information to be integrated with phenotypic and genomic data to scrutinize the biological basis for adaptation and plasticity/resilience in livestock populations" is a key question in farm animal genomics research (Bruford et al., 2015). Until now, epigenetic analyses on livestock have been performed almost exclusively in stress-controlled conditions in relation with agronomic traits such as reproduction (Lan et al., 2013), production (Peñagaricano et al., 2014; X. Wang et al., 2015) and disease sensibility (Doherty et al., 2014). To our knowledge, only the study by Sevane et al. (2018) searched for DNA methylation in relation with climate variation by comparing cattle adapted to tropical and temperate climates.

In this context we aimed at characterizing both epigenetic and genetic variations with regards to climate variation in a single framework. Our study focused on sheep and goats sampled in Morocco where farming is mainly carried out in traditional ways with a strong impact of environmental conditions upon animals. Within each species, our goal was to identify candidate genomic regions involved in both genetic (selection on Single Nucleotide Polymorphism, SNP) and epigenetic (differential DNA methylation) mechanisms related to environmental variations, using whole genome comparisons of groups of individuals from locations with contrasted annual temperature range.

MATERIALS AND METHODS

Sampling

Morocco has very contrasted climates including hot/cold desert and semi-arid, warm-summer Mediterranean, dry-summer subarctic (Born et al., 2008). In this context, we sampled sheep (*Ovis aries*) and goats (*Capra hircus*) from locations with contrasted annual temperature variations (**Figure 1**, bioclimatic variable BIO7 from the WorldClim dataset, www.worldclim.org/bioclim, Hijmans et al., 2005). This variable is the difference between the maximum temperature of the warmest month and the minimum temperature of the coldest month. It is correlated with several temperature variables, sunshine and elevation (**Supplementary Table S1**). Twenty-one unrelated goats (10 and 11 from low and high annual temperature ranges, respectively) and 22 unrelated sheep (12 and 10 from low and high annual temperature ranges, respectively) were selected (**Figure 1**; **Supplementary Table S2**) from the samples collected between January 2008 and March 2012 in the frame of the NextGen European project (Grant Agreement no. 244356), in accordance with ethical regulations of the European Union Directive 86/609/EEC, and for which Whole

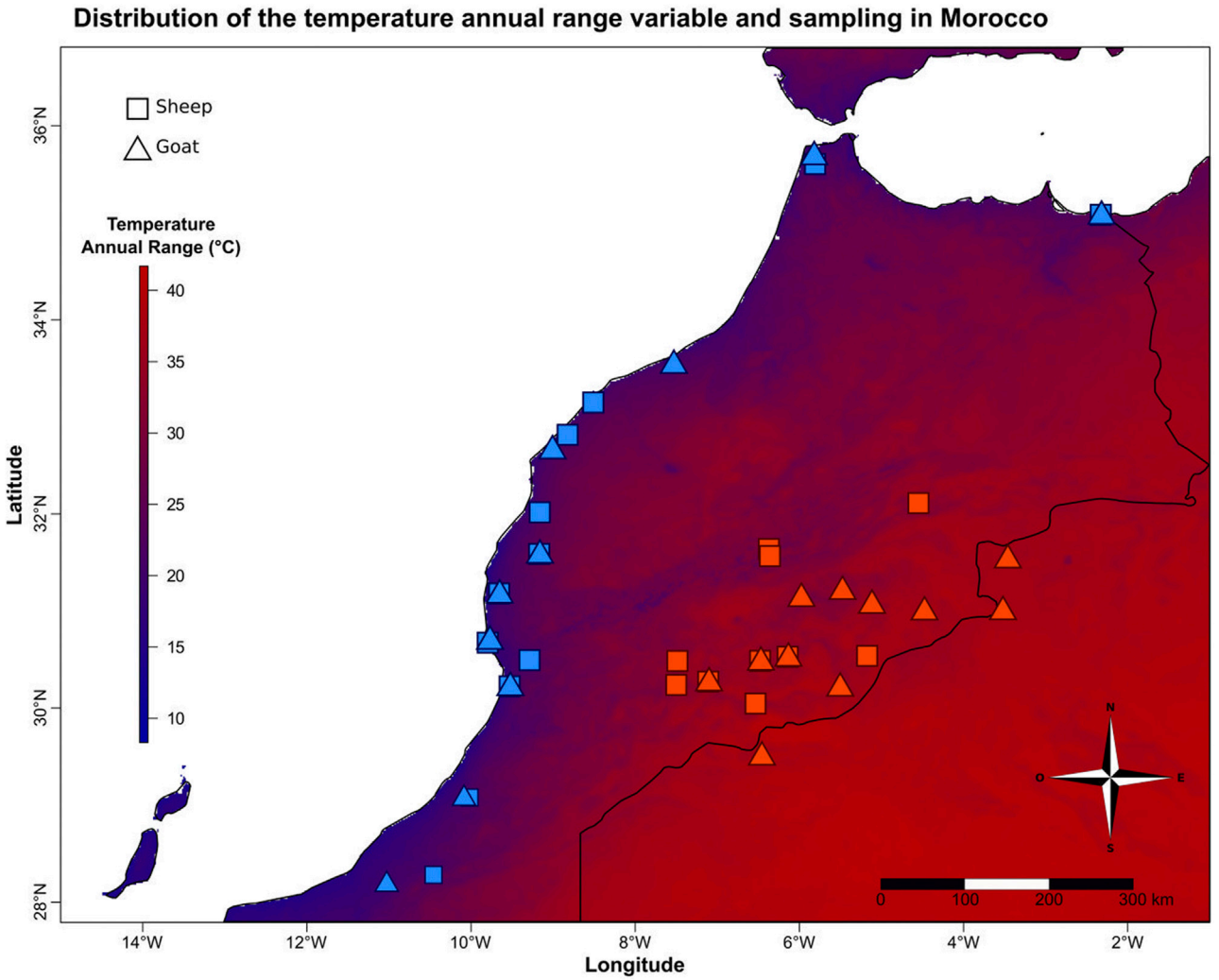

FIGURE 1 | Variation of the annual temperature range and sampling of small ruminants in Morocco. Distribution of sheep and goats sampled at each end of the temperature gradient in Morocco. The color gradient corresponds to the variation of the annual temperature range (BIO7).

Genome Sequences were already available (projects.ensembl.org/nextgen/). The average annual temperature range between the two groups is 20 vs. 39.5°C, which is almost twice for the animals living in Eastern Morocco.

Whole Genome Sequences

Dataset

We extracted the states of SNPs from the variant files generated by the NextGen project. Whole Genome Sequences (WGS) at 12–14 X coverage corresponding to the studied individuals were retrieved from the European Nucleotide Archive (www.ebi.ac.uk/ena, accession code PRJEB7436). The sequences, produced and filtered following the protocol in Alberto et al. (2018), were aligned on the reference genomes CHIR v1.0 (GenBank assembly GCA_000317765.147, Dong et al., 2013) for goats and OAR v3.1 (GenBank assembly GCA_000298735.1, Jiang et al., 2014), for sheep.

Detection of Selection Signatures

We kept SNPs with a missing data rate inferior to 30% and a minor allele frequency above 1%. Then we calculated the F_{ST} value (Weir and Cockerham, 1984) between the two groups (low vs. high temperature annual range for each species) for each SNP using the weir-fst-pop function in vcftools version 0.1.16 (Danecek et al., 2011). Candidate genomic regions were defined by 1) merging SNPs with the highest F_{ST} values (i.e., top 0.025% and top 0.001% for the pathway and gene approaches, respectively) closer than 5,000 bases from each other and then 2) adding 2,500 bases at each side of the region obtained. At the end 3) the genes overlapping these regions were extracted.

To annotate the genomic regions, two gene ontology analyses were performed with the GO Ontology database released on 2020-08-10 (Carbon and Mungall, 2018). The *Bos taurus* reference was used to annotate the candidate genes. We do

not expect any major issue due to the use of *B. taurus* as a reference as the genome synteny between these three Bovidae species is very high (see Jiang et al., 2014).

1) Enrichment analysis: it was performed on the list of genes extracted with the 0.025% threshold. For that, we used the statistical over-representation test in PANTHER (version 15.0, Mi et al., 2019) and the *Bos taurus* reference. Each biological process with a significant Fisher's exact test and a False Discovery Rate (FDR) lower than 0.05 was annotated. We kept the more basal significant biological pathway and searched the GO terms with the highest hierarchical level (i.e., child term of "biological process") associated.
2) Candidate gene identification: we extracted the GO term corresponding to the top genes using a more stringent threshold (0.001%) to avoid the false positive and highlight the most impacted genes. In addition, a specific gene bibliography was done using Google Scholar with the gene's name and "goat/sheep" or "livestock" or "mammals" as keywords, to determine what could be their effect on livestock phenotypes. These were grouped into large phenotype categories such as "Environmental perception," "Immunity," "Production" (genes involved in milk or meat production), "Reproduction" and "Other."

DNA Methylated Regions

DNA Extraction, Library Preparation and Sequencing

DNA was extracted from ear biopsies collected by the Nextgen consortium (see above) using the isolation of genomic DNA from tissues protocol from QIAamp® DNA micro kit (QIAgen, Germantown, MD, United States). Methylated DNA immunoprecipitation sequencing (MeDIP-Seq) was performed at the GeT-PlaGe core facility, INRAE Toulouse (http://www.get.genotoul.fr). Methylated DNA libraries were prepared according to Bioo Scientific's protocol using the Bioo Scientific NEXTflex™ Methyl-seq Library Prep Kit for Illumina Sequencing. Briefly, DNA was fragmented by sonication on a covaris M220, size selection was performed using AMPure XP beads and adapters were ligated to be sequenced. 3 µg of each library was diluted in 450 µl of TE buffer, denatured in boiling water for 10 min and immediately cooled in ice for 10 min. Then 50 µL of 10X concentrated IP buffer was added to the mixture as well as 1 µg of anti-5-methylcytosine monoclonal antibody (clone 33D3, Diagenode®). After overnight incubation at 4°C with agitation, the DNA-antibody complexes were purified using 40 µL of "Dynabeads-ProteinG" previously washed in PBS. The DNA-antibody complexes were washed twice in 700ul of IP buffer and then the DNA was purified using the iPure kit (Diagenode®). Control or input DNA samples were not coupled to the antibodies and were therefore not enriched in epigenetic markers. Then, 12 cycles of PCR were performed. Library quality was assessed using an Advanced Analytical Fragment Analyzer and libraries were quantified by QPCR using the Kapa Library Quantification Kit. MeDIP-Seq was performed on an Illumina® HiSeq 2500 with the Illumina Reagent Kits. At the end, the sequenced reads correspond to fragment of methylated DNA.

Dataset Preparation

Illumina paired-end reads (mean size of 125 bp) from all animals were aligned on the same reference genomes as WGS data by using the default parameters of BWA-MEM version 0.7.12 (Li and Durbin, 2009). The function multicov from bedtools version 2.29.2 (Quinlan and Hall, 2010) was used to report the MeDIP-Seq coverage of alignment files inside windows of 100 non-overlapping base pairs along the genome. The same function was used to extract the WGS coverage. In these windows we also extracted the number of CpG from the reference genomes by using the MEDIPS package (Lienhard et al., 2014) on *R* software version 3.5.3 (R Core Team, 2019).

We compared genome wide coverage profiles of MeDIP-Seq by using the MEDIPS.correlation function from the MEDIPS package with the uniq, extent, and shift parameters equal to 0 and the window size parameter equal to 100, to keep all MeDIP-Seq reads in each 100 bases window on the genome. We verified with a pairwise Pearson test that the number of read by windows for one animal had a correlation coefficient greater than 0.6 with all other individuals of the same species.

The numbers of WGS and MeDIP-Seq reads in each window were normalized between individuals with default parameters of calcNormFactors function (edgeR package, Robinson et al., 2010). Then, normalized MeDIP-Seq read counts were divided by normalized WGS read counts in order to get rid of the effect of sequencing coverage variations (e.g., due to repeated regions). Normalized values have been rounded to two decimals. Four outlying windows (2 for each species) with 10 times more MeDIP-Seq reads than the others were removed as this may be due to PCR or alignment errors.

Whole Epigenome Correlations

For each species, correlations between the number of CpG and that of MeDIP-Seq reads found in the 100 bp window were tested using a Pearson's correlation test. We used the same test to test the correlation between species for each of these two variables. For that we used the correspondence between *Capra* and *Ovis* orthologous regions resulting from the cross-alignment between the two reference genomes CHIR_1.0 and OAR v3.1 performed by Alberto et al. (2018). Results were represented as density plots performed with the hexbin package (Carr et al., 2019).

Detection of Differentially Methylated Regions

We used the edgeR package to detect DMRs. Starting from non-standardized MeDIP-Seq read counts, we kept the 100 bp windows for which at least 10% of the samples had at least 1 count-per-million (1 read per million windows, this accommodates differences in library sizes between sample) and we normalized the read counts with the calcNormFactors function with default parameters. We estimated common dispersion with the estimateGLMRobustDisp function which mitigates the influence of outliers (X. Zhou et al., 2014) and we tested the occurrence of differential methylation between groups with the default parameters of the exactTest function.

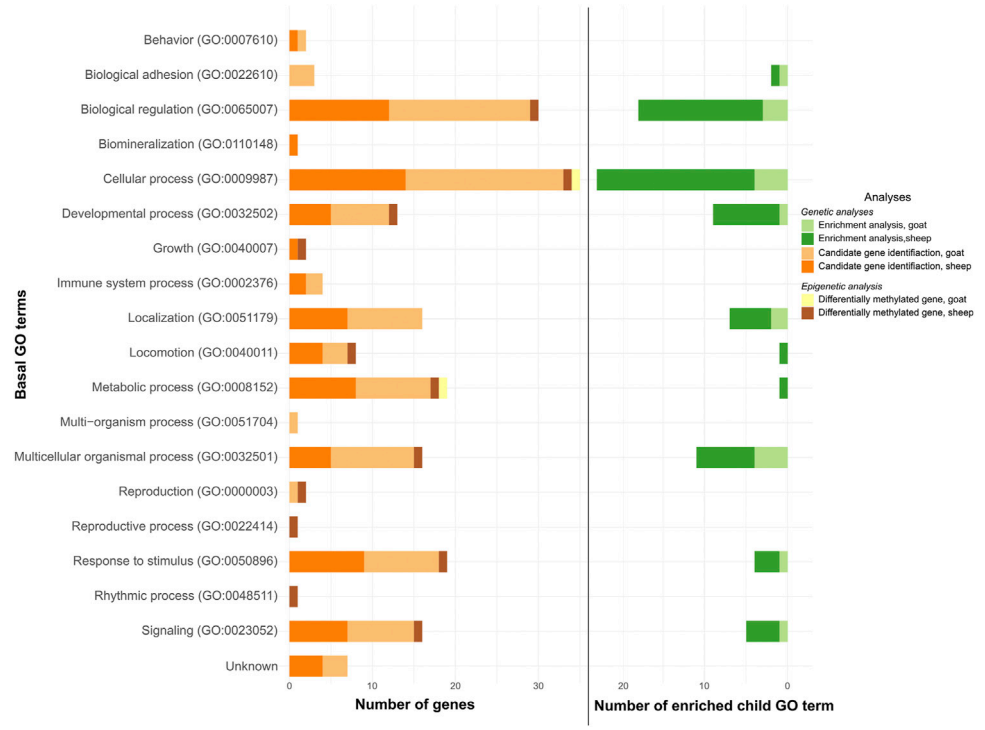

FIGURE 2 | GO terms putatively involved in the adaptation and acclimation of small ruminants to temperature variations. Number of genes or child GO terms per basal GO terms for candidate gene identification (under selection -oranges- and differentially methylated -yellow/brown-) and enrichment analysis (greens), respectively.

The *p*-values obtained were converted into q-values with the Bioconductor package qvalue (Storey et al., 2015) and the ones showing q-values inferior to 0.01 were selected for the rest of the analysis. Finally, we merged significant windows distant than less than 5,000 bases into regions.

In order to delineate the peaks of methylation occurring for some of the individuals within these regions, we used the changepoint package (Killick and Eckley, 2014) which detects breakpoints in time series. We looked for all possible breakpoints within a region using the cpt.meanvar function that identify change in mean and variance of the sum of coverages for all animals of each group. We kept the region between two breakpoints that contained at least one window detected as significant with edgeR, thus, the delimited regions were merged with the bedr package (Waggott et al., 2017) when they overlapped. Then, we used a Wilcoxon test to detect peaks with different methylation levels between the two groups. Among those peaks, we discarded those where the number of MeDIP-Seq reads was related to the DNA coverage in the WGS data, because here high MeDIP-Seq counts would result from the high DNA coverage (e.g., due to repeated sequences). The regions containing the remaining peaks were mapped on the reference genomes and included or overlapping genes were identified. We used the Functional classification viewed in gene list analysis in PANTHER to extract the Biological Pathways annotated in the Gene Ontology Database associated to these genes.

RESULTS

Detection of Selection Signatures

The F_{ST} values at each genome position and their overall distribution are presented for goats and sheep in **Supplementary Figures S1, S2**, respectively. Almost 94,5% of the F_{ST} values are less or equal to 0.1 for both species. The top 0.025 and 0.001% thresholds correspond to F_{ST} value of 0.4 and 0.55 respectively.

Enrichment Analysis

The 5,000 SNPs for goat and 6,000 SNPs for sheep which harbored the highest 0.025% F_{ST} values (**Supplementary Figures S3, S4**) corresponded to 442 and 489 genes, respectively, putatively under selection.

The enriched pathways were associated with 8 and 10 GO terms in goat (**Figure 2**, light green color; **Supplementary Table S3**) and sheep (**Figure 2**, dark green color; **Supplementary Table S4**), respectively. Eight were common to both species.

Among the 0.025% top genes (i.e., 931 genes selected for the enrichment analysis), 41 were orthologous between sheep and goats. When we did the same enrichment analysis (described above) with the 41 genes shared by sheep and goats we obtained 21 GO terms. Four of them had the "Nervous system development" in their parental terms; and the last three were associated with cellular processes, one of which was related to synapses organization.

TABLE 1 | Phenotype categories in which the candidate genes (top 0.001%) are involved.

Species	Phenotype categories	Genes
Goat	Environmental perception	TMTC2, EDIL3, SASH1
	Immunity	—
	Production	CP, SLC9A9, PPFIA2, TMTC2, GALNTL6, CTSB, SASH1, PLCG2, MYADM
	Reproduction	ADGRB3, USH1C, PLCG2, MYADM, SERPINB7
	Other	TRABD2B, MVB12B, KLF12, TNFSF9, CALHM3, SORCS1, RAB30
	Unknown	LOC102184299, LOC106503718
Sheep	Environmental perception	NOX3, KSR2
	Immunity	SAMD12, SEMA5A, SEZ6L
	Production	BMPER, KSR2, PTPRE
	Reproduction	GCSAML, KSR2, BMPER
	Other	AGO3, IQCJ, KCNG3, ANO6, CASP8AP2, CNKSR3, ENGASE, KCNA4
	Unknown	LOC101107868, LOC101119001

Candidate Gene Identification

The top 0.001% F_{ST} values corresponded to 194 SNPs for goat and 146 SNPs for sheep distributed in 56 and 42 genomic regions, respectively. Their location in the species genome is detailed in **Supplementary Tables S5, S6** for goat and sheep, respectively. In goats, 34 of the 56 regions were intergenic and the others were related to 22 genes (**Table 1**). For sheep, we obtained 24 intergenic regions and 18 genes (**Table 1**), which differed between the two environmental groups. No genes were common between the two species.

The 22 genes were linked to 15 GO terms in goat (**Figure 2**, light orange color; **Supplementary Table S5**) and the 18 genes in sheep were associated with 14 GO terms (**Figure 2**, dark orange color; **Supplementary Table S6**). Twelve of these GO terms were common to both species.

In livestock, several of these genes are known to be involved in phenotypic changes in relation with the perception of the environment (3 genes in goat, 2 in sheep), immunity (1 in goat, 3 in sheep), production (9 in goat, 2 in sheep), reproduction (5 in goat and 3 in sheep). The "Other" category brings together the genes whose effect on the phenotype was not documented in livestock species. We also detected genes that were uncharacterized (**Table 1**, **Supplementary Tables S5, S6**).

Distribution of CpG and Methylated Regions in Both Species

The majority of the 22 million windows (i.e., 96% for each species) carried one methylated read, 3 and 2% more than one, and 1 and 2% none, for goat and sheep, respectively. Concerning the number of CpG, 47% of the windows did not have any, 50% had between 1 and 4, and 3% had more than 4 in both species.

Genomic windows of 100 bp that were orthologous between *Capra* and *Ovis* showed a correlation between species for the number of methylated read (R = 0.81, p-value < 2.10^{-16}, **Figure 3A**) and the number of CpGs (R = 0.67, p-value < 2.10^{-16}, **Figure 3B**) A large number of windows had few CpG and methylated reads (25% carried less than 3 CpGs and 1 methylated read). Within species, we found a correlation between the number of CpGs and that of methylated reads, R = 0.49 (p-value < 2.10^{-16}, **Figure 3C**) for goats and R = 0.44 (p-value < 2.10^{-16}, **Figure 3D**) for sheep.

For the detection of DMRs, 20,256 windows in goat and 16,281 windows in sheep were kept out of more than 24 millions. The edgeR analysis revealed 20 and 7 windows with significant differences in methylation levels between the two groups (i.e., low vs high annual temperature variations) in goat and sheep, respectively. Merging candidate windows closer than 5,000 bp, we obtained 8 and 4 regions in which we detected 8 and 5 methylation peaks for goat and sheep, respectively. From these, 4 and 2 peaks respectively showed significant differential MeDIP-Seq coverage between the 2 groups (low vs. high temperature annual range). For 4 peaks, the differential MeDIP-Seq coverage was concomitant to a differential WGS coverage, indicating that the differentiation would probably result from repetition polymorphism. Then only one region per species was kept as resulting from a differential methylation. They were associated to AGPTA4 in goat and SLIT3 in sheep (**Figure 4**).

The GO terms associated with these genes were "Cellular process," and "Metabolic process" for AGPAT4 (**Figure 2**, yellow color); and "Biological regulation," "Cellular process," "Developmental process," "Growth," "Locomotion," "Metabolic process," "Multicellular organismal process," "Reproduction," "Reproductive process," "Response to stimulus," "Rhythmic process," and "Signaling" for SLIT3 (**Figure 2**, brown color).

DISCUSSION

To our knowledge, this is the first study to investigate simultaneously the genetic and epigenetic variations putatively involved in the adjustment of phenotypes to field conditions in small ruminants. Until now, few publications jointly addressed both mechanisms. Some studies highlighted significant association between local genetic variation and the presence of DMRs like that of Eichten et al. (2013) who found that 51% of the DMRs between maize populations were associated with local SNPs. Foust et al. (2016) and Herrera and Bazaga (2010) found a correlation between AFLP and MS-AFLP marks in two salt marsh perennials (*Spartina alterniflora* and *Borrichia frutescens*) and violet populations (*Viola cazorlensis*), respectively. In our study, we investigated both genetic and epigenetic variations in sheep

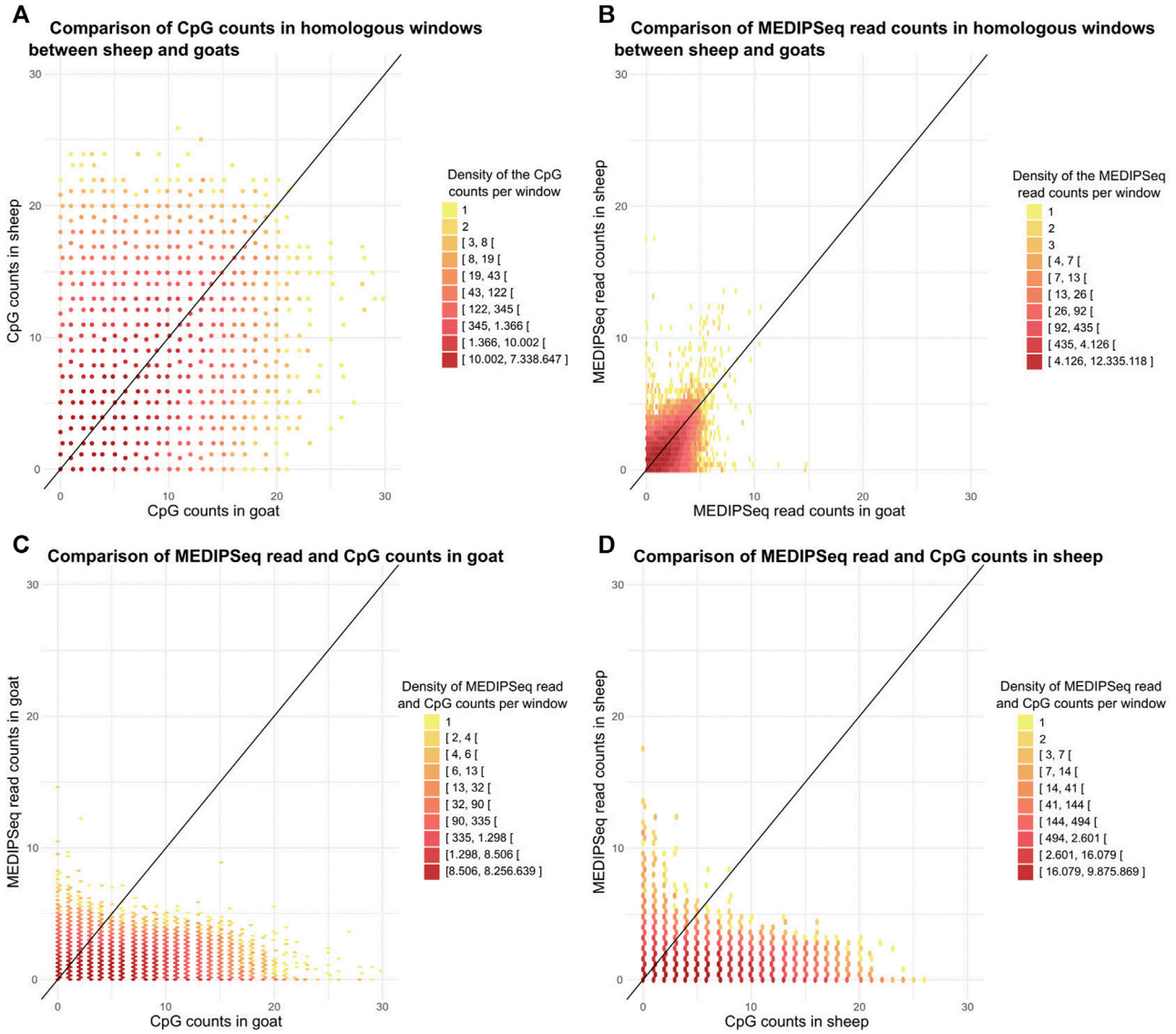

FIGURE 3 | Characterization and comparison of sheep and goats methylomes. Relationship between the number of **(A)**. CpG, or **(B)**. MeDIP-Seq reads per 100 bp window between the two species across the genome. Relation between the number of MeDIP-Seq reads and the number of CpG per 100 bp widow for **(C)**. goat, and **(D)**. sheep. The colors represent the number of points and the black line (y = x) is represented.

and goats originating from field populations. The sampling was carried out in order to avoid a breed effect as much as possible. However, given the specific distribution and adaptation of some breeds, several Ghazalia goats and D'man sheep were part of the groups submitted to high temperature annual range (c.a. 39.5°C) as opposed to that with low temperature annual range (c.a. 20°C). They represented less than half of the individuals (3/11 and 4/10 in goats and sheep, respectively) and had similar methylation and F_{ST} patterns to the other breeds from the same group. Thus, the mechanisms inferred would be shared by the variety of Moroccan breeds represented in our sampling. Our study points out several genes which polymorphism or methylation level is contrasted between the two groups and may contribute to the adaptation and acclimatization to these specific environments. The candidate genes putatively under selection are primarily involved in environmental perception, immunity, production and reproduction, and their variations could be driven by different environmental factors. First, a direct effect of high temperatures might operate through thermoregulation by reducing body temperature, through the decrease of metabolic heat and the increase of heat dissipation. This might be obtained by avoiding solar radiations (Al-Tamimi, 2007), increasing respiration or sweating and reducing food intake (Marai et al., 2007; Sejian et al., 2010). These adjustments might reduce the energy allocated to other biological processes including productivity (Tao et al., 2011; Dado-Senn et al., 2019) or reproduction (Meyerhoeffer et al., 1985; Monteiro et al., 2016; Krishnan et al., 2017). In relation with temperature annual range, we found candidate

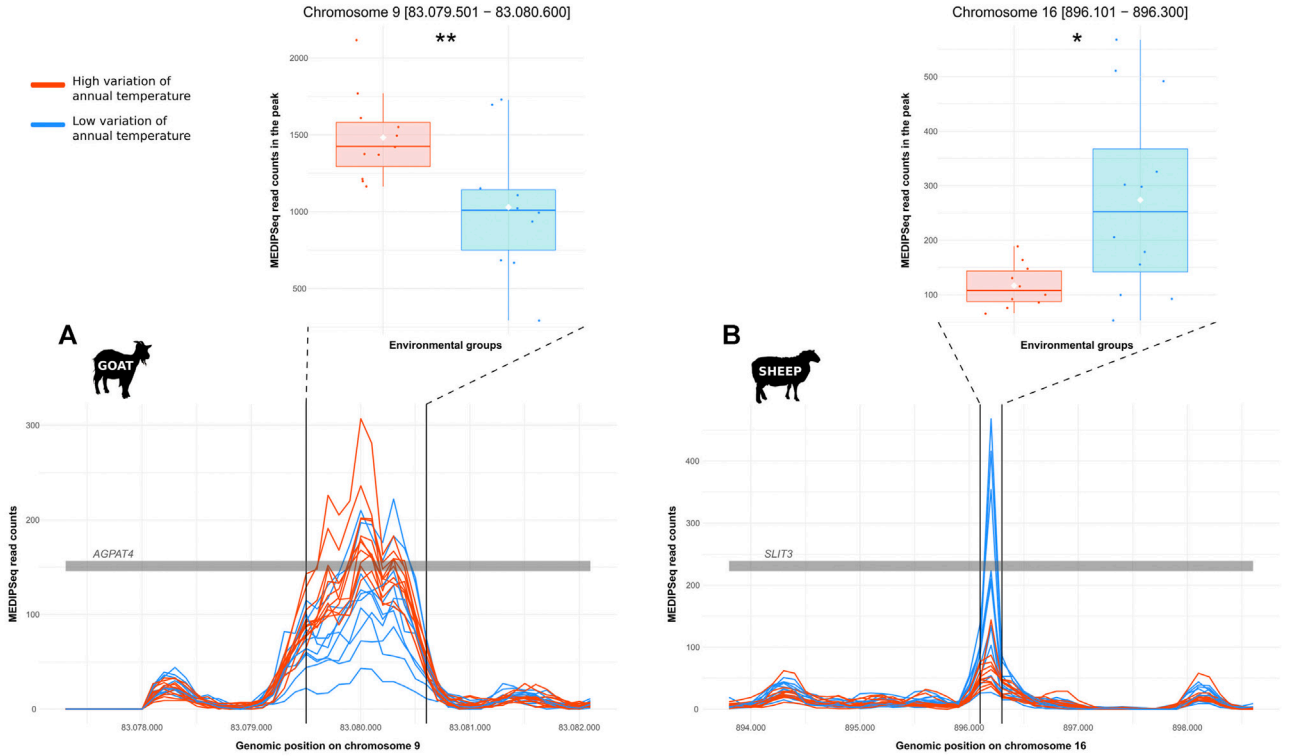

FIGURE 4 | Differentially methylated regions putatively involved in temperature acclimation in small ruminants. DMRs between genomes from locations with low (blue) and high (red) temperature annual range for **(A)**. goat or **(B)**. sheep. Each line represents the variations of the level of DNA methylation along a gene for one individual; the grey bar represents part of the gene; the two black lines delineate the peak on which the differential level of methylation was tested. Boxplots represent the level of methylation (MeDIP-Seq coverage) for each group compared with a Wilcoxon test (Significance codes: ** = p-value ≤ 0.01, * = p-value ≤ 0.05).

genes involved in production, including milk production for CP, MYAMD, PTPRE, SASH1 and SLC9A9 (Hussein and Staufenbiel, 2012; Dong et al., 2015; Laodim et al., 2017; Shin et al., 2017; Xiang et al., 2017), mammary gland development for PPFIA2 (Mota et al., 2020) and morphology for TMTC2 (Abo-Ismail et al., 2017; Z.-H. Fang and Pausch, 2019). In addition, CTSB, GALNTL6, and KSR2 genes were related to meat production (Russo et al., 2002; Doran et al., 2014; Puig-Oliveras et al., 2014; Lukić et al., 2020), and BMPER and KSR2 with body size (Puig-Oliveras et al., 2014; Zhao et al., 2015; Lukić et al., 2020). Genes involved in reproduction were also impacted. ADGRB3, GCSAML, and KSR2 affect fertility (Pimentel et al., 2011; X. Wang et al., 2015; Mohammadi et al., 2020), while PLCG2 and SERPINB7 affect gestation (Vallée et al., 2003; Ponsuksili et al., 2012). Moreover offspring survival can also be impacted through litter size via BMPER and USH1C (Kwon et al., 2016; Bolormaa et al., 2017; Mota et al., 2020) and offspring weight with MYADM (Gonzalez et al., 2020). Also, we found two candidate genes related to environmental perception. NOX3 is involved in the response to stresses in general (Lambeth, 2004) and has been related with heat stress in chicken (Aggrey et al., 2018), and KSR2 plays a role in the energy balance by adapting feeding behavior and thermogenesis in mice (Guo et al., 2017). Second, environmental parameters covarying with the annual temperature range (BIO7) may be prominent in driving the selection on the genes which polymorphism was contrasted between the two groups. TMTC2 and EDIL3 are involved in adaptation to high altitude in sheep and cattle (Ji Yang et al., 2016; Pierce et al., 2020), this later variable being correlated with BIO7 in Morocco. In addition, BIO7 variable is correlated with sunshine which impacts directly skin pigmentation, which would explain the detection of SASH1, which regulates melanocytes migration in relation with the production and transfer of melanin to protect skin from UV radiations (D. Zhou et al., 2013). Third, the temperature regime may have indirect impacts on phenotypes by driving other environmental factors. Especially, it determines plant resources (i.e., food quality and quantity) and is related to water availability, and thus strongly impacts the diet of free ranging livestock. The lack of resources has a negative impact on, e.g., reproduction (Martin et al., 2004; J. J. Robinson et al., 2006) and lactation (Razzaghi et al., 2016). In contrast, supplementation, which has a positive effect on milk yield and composition in sheep, could temperate this negative impact on lactation (Caroprese et al., 2011). As a candidate gene, we found PLCG2 which is regulated in the cattle liver according to feed intake (Salleh et al., 2017). Likewise, genes involved in immune functions may be selected by multi-factorial drivers related to nutrition or the pathogenic context in relation with climate (Bett et al., 2017). This could be the case for SAMD12 playing a role in pancreatitis in human (Giri et al., 2016),

SEZ6L associated with foot-and-mouth disease (Lee et al., 2015) or SEMA5A with mastitis in cattle (Sugimoto et al., 2006).

Besides the candidate genes putatively involved in local adaptation through allelic selection, we identified two strong candidates in relation with an epigenetic response. The global epigenomic context was similar in sheep and goats, and the level of DNA methylation along the genome was correlated to the CpG content. This is consistent with what is known about mammalian epigenomes where the CpG methylation generally covers the entire nuclear genome, with the exception of CpG rich regions near the promoters of active genes (Suzuki and Bird, 2008; Feng et al., 2010). We found two genes differentially methylated according to the provenance of the individual, namely AGPAT4 (1-acylglycerol-3-phosphate O-acyltransferase 4) in goats, and SLIT3 (Slit Guidance Ligand 3) in sheep. AGPAT4 is part of the 1-acyl-sn-glycerol 3-phosphate acyltransferases family which is involved in triglyceride synthesis (Lu et al., 2005; Takeuchi and Reue, 2009), with a role in the milk composition of ruminants (Bionaz and Loor, 2008; He et al., 2011; Jie Yang et al., 2016). It has an organ-related level of expression in human (Lu et al., 2005; Prasad et al., 2011) and plays a role in muscle development and meat quality in cattle (X. Fang et al., 2017). In addition, the variation of AGPAT genes expression in the mammary gland has been related to milk fat depression in dairy cows (Peterson et al., 2003). On its side, SLIT3 takes place into the SLIT-ROBO interactions which are involved in the development of the nervous system (K. H. Wang et al., 1999; Andrews et al., 2007), cardiac development (Liu et al., 2003; Mommersteeg et al., 2015) and reproduction, especially menstrual cycle (Dickinson and Duncan, 2010) and fertility (Amundson et al., 2015). SLIT3 was also shown to be differentially methylated and upregulated in goat mammary glands during lactation compared to the dry period (Zhang et al., 2017). Such DMRs related to milk quality and lactation might result directly from heat stress. Indeed, it was shown that heifers submitted to a thermic shock during fetal development had smaller mammary gland alveoli, resulting in lower milk production (Skibiel et al., 2018; Dahl et al., 2019). The DMRs observed could also result from indirect effects of the climate through a variation of the diet resulting from a shift in plant resources. Peterson et al. (2003) showed that the concentrate forage diet influences the mRNA abundance of AGPAT genes in mammary glands in dairy cows, involving a modification of milk fat. The regulation of the SLIT-ROBO pathway is also affected by the diet (Amundson et al., 2015). Indeed, the diet can directly influence the levels of DNA methylation (Anderson et al., 2012) either due to direct exposure of the individuals or to parental effects. For example, a heat-stress-induced perturbation of mammary glands development in the fetus, related to differential methylation and differential gene expression (Skibiel et al., 2018), results in effects persisting over the lactations and in the next generation too (Ouellet et al., 2020). Maternal nutrition also affects DNA methylation and gene expression of lambs, especially in muscle and adipose tissues (Lan et al., 2013; Peñagaricano et al., 2014; Namous et al., 2018). Obviously, we found differential methylation in genes related to lactation in ear biopsies, not mammary glands. However, sweat glands localized in the dermis and mammary glands have a common tissue origin, and recent studies showed that the most similar tissues had comparable CpG methylation patterns (Lokk et al., 2014). Given the epigenetic memory, embryonic patterns may be retained in adult tissues (Hon et al., 2013). In this case, an early stress (e.g., *in utero*) could modify the methylation pattern of embryonic cells, which could be retained in structures with a common origin. Further characterization of the methylation patterns in different organs, including mammary glands, is needed to confirm the possible role of these mechanisms.

We found that candidate genes for local adaptation in response to different temperature regimes were related to a large variety of biological pathways which were mainly involved in functions such as immunity, perception of the environment and reproduction, or related to productivity traits. While these functions were affected in both sheep and goats, the best candidate genes (i.e., top 0.001%) identified were different, confirming that alternative genes might be selected under the same environmental constraints for close species (e.g., Benjelloun, 2015; Raeymaekers et al., 2017) or even populations of the same species (Manceau et al., 2010; Benjelloun et al., 2015). However, out of the 0.025% top genes (i.e., 931 genes selected for the enrichment analyses), 41 were orthologous between sheep and goats and mainly related to the nervous system development. Interestingly, the present study also investigated epigenetic variations in the individuals displaying candidate genes for local adaptation. Our deliberately stringent approach to avoid false positive, which identified only one candidate DMR per species, prevents a global view but already points out similarities between traits impacted by genetic and epigenetic mechanisms. In sheep, genes impacted by selection such as BMPER, GCSAML, KSR2, PTPRE affect embryo quality, litter size and lactation, which are also influenced by SLIT3 that was differentially methylated. Likewise, in goats, AGPAT4 that was differentially methylated has an effect on lipid metabolism, meat quality and the presence of lipids in milk, these traits being influenced by several candidate genes such as CTSB, MYADM, SERPINB7, USH1C, GALNTL6 and CALHM3.

CONCLUSION

This study showed differentially methylated genes and candidate genes under selection related to a differential regime of temperatures underwent by sheep and goats. Genetic and epigenetic mechanisms share several target functions including lactation and milk composition. These adjustments of phenotypes to the two contrasted environments may result from drivers related to temperature either directly or indirectly through e.g., the variation of plant resources conditioning the diet. We do not yet have enough evidence to say that these results reflect the joint effects of selection and epigenetic regulation on the same traits. However, such combined effects might play an important role in evolution (Richards, 2011), and we can wonder whether DMRs might cushion phenotypic variations due to prior selection of genes or reflect the persistence of adaptive plasticity as long as adaptation is not complete? Deciphering the relative impact of climate and e.g., differential diets on the genetic and epigenetically-induced variations of phenotypes will be a major stake for developing future breeding strategies in the context of global change.

DATA AVAILABILITY STATEMENT

The genetic dataset (fastq of Whole Genome Sequencing) analyzed in this study can be found in the European Nucleotide Archive (www.ebi.ac.uk/ena, accession code PRJEB7436). The epigenetic dataset (fastq of the MeDIP-seq) generated for this study can be found in the European Nucleotide Archive (www.ebi.ac.uk/ena, accession code PRJEB47531).

ETHICS STATEMENT

The animal study was reviewed and approved by European Union Directive 86/609/EEC.

AUTHOR CONTRIBUTIONS

FP and FB supervised the project. FP and BB contributed to the design project. BB conducted the sampling of Moroccan small ruminants. FA, FP, and LD performed DNA extraction and managed the DNA samples. LD, CG, MK, PD, and FB performed the analyzes. LD and FP wrote the paper.

FUNDING

We thank the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation for financing LD. The LECA is part of the Labex OSUG@2020 (ANR10 LABX56). This study was partly funded by the projects Climgen (Program FACE ERA-NET PLUS 2014, ANR- 14-JAFC-0002) and MethylGoat (AGIR-2014, Univ. Grenoble Alpes). Computations were performed using the CIMENT infrastructure, which is supported by the Auvergne-Rhône-Alpes region (CPER07_13 CIRA).

ACKNOWLEDGMENTS

We thank Florence Plisson and Sylvie Veyrenc for their help during the DNA manipulations (extraction, MeDIP-Seq); and Hervé Acloque for help in setting up the MeDIP-Seq approach. This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr), and was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). All the computations presented in this paper were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is partly supported by the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.745284/full#supplementary-material

REFERENCES

Abo-Ismail, M. K., Brito, L. F., Miller, S. P., Sargolzaei, M., Grossi, D. A., Moore, S. S., et al. (2017). Genome-wide Association Studies and Genomic Prediction of Breeding Values for Calving Performance and Body Conformation Traits in Holstein Cattle. *Genet. Sel Evol.* 49 (1), 82. doi:10.1186/s12711-017-0356-8

Aggrey, S. E., Habashy, W. S., Milfort, M. C., Adomako, K., Fuller, A. L., and Rekaya, R. (2018). "Molecular and Cellular Mechanisms that Underlie Genes and Antioxidant Enzyme Activities in Meat-type Birds during Heat Stress," in World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.

Al-Tamimi, H. J. (2007). Thermoregulatory Response of Goat Kids Subjected to Heat Stress. *Small Ruminant Res.* 71 (1), 280–285. doi:10.1016/j.smallrumres.2006.04.013

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

Amundson, O. L., Fountain, T. H., Larimore, E. L., Richardson, B. N., McNeel, A. K., Wright, E. C., et al. (2015). Postweaning Nutritional Programming of Ovarian Development in Beef Heifers1,2. *J. Anim. Sci.* 93 (11), 5232–5239. doi:10.2527/jas.2015-9067

Anderson, O. S., Sant, K. E., and Dolinoy, D. C. (2012). Nutrition and Epigenetics: An Interplay of Dietary Methyl Donors, One-Carbon Metabolism and DNA Methylation. *J. Nutr. Biochem.* 23 (8), 853–859. doi:10.1016/j.jnutbio.2012.03.003

Andrews, W. D., Barber, M., and Parnavelas, J. G. (2007). Slit?Robo Interactions during Cortical Development. *J. Anat.* 211 (2), 188–198. doi:10.1111/j.1469-7580.2007.00750.x

Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F. d. r., Coissac, E., Stucki, S., et al. (2015). Characterizing Neutral Genomic Diversity and Selection Signatures in Indigenous Populations of Moroccan Goats (*Capra hircus*) Using WGS Data. *Front. Genet.* 6, 107. doi:10.3389/fgene.2015.00107

Benjelloun, B. (2015). "Diversité des génomes et adaptation locale des petits ruminants d'un pays méditerranéen: Le Maroc,". doctoral thesis (Grenoble (France): Université Grenoble Alpes). http://www.theses.fr/2015GREAV011.

Bett, B., Kiunga, P., Gachohi, J., Sindato, C., Mbotha, D., Robinson, T., et al. (2017). Effects of Climate Change on the Occurrence and Distribution of Livestock Diseases. *Prev. Vet. Med.* 137, 119–129. doi:10.1016/j.prevetmed.2016.11.019

Bionaz, M., and Loor, J. J. (2008). ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 Are the Most Abundant Isoforms in Bovine Mammary Tissue and Their Expression Is Affected by Stage of Lactation. *J. Nutr.* 138 (6), 1019–1024. doi:10.1093/jn/138.6.1019

Bolormaa, S., Brown, D. J., Swan, A. A., van der Werf, J. H. J., Hayes, B. J., and Daetwyler, H. D. (2017). Genomic Prediction of Reproduction Traits for Merino Sheep. *Anim. Genet.* 48 (3), 338–348. doi:10.1111/age.12541

Born, K., Christoph, M., Fink, A. H., Knippertz, P., Paeth, H., and Speth, P. (2008). "Moroccan Climate in the Present and Future: Combined View from Observational Data and Regional Climate Scenarios," in *Climatic Changes and Water Resources in the Middle East and North Africa*. Editors F. Zereini and H. Hötzl (Springer Berlin Heidelberg), 29–45. doi:10.1007/978-3-540-85047-2_4

Bossdorf, O., Richards, C. L., and Pigliucci, M. (2008). Epigenetics for Ecologists. *Ecol. Lett.* 11, 106–115. doi:10.1111/j.1461-0248.2007.01130.x

Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F. J., et al. (2015). Prospects and Challenges for the Conservation of Farm Animal Genomic Resources, 2015-2025. *Front. Genet.* 6, 314. doi:10.3389/fgene.2015.00314

Carbon, S., and Mungall, C. (2018). *Gene Ontology Data Archive* [Data Set]. *Zenodo*. doi:10.5281/zenodo.3980761

Caroprese, M., Albenzio, M., Bruno, A., Fedele, V., Santillo, A., and Sevi, A. (2011). Effect of Solar Radiation and Flaxseed Supplementation on Milk Production and Fatty Acid Profile of Lactating Ewes under High Ambient Temperature. *J. Dairy Sci.* 94 (8), 3856–3867. doi:10.3168/jds.2010-4067

Carr, D., Lewin-Koh, N., Maechler, M., and Sarkar, D. (2019). Hexbin: Hexagonal Binning Routines. (R package version 1.28.0) [Computer software]. https://CRAN.R-project.org/package=hexbin.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., and Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science* 333 (6045), 1024–1026. doi:10.1126/science.1206432

Cooney, C. A., Dave, A. A., and Wolff, G. L. (2002). Maternal Methyl Supplements in Mice Affect Epigenetic Variation and DNA Methylation of Offspring. *J. Nutr.* 132 (8), 2393S–2400S. doi:10.1093/jn/132.8.2393S

Dado-Senn, B., Skibiel, A. L., Fabris, T. F., Dahl, G. E., and Laporta, J. (2019). Dry Period Heat Stress Induces Microstructural Changes in the Lactating Mammary Gland. *PLOS ONE* 14 (9), e0222120. doi:10.1371/journal.pone.0222120

Dahl, G. E., Skibiel, A. L., and Laporta, J. (2019). In Utero Heat Stress Programs Reduced Performance and Health in Calves. *Vet. Clin. North America: Food Anim. Pract.* 35 (2), 343–353. doi:10.1016/j.cvfa.2019.02.005

Danchin, É., Charmantier, A., Champagne, F. A., Mesoudi, A., Pujol, B., and Blanchet, S. (2011). Beyond DNA: Integrating Inclusive Inheritance into an Extended Theory of Evolution. *Nat. Rev. Genet.* 12 (7), 475–486. doi:10.1038/nrg3028

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

Dickinson, R. E., and Duncan, W. C. (2010). The SLIT-ROBO Pathway: a Regulator of Cell Function with Implications for the Reproductive System. *Reproduction* 139 (4), 697–704. doi:10.1530/REP-10-0017

Doherty, R., Farrelly, C. O., and Meade, K. G. (2014). Comparative Epigenetics: Relevance to the Regulation of Production and Health Traits in Cattle. *Anim. Genet.* 45, 3–14. doi:10.1111/age.12140

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and Automated Whole-Genome Optical Mapping of the Genome of a Domestic Goat (*Capra hircus*). *Nat. Biotechnol.* 31 (2), 135–141. doi:10.1038/nbt.2478

Dong, Y., Zhang, X., Xie, M., Arefnezhad, B., Wang, Z., Wang, W., et al. (2015). Reference Genome of Wild Goat (capra Aegagrus) and Sequencing of Goat Breeds Provide Insight into Genic Basis of Goat Domestication. *BMC Genomics* 16 (431). doi:10.1186/s12864-015-1606-1

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 (1), 837. doi:10.1186/1471-2164-15-837

Eichten, S. R., Briskine, R., Song, J., Li, Q., Swanson-Wagner, R., Hermanson, P. J., et al. (2013). Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. *Plant Cell* 25 (8), 2783–2797. doi:10.1105/tpc.113.114793

Fang, X., Zhao, Z., Yu, H., Li, G., Jiang, P., Yang, Y., et al. (2017). Comparative Genome-wide Methylation Analysis of Longissimus Dorsi Muscles between Japanese Black (Wagyu) and Chinese Red Steppes Cattle. *PLoS ONE* 12 (8), e0182492. doi:10.1371/journal.pone.0182492

Fang, Z.-H., and Pausch, H. (2019). Multi-trait Meta-Analyses Reveal 25 Quantitative Trait Loci for Economically Important Traits in Brown Swiss Cattle. *BMC Genomics* 20 (1), 695. doi:10.1186/s12864-019-6066-6

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., et al. (2010). Conservation and Divergence of Methylation Patterning in Plants and Animals. *Proc. Natl. Acad. Sci.* 107 (19), 8689–8694. doi:10.1073/pnas.1002720107

Foust, C. M., Preite, V., Schrey, A. W., Alvarez, M., Robertson, M. H., Verhoeven, K. J. F., et al. (2016). Genetic and Epigenetic Differences Associated with Environmental Gradients in Replicate Populations of Two Salt Marsh Perennials. *Mol. Ecol.* 25 (8), 1639–1652. doi:10.1111/mec.13522

Gibney, E. R., and Nolan, C. M. (2010). Epigenetics and Gene Expression. *Heredity* 105 (1), 4–13. doi:10.1038/hdy.2010.54

Giri, A. K., Midha, S., Banerjee, P., Agrawal, A., Mehdi, S. J., Dhingra, R., et al. (2016). Common Variants in CLDN2 and MORC4 Genes Confer Disease Susceptibility in Patients with Chronic Pancreatitis. *PLOS ONE* 11 (1), e0147345. doi:10.1371/journal.pone.0147345

Gonzalez, M., Villa, R., Villa, C., Gonzalez, V., Montano, M., Medina, G., et al. (2020). Inspection of Real and Imputed Genotypes Reveled 76 SNPs Associated to Rear Udder Height in Holstein Cattle. *J. Adv. Vet. Anim. Res.* 7 (2), 234. doi:10.5455/javar.2020.g415

Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., et al. (2014). Whole-genome Sequencing of Six Dog Breeds from Continuous Altitudes Reveals Adaptation to High-Altitude Hypoxia. *Genome Res.* 24, 1308–1315. doi:10.1101/gr.171876.113

Guo, L., Costanzo-Garvey, D. L., Smith, D. R., Neilsen, B. K., MacDonald, R. G., and Lewis, R. E. (2017). Kinase Suppressor of Ras 2 (KSR2) Expression in the Brain Regulates Energy Balance and Glucose Homeostasis. *Mol. Metab.* 6 (2), 194–205. doi:10.1016/j.molmet.2016.12.004

He, C., Wang, C., Chang, Z. H., Guo, B. L., Li, R., Yue, X. P., et al. (2011). AGPAT6 Polymorphism and its Association with Milk Traits of Dairy Goats. *Genet. Mol. Res.* 10 (4), 2747–2756. doi:10.4238/2011.November.4.8

Henriques, D., Wallberg, A., Chávez-Galarza, J., Johnston, J. S., Webster, M. T., and Pinto, M. A. (2018). Whole Genome SNP-Associated Signatures of Local Adaptation in Honeybees of the Iberian Peninsula. *Sci. Rep.* 8, 11145. doi:10.1038/s41598-018-29469-5

Herrera, C. M., and Bazaga, P. (2010). Epigenetic Differentiation and Relationship to Adaptive Genetic Divergence in Discrete Populations of the Violet Viola Cazorlensis. *New Phytol.* 187 (3), 867–876. doi:10.1111/j.1469-8137.2010.03298.x

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *Int. J. Climatol.* 25 (15), 1965–1978. doi:10.1002/joc.1276

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., et al. (2013). Epigenetic Memory at Embryonic Enhancers Identified in DNA Methylation Maps from Adult Mouse Tissues. *Nat. Genet.* 45, 1198–1206. doi:10.1038/ng.2746

Hussein, H. A., and Staufenbiel, R. (2012). Variations in Copper Concentration and Ceruloplasmin Activity of Dairy Cows in Relation to Lactation Stages with Regard to Ceruloplasmin to Copper Ratios. *Biol. Trace Elem. Res.* 146 (1), 47–52. doi:10.1007/s12011-011-9226-3

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

Killick, R., and Eckley, I. A. (2014). Changepoint: AnRPackage for Changepoint Analysis. *J. Stat. Soft.* 58 (3), 1–19. doi:10.18637/jss.v058.i03

Kim, G., Clarke, C. R., Larose, H., Tran, H. T., Haak, D. C., Zhang, L., et al. (2017). Herbicide Injury Induces DNA Methylome Alterations in Arabidopsis. *PeerJ* 5, e3560. doi:10.7717/peerj.3560

Krishnan, G., Bagath, M., Pragna, P., Vidya, M. K., Aleena, J., Archana, P. R., et al. (2017). "Mitigation of the Heat Stress Impact in Livestock Reproduction," In *Payan Carreira, R. (IntechOpen).* Theriogenology Available at: https://www.intechopen.com/chapters/55491.

Kwon, S. G., Hwang, J. H., Park, D. H., Kim, T. W., Kang, D. G., Kang, K. H., et al. (2016). Identification of Differentially Expressed Genes Associated with Litter Size in Berkshire Pig Placenta. *PLOS ONE* 11 (4), e0153311. doi:10.1371/journal.pone.0153311

Lambeth, J. D. (2004). NOX Enzymes and the Biology of Reactive Oxygen. *Nat. Rev. Immunol.* 4 (3), 181–189. doi:10.1038/nri1312

Lan, X., Cretney, E. C., Kropp, J., Khateeb, K., Berg, M. A., Peñagaricano, F., et al. (2013). Maternal Diet during Pregnancy Induces Gene Expression and DNA Methylation Changes in Fetal Tissues in Sheep. *Front. Genet.* 4, 49. doi:10.3389/fgene.2013.00049

Laodim, T., Elzo, M. A., Koonawootrittriron, S., Suwanasopee, T., and Jattawa, D. (2017). Identification of SNP Markers Associated with Milk and Fat Yields in Multibreed Dairy Cattle Using Two Genetic Group Structures. *Livestock Sci.* 206, 95–104. doi:10.1016/j.livsci.2017.10.015

Lea, A. J., Altmann, J., Alberts, S. C., and Tung, J. (2016). Resource Base Influences Genome-wide DNA Methylation Levels in Wild Baboons (*Papio cynocephalus*). *Mol. Ecol.* 25 (8), 1681–1696. doi:10.1111/mec.13436

Lee, B.-Y., Lee, K.-N., Lee, T., Park, J.-H., Kim, S.-M., Lee, H.-S., et al. (2015). Bovine Genome-wide Association Study for Genetic Elements to Resist the Infection of Foot-And-Mouth Disease in the Field. *Asian Australas. J. Anim. Sci.* 28 (2), 166–170. doi:10.5713/ajas.14.0383

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

Lienhard, M., Grimm, C., Morkel, M., Herwig, R., and Chavez, L. (2014). MEDIPS: Genome-wide Differential Coverage Analysis of Sequencing Data Derived from DNA Enrichment Experiments. *Bioinformatics* 30 (2), 284–286. doi:10.1093/bioinformatics/btt650

Liu, J., Zhang, L., Wang, D., Shen, H., Jiang, M., Mei, P., et al. (2003). Congenital Diaphragmatic Hernia, Kidney Agenesis and Cardiac Defects Associated with Slit3-Deficiency in Mice. *Mech. Dev.* 120 (9), 1059–1070. doi:10.1016/S0925-4773(03)00161-8

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., et al. (2014). DNA Methylome Profiling of Human Tissues Identifies Global and Tissue-specific Methylation Patterns. *Genome Biol.* 15 (4), 3248. doi:10.1186/gb-2014-15-4-r54

Lu, B., Jiang, Y. J., Man, M. Q., Brown, B., Elias, P. M., and Feingold, K. R. (2005). Expression and Regulation of 1-Acyl-Sn-Glycerol- 3-phosphate Acyltransferases in the Epidermis. *J. Lipid Res.* 46 (11), 2448–2457. doi:10.1194/jlr.M500258-JLR200

Lukić, B., Ferenčaković, M., Šalamon, D., Čačić, M., Orehovački, V., Iacolina, L., et al. (2020). Conservation Genomic Analysis of the Croatian Indigenous Black Slavonian and Turopolje Pig Breeds. *Front. Genet.* 11, 261. doi:10.3389/fgene.2020.00261

Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., and Hoekstra, H. E. (2010). Convergence in Pigmentation at Multiple Levels: Mutations, Genes and Function. *Phil. Trans. R. Soc. B* 365 (1552), 2439–2450. doi:10.1098/rstb.2010.0104

Marai, I. F. M., El-Darawany, A. A., Fadiel, A., and Abdel-Hafez, M. A. M. (2007). Physiological Traits as Affected by Heat Stress in Sheep-A Review. *Small Ruminant Res.* 71 (1–3), 1–12. doi:10.1016/j.smallrumres.2006.10.003

Martin, G. B., Rodger, J., and Blache, D. (2004). Nutritional and Environmental Effects on Reproduction in Small Ruminants. *Reprod. Fertil. Dev.* 16 (4), 491. doi:10.1071/RD04035

McDonald, K. W., McClure, C. J. W., Rolek, B. W., and Hill, G. E. (2012). Diversity of Birds in Eastern North America Shifts north with Global Warming. *Ecol. Evol.* 2 (12), 3052–3060. doi:10.1002/ece3.410

Meyerhoeffer, D. C., Wettemann, R. P., Coleman, S. W., and Wells, M. E. (1985). Reproductive Criteria of Beef Bulls during and after Exposure to Increased Ambient Temperature. *J. Anim. Sci.* 60 (2), 352–357. doi:10.2527/jas1985.602352x

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P. D. (2019). PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools. *Nucleic Acids Res.* 47 (D1), D419–D426. doi:10.1093/nar/gky1038

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

Mommersteeg, M. T. M., Yeh, M. L., Parnavelas, J. G., and Andrews, W. D. (2015). Disrupted Slit-Robo Signalling Results in Membranous Ventricular Septum Defects and Bicuspid Aortic Valves. *Cardiovasc. Res.* 106 (1), 55–66. doi:10.1093/cvr/cvv040

Monteiro, A. P. A., Tao, S., Thompson, I. M. T., and Dahl, G. E. (2016). In Utero heat Stress Decreases Calf Survival and Performance through the First Lactation. *J. Dairy Sci.* 99 (10), 8443–8450. doi:10.3168/jds.2016-11072

Mota, L. F. M., Lopes, F. B., Fernandes Júnior, G. A., Rosa, G. J. M., Magalhães, A. F. B., Carvalheiro, R., et al. (2020). Genome-wide Scan Highlights the Role of Candidate Genes on Phenotypic Plasticity for Age at First Calving in Nellore Heifers. *Sci. Rep.* 10 (1), 6481. doi:10.1038/s41598-020-63516-4

Namous, H., Peñagaricano, F., Del Corvo, M., Capra, E., Thomas, D. L., Stella, A., et al. (2018). Integrative Analysis of Methylomic and Transcriptomic Data in Fetal Sheep Muscle Tissues in Response to Maternal Diet during Pregnancy. *BMC Genomics* 19 (1), 123. doi:10.1186/s12864-018-4509-0

Ouellet, V., Laporta, J., and Dahl, G. E. (2020). Late Gestation Heat Stress in Dairy Cows: Effects on Dam and Daughter. *Theriogenology* 150, 471–479. doi:10.1016/j.theriogenology.2020.03.011

Peñagaricano, F., Wang, X., Rosa, G. J., Radunz, A. E., and Khatib, H. (2014). Maternal Nutrition Induces Gene Expression Changes in Fetal Muscle and Adipose Tissues in Sheep. *BMC Genomics* 15 (1), 1034. doi:10.1186/1471-2164-15-1034

Peterson, D. G., Matitashvili, E. A., and Bauman, D. E. (2003). Diet-Induced Milk Fat Depression in Dairy Cows Results in Increased Trans-10, Cis-12 CLA in Milk Fat and Coordinate Suppression of mRNA Abundance for Mammary Enzymes Involved in Milk Fat Synthesis. *J. Nutr.* 133 (10), 3098–3102. doi:10.1093/jn/133.10.3098

Pierce, C. F., Speidel, S. E., Coleman, S. J., Enns, R. M., Bailey, D. W., Medrano, J. F., et al. (2020). Genome-wide Association Studies of Beef Cow Terrain-Use Traits Using Bayesian Multiple-SNP Regression. *Livestock Sci.* 232, 103900. doi:10.1016/j.livsci.2019.103900

Pimentel, E. C. G., Bauersachs, S., Tietze, M., Simianer, H., Tetens, J., Thaller, G., et al. (2011). Exploration of Relationships between Production and Fertility Traits in Dairy Cattle via Association Studies of SNPs within Candidate Genes Derived by Expression Profiling. *Anim. Genet.* 42 (3), 251–262. doi:10.1111/j.1365-2052.2010.02148.x

Ponsuksili, S., Murani, E., Schwerin, M., Schellander, K., Tesfaye, D., and Wimmers, K. (2012). Gene Expression and DNA-Methylation of Bovine Pretransfer Endometrium Depending on its Receptivity after In Vitro-Produced Embryo Transfer. *PLOS ONE* 7 (8), e42402. doi:10.1371/journal.pone.0042402

Prasad, S. S., Garg, A., and Agarwal, A. K. (2011). Enzymatic Activities of the Human AGPAT Isoform 3 and Isoform 5: Localization of AGPAT5 to Mitochondria. *J. Lipid Res.* 52 (3), 451–462. doi:10.1194/jlr.M007575

Puig-Oliveras, A., Ballester, M., Corominas, J., Revilla, M., Estellé, J., Fernández, A. I., et al. (2014). A Co-association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. *PLOS ONE* 9 (12), e114862. doi:10.1371/journal.pone.0114862

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

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I., Hellemans, B., Maes, G. E., et al. (2017). Adaptive and Non-adaptive Divergence in a Common Landscape. *Nat. Commun.* 8 (1), 267. doi:10.1038/s41467-017-00256-6

Razzaghi, A., Valizadeh, R., Naserian, A. A., Mesgaran, M. D., Carpenter, A. J., and Ghaffari, M. H. (2016). Effect of Dietary Sugar Concentration and sunflower Seed Supplementation on Lactation Performance, Ruminal Fermentation, Milk Fatty Acid Profile, and Blood Metabolites of Dairy Cows. *J. Dairy Sci.* 99 (5), 3539–3548. doi:10.3168/jds.2015-10565

Richards, E. J. (2011). Natural Epigenetic Variation in Plant Species: a View from the Field. *Curr. Opin. Plant Biol.* 14 (2), 204–209. doi:10.1016/j.pbi.2011.03.009

Robinson, J. J., Ashworth, C. J., Rooke, J. A., Mitchell, L. M., and McEvoy, T. G. (2006). Nutrition and Fertility in Ruminant Livestock. *Anim. Feed Sci. Tech.* 126 (3), 259–276. doi:10.1016/j.anifeedsci.2005.08.006

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26 (1), 139–140. doi:10.1093/bioinformatics/btp616

Russo, V., Fontanesi, L., Davoli, R., Nanni Costa, L., Cagnazzo, M., Buttazzoni, L., et al. (2002). Investigation of Candidate Genes for Meat Quality in Dry-Cured Ham Production: the Porcine Cathepsin B (CTSB) and Cystatin B (CSTB) Genes. *Anim. Genet.* 33 (2), 123–131. doi:10.1046/j.1365-2052.2002.00835.x

Salleh, M. S., Mazzoni, G., Höglund, J. K., Olijhoek, D. W., Lund, P., Løvendahl, P., et al. (2017). RNA-seq Transcriptomics and Pathway Analyses Reveal Potential Regulatory Genes and Molecular Mechanisms in High- and Low-Residual Feed Intake in Nordic Dairy Cattle. *BMC Genomics* 18 (1), 258. doi:10.1186/s12864-017-3622-9

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological Genomics of Local Adaptation. *Nat. Rev. Genet.* 14 (11), 807–820. doi:10.1038/nrg3522

Sejian, V., Maurya, V. P., and Naqvi, S. M. K. (2010). Adaptability and Growth of Malpura Ewes Subjected to thermal and Nutritional Stress. *Trop. Anim. Health Prod.* 42 (8), 1763–1770. doi:10.1007/s11250-010-9633-z

Sevane, N., Martínez, R., and Bruford, M. W. (2018). Genome-wide Differential DNA Methylation in Tropically Adapted Creole Cattle and Their Iberian Ancestors. *Anim. Genet.* 50 (1), 15–26. doi:10.1111/age.12731

Shin, D., Lee, C., Park, K.-D., Kim, H., and Cho, K.-h. (2017). Genome-association Analysis of Korean Holstein Milk Traits Using Genomic Estimated Breeding Value. *Asian-australas J. Anim. Sci.* 30 (3), 309–319. doi:10.5713/ajas.15.0608

Skibiel, A. L., Dado-Senn, B., Fabris, T. F., Dahl, G. E., and Laporta, J. (2018). In Utero exposure to thermal Stress Has Long-Term Effects on Mammary Gland Microstructure and Function in Dairy Cattle. *PLOS ONE* 13 (10), e0206046. doi:10.1371/journal.pone.0206046

Skibiel, A. L., Peñagaricano, F., Amorín, R., Ahmed, B. M., Dahl, G. E., and Laporta, J. (2018). In Utero Heat Stress Alters the Offspring Epigenome. *Sci. Rep.* 8 (1), 14609. doi:10.1038/s41598-018-32975-1

Storey, J. D., Bass, A. J., Dabney, A., and Robinson, D. (2015). Qvalue: Q-Value Estimation for False Discovery Rate Control. R package version 2.8.0. . Http://Github.Com/Jdstorey/Qvalue.

Sugimoto, M., Fujikawa, A., Womack, J. E., and Sugimoto, Y. (2006). Evidence that Bovine Forebrain Embryonic Zinc finger-like Gene Influences Immune Response Associated with Mastitis Resistance. *Proc. Natl. Acad. Sci.* 103 (17), 6454–6459. doi:10.1073/pnas.0601015103

Suzuki, M. M., and Bird, A. (2008). DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nat. Rev. Genet.* 9 (6), 465–476. doi:10.1038/nrg2341

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are Cattle, Sheep, and Goats Endangered Species? *Mol. Ecol.* 17 (1), 275–284. doi:10.1111/j.1365-294X.2007.03475.x

Takeuchi, K., and Reue, K. (2009). Biochemistry, Physiology, and Genetics of GPAT, AGPAT, and Lipin Enzymes in Triglyceride Synthesis. *Am. J. Physiology-Endocrinology Metab.* 296 (6), E1195–E1209. doi:10.1152/ajpendo.90958.2008

Tao, S., Bubolz, J. W., do Amaral, B. C., Thompson, I. M., Hayen, M. J., Johnson, S. E., et al. (2011). Effect of Heat Stress during the Dry Period on Mammary Gland Development. *J. Dairy Sci.* 94 (12), 5976–5986. doi:10.3168/jds.2011-4329

Vallée, M., Beaudry, D., Roberge, C., Matte, J. J., Blouin, R., and Palin, M.-F. (2003). Isolation of Differentially Expressed Genes in Conceptuses and Endometrial Tissue of Sows in Early Gestation1. *Biol. Reprod.* 69 (5), 1697–1706. doi:10.1095/biolreprod.103.019307

Waggott, D., Haider, S., and Boutros, P. C. (2017). Bedr: Genomic Region Processing Using Tools Such as "BEDTools", "BEDOPS" and "Tabix". R package version 1.0.4. Https://CRAN.R-Project.Org/Package=bedr.

Wang, K. H., Brose, K., Arnott, D., Kidd, T., Goodman, C. S., Henzel, W., et al. (1999). Biochemical Purification of a Mammalian Slit Protein as a Positive Regulator of Sensory Axon Elongation and Branching. *Cell* 96 (6), 771–784. doi:10.1016/S0092-8674(00)80588-7

Wang, X., Lan, X., Radunz, A. E., and Khatib, H. (2015). Maternal Nutrition during Pregnancy Is Associated with Differential Expression of Imprinted Genes and DNA Methyltranfereases in Muscle of Beef Cattle Offspring1. *J. Anim. Sci.* 93 (1), 35–40. doi:10.2527/jas.2014-8148

Weir, B. S., and Cockerham, C. C. (1984). Estimating F -Statistics for the Analysis of Population Structure. *Evolution* 38 (6), 1358–1370. doi:10.2307/240864110.1111/j.1558-5646.1984.tb05657.x

Xiang, R., MacLeod, I. M., Bolormaa, S., and Goddard, M. E. (2017). Genome-wide Comparative Analyses of Correlated and Uncorrelated Phenotypes Identify Major Pleiotropic Variants in Dairy Cattle. *Sci. Rep.* 7 (1), 9248. doi:10.1038/s41598-017-09788-9

Yang, J., Jiang, J., Liu, X., Wang, H., Guo, G., Zhang, Q., et al. (2016). Differential Expression of Genes in Milk of Dairy Cattle during Lactation. *Anim. Genet.* 47 (2), 174–180. doi:10.1111/age.12394

Yang, J., Li, W.-R., Lv, F.-H., He, S.-G., Tian, S.-L., Peng, W.-F., et al. (2016). Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol. Biol. Evol.* 33 (10), 2576–2592. doi:10.1093/molbev/msw129

Zeder, M. A. (2008). Domestication and Early Agriculture in the Mediterranean Basin: Origins, Diffusion, and Impact. *Proc. Natl. Acad. Sci.* 105 (33), 11597–11604. doi:10.1073/pnas.0801317105

Zhang, X., Zhang, S., Ma, L., Jiang, E., Xu, H., Chen, R., et al. (2017). Reduced Representation Bisulfite Sequencing (RRBS) of Dairy Goat Mammary Glands Reveals DNA Methylation Profiles of Integrated Genome-wide and Critical Milk-Related Genes. *Oncotarget* 8 (70), 115326–115344. doi:10.18632/oncotarget.23260

Zhao, C., Gui, L., Li, Y., Plath, M., and Zan, L. (2015). Associations between Allelic Polymorphism of the BMP Binding Endothelial Regulator and Phenotypic Variation of Cattle. *Mol. Cell Probes* 29 (6), 358–364. doi:10.1016/j.mcp.2015.09.007

Zhou, D. a., Wei, Z., Deng, S., Wang, T., Zai, M., Wang, H., et al. (2013). SASH1 Regulates Melanocyte Transepithelial Migration through a Novel Gαs-SASH1-IQGAP1-E-Cadherin Dependent Pathway. *Cell Signal.* 25 (6), 1526–1538. doi:10.1016/j.cellsig.2012.12.025

Zhou, X., Lindsay, H., and Robinson, M. D. (2014). Robustly Detecting Differential Expression in RNA Sequencing Data Using Observation Weights. *Nucleic Acids Res.* 42 (11), e91. doi:10.1093/nar/gku310

Zoldoš, V., Biruš, I., Muratović, E., Šatović, Z., Vojta, A., Robin, O., et al. (2018). Epigenetic Differentiation of Natural Populations of Lilium Bosniacum Associated with Contrasting Habitat Conditions. *Genome Biol. Evol.* 10 (1), 291–303. doi:10.1093/gbe/evy010

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: 15 October 2021
doi: 10.3389/fgene.2021.759831

Ancient Mitogenomes Provide New Insights into the Origin and Early Introduction of Chinese Domestic Donkeys

Linying Wang[1], Guilian Sheng[2,3], Michaela Preick[4], Songmei Hu[5], Tao Deng[6], Ulrike H. Taron[4], Axel Barlow[4,7], Jiaming Hu[8], Bo Xiao[8], Guojiang Sun[1], Shiwen Song[2], Xindong Hou[2,3], Xulong Lai[3,8], Michael Hofreiter[4] and Junxia Yuan[1,3]**

[1]Faculty of Materials Science and Chemistry, China University of Geosciences, Wuhan, China, [2]School of Environmental Studies, China University of Geosciences, Wuhan, China, [3]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, China, [4]Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [5]Shaanxi Provincial Institute of Archaeology, Xi'an, China, [6]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Beijing, China, [7]School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom, [8]School of Earth Sciences, China University of Geosciences, Wuhan, China

OPEN ACCESS

Edited by:
Hai Xiang,
Foshan University, China

Reviewed by:
Takahiro Yonezawa,
Tokyo University of Agriculture, Japan
Qianjun Zhao,
Institute of Animal Sciences (CAAS), China

***Correspondence:**
Michael Hofreiter
michael.hofreiter@uni-potsdam.de
Junxia Yuan
yuanjx@cug.edu.cn

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

Received: *17 August 2021*
Accepted: *30 September 2021*
Published: *15 October 2021*

Citation:
Wang L, Sheng G, Preick M, Hu S, Deng T, Taron UH, Barlow A, Hu J, Xiao B, Sun G, Song S, Hou X, Lai X, Hofreiter M and Yuan J (2021) Ancient Mitogenomes Provide New Insights into the Origin and Early Introduction of Chinese Domestic Donkeys. Front. Genet. 12:759831. doi: 10.3389/fgene.2021.759831

Both molecular data and archaeological evidence strongly support an African origin for the domestic donkey. Recent genetic studies further suggest that there were two distinct maternal lineages involved in its initial domestication. However, the exact introduction time and the dispersal process of domestic donkeys into ancient China are still unresolved. To address these questions, we retrieved three near-complete mitochondrial genomes from donkey specimens excavated from Gaoling County, Shaanxi Province, and Linxia Basin, Gansu Province, China, dated at 2,349-2,301, 469-311, and 2,160-2,004 cal. BP, respectively. Maximum-likelihood and Bayesian phylogenetic analyses reveal that the two older samples fall into the two different main lineages (i.e., clade I and clade II) of the domestic donkey, suggesting that the two donkey maternal lineages had been introduced into Midwestern China at least at the opening of Silk Road (approximately the first century BC). Bayesian analysis shows that the split of the two donkey maternal lineages is dated at 0.323 Ma (95% CI: 0.583–0.191 Ma) using root-tip dating calibrations based on near-complete mitogenomes, supporting the hypothesis that modern domestic donkeys go back to at least two independent domestication events. Moreover, Bayesian skyline plot analyses indicate an apparent female population increase between 5,000 and 2,500 years ago for clade I followed by a stable population size to the present day. In contrast, clade II keeps a relatively stable population size over the past 5,000 years. Overall, our study provides new insights into the early domestication history of Chinese domestic donkeys.

Keywords: Chinese domestic donkeys, ancient DNA, mitochondrial genome, maternal lineage, divergence time

INTRODUCTION

The domestication of the donkey (*Equus asinus*) is a vital event in human history, which played a significant role in the development of human civilization (Liu et al., 2010; Bai et al., 2017; Wang et al., 2020). Donkey deeply transformed ancient societies and land-based transport in Africa and Eurasia, contributed to the growth of the early Egyptian State, and allowed the development of mobile pastoralism and ancient overland trade routes (Denbow et al., 1993; Denham and Iriarte, 2007; Rossel et al., 2008; Kimura et al., 2010). The domestication of the donkey therefore probably indicates a major cultural shift away from sedentary, agrarian life-styles towards more migration and trade in ancient times (Beja-Pereira et al., 2004; Han et al., 2014). However, compared with the other domesticated species of the genus *Equus*, i.e., the horse, the domestic donkey is greatly underrepresented in the scientific literature (Blench, 2000; Lu et al., 2008; Ma et al., 2020). In the last decades with the promotion of agricultural mechanization and the rapid development of the transportation industry in modern society, the role of domestic donkeys as a means of transportation is decreasing and the number of donkeys has declined greatly (Xie, 1987). Despite these developments, currently donkeys still remain an essential means of transport for people living in mountain areas, deserts, and underdeveloped regions of the world (Blench, 2000; Smith and Pearson, 2005; Kimura et al., 2010; Ma et al., 2020).

Archeological evidence suggests an African origin for the donkey (Epstein, 1971; Clutton-Brock, 1992; Rossel et al., 2008). The earliest domestic donkey remains, 5,000-year-old ass skeletons, have been excavated from an early pharaonic mortuary complex at Abydos, Middle Egypt, which exhibit a range of osteopathologies consistent with load carrying (Rossel et al., 2008). However, it is often difficult to determine whether the remains from early phases of animal domestication originate from animals that have been domesticated or not (Peters and von der Driesc, 1997; Rossel et al., 2008). Compared with the horse, donkey remains are relatively rare in archaeological sites and are not easily distinguished from the former based on morphological characters alone (Han et al., 2014). Therefore, the available morphological evidence provides limited information about the timing and location of donkey domestication.

Mitochondrial and nuclear DNA have revealed that domestic donkeys originated from African wild asses (Ivankovic et al., 2002; Beja-Pereira et al., 2004; Kimura et al., 2010; Ma et al., 2020; Wang et al., 2020). Mitochondrial DNA studies showed that domestic donkeys harbored two distinct lineages (i.e., clade I and clade II). Clade I (Nubian lineage) contains domestic donkeys and the Nubian wild ass (*Equus africanus africanus*), while clade II (unknown origin) probably derived from a now extinct African wild ass population, which might have been close to the Somali wild ass (*Equus africanus somaliensis*) (Kimura et al., 2010, 2013; Ma et al., 2020). Wang et al. (2020) recently analyzed 126 modern domestic donkey nuclear genomes. Their *D-statistic* analysis showed an African domestication of donkeys, consistent with the results from mitochondrial DNA, and indicated its subsequent spread to Europe and Asia. In addition, the principal component analysis (PCA) suggested that domestic donkeys are divided into three main clusters on the nuclear level, i.e., a Tropical Africa cluster, a North Africa and Eurasia cluster and an Australia cluster. Wang et al. (2020) finally found that domestic donkeys showed reduced levels of Y chromosome variability, which might indicate a discordance of paternal and maternal histories of donkeys, similar to the domestic horse (Lindgren et al., 2004; Lippold et al., 2011).

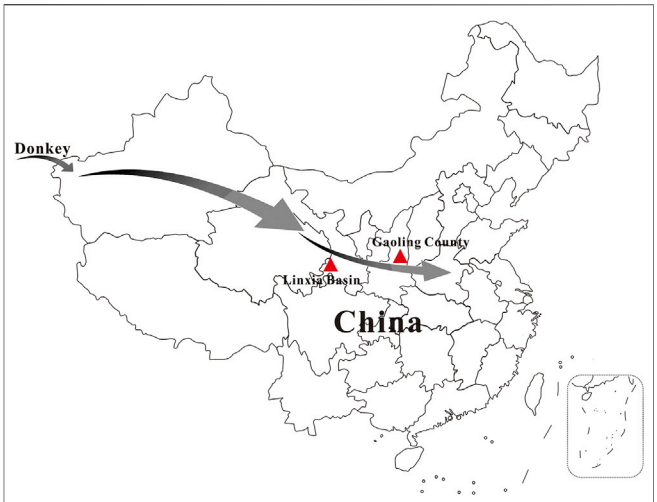

FIGURE 1 | Sampling locations of Chinese ancient donkeys analyzed in this study. Sample LXH1 was excavated from Linxia Basin, specimens SG1 and SG3 were collected from Gaoling County. The arrows indicate possible dispersal route of domestic donkey to ancient China.

The history of domestic donkey in China dates back more than 4,000 years (Zheng, 1980; Xie, 1987; Chen et al., 2010). According to literature records, domestic donkeys were bred in present-day Shache County, Xinjiang Uygur Autonomous Region, Northwestern China as early as in the Yin and Shang Dynasties (1,300-1,046 BC) (Yang and Hong, 1989). Regarding the origin of the Chinese domestic donkey, there are two main views: 1) Due to morphological similarities to Asian wild asses, e.g., in fur color, some researchers believed that Chinese domestic donkeys might have originated from Mongolian wild ass (*Equus hemionus*) (Xie, 1987; Liu et al., 2010). 2) In contrast, genetic studies suggested that Chinese domestic donkeys originate from African wild asses (Sun et al., 2007; Han et al., 2014; Ma et al., 2020). Wang et al. (2020) analyzed mitochondrial DNA and nuclear genomes of Chinese local donkey breeds and revealed that Chinese donkeys are closer to African wild asses than to Asian wild asses (*E. hemionus* and *Equus kiang*). So far, most molecular studies on Chinese domestic donkey focus on modern specimens. The only public report on genetic analyses of Chinese ancient donkeys has been presented by Han et al. (2014), but only mitochondrial DNA D-loop and cytochrome *b* gene fragments were obtained, with the dates of the analyzed samples ranging between 1,200–550 years before present (BP). Han et al. (2014) found that the ancient specimens represent both donkey mitochondrial maternal lineages, i.e., the Nubian lineage (clade Ⅰ) and the lineage of unknown origin (clade Ⅱ). Unfortunately, due to a lack of

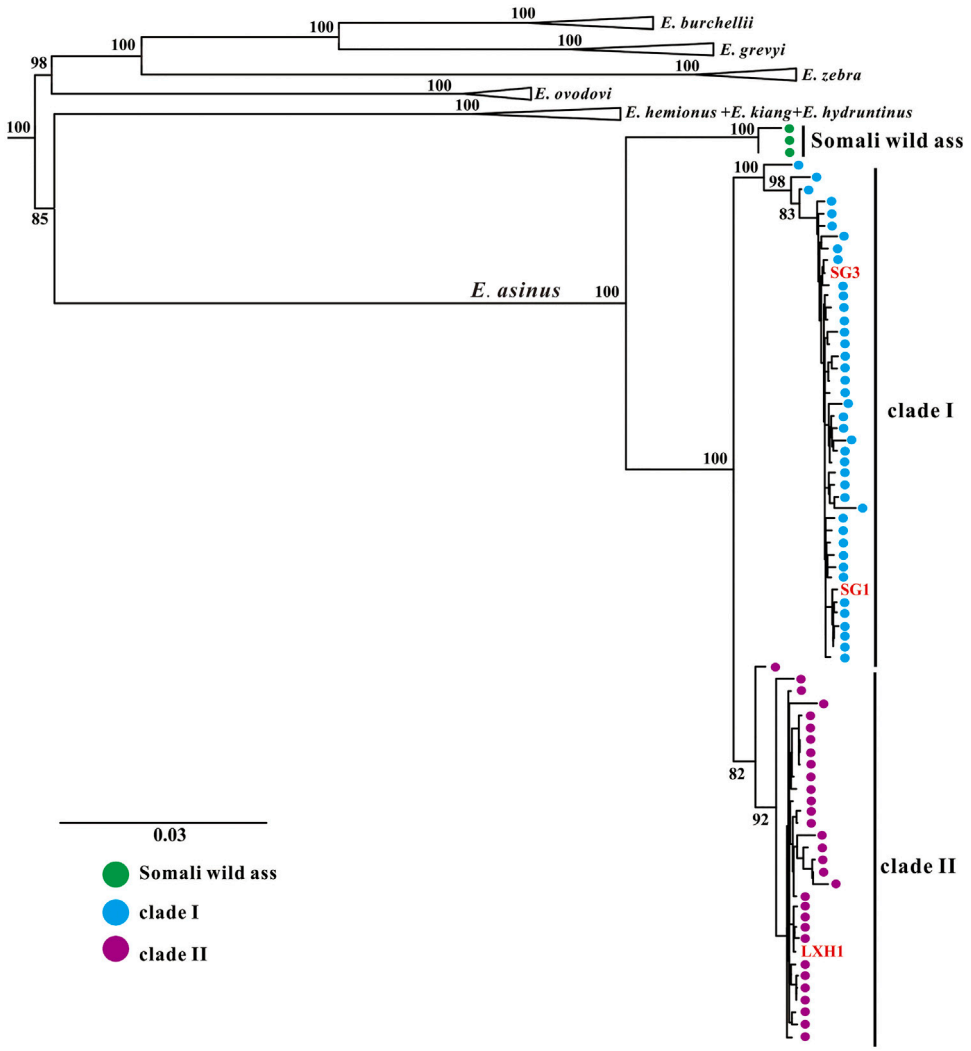

FIGURE 2 | Maximum-likelihood phylogenetic tree of the genus Equus based on near-complete mitochondrial genomes. *E. caballus* was selected as outgroup (not shown here). Branch labels show bootstrap values derived from 1,000 replications.

genetic information from earlier Chinese donkeys, we know little about the initial dispersal process of donkeys into China.

In this study, we retrieved three near-complete mitogenomes from archaeological donkey specimens excavated from Midwestern China, investigated the phylogenetic status of the analyzed individuals and estimated the divergence time of the two donkey lineages. We also carried out a Bayesian skyline plot (BSP) analysis to assess donkey population dynamics. Overall, our study provides new insights into the early domestication history of Chinese donkeys.

RESULTS AND DISCUSSION

Three ancient donkey tooth samples are included in this study. Two specimens (SG1 and SG3) were excavated from Gaoling County, Shaanxi Province, China and one sample (LXH1) was collected from Linxia Basin, Gansu Province, China (**Figure 1**). ^{14}C dating of the samples was performed by accelerator mass spectrometry (AMS) at the Archaeological Geochronology Laboratory of Peking University (PKUAMS, China). Calibration was done using IntCal13 (Reimer et al., 2013), yielding ages of 2,349-2,301 (SG1), 469-311 (SG3), and 2,160-2,004 (LXH1) cal. BP, respectively. Detailed information on the samples is listed in **Supplementary Table S1**. Using hybridization capture technology and an *E. asinus* mitogenome (GenBank No. X97337) as reference, we successfully retrieved three near-complete mitochondrial genomes from the analyzed samples with a mean depth of 79.5-, 37.2- and 36.8-fold, respectively (**Supplementary Table S1**). Mitochondrial DNA (mtDNA) fragments show damage patterns characteristic for ancient DNA (Briggs et al., 2007) (**Supplementary Figure S1**), supporting the obtained sequences as derived from authentic ancient DNA.

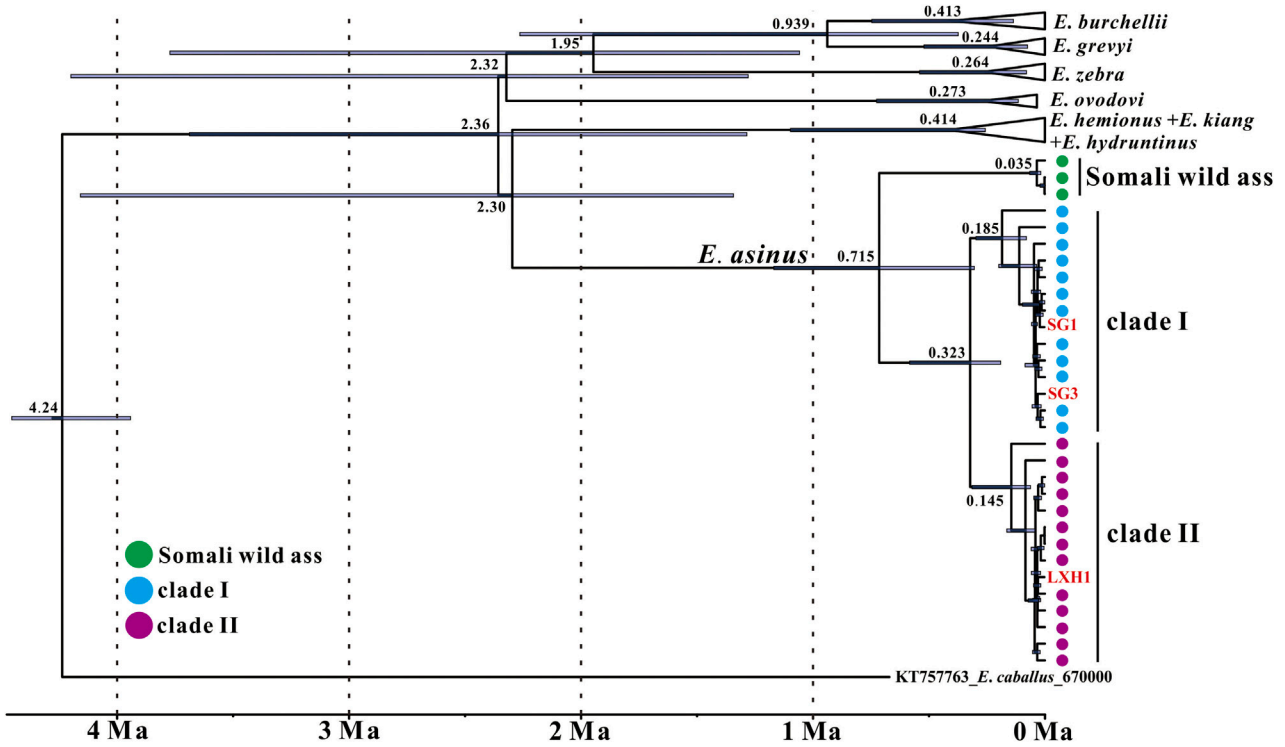

FIGURE 3 | Maximum clade credibility tree of the genus Equus as recovered with BEAST based on near-complete mitochondrial genomes. Node heights are centered on the median posterior age estimates with blue bars showing 95% credibility intervals of the divergence times. Tip dates of samples used in the molecular clock analysis are listed in **Supplementary Table S2**.

Early Dispersal of Domestic Donkey to Ancient China

We reconstructed phylogenetic trees using these newly obtained mitogenomes together with *Equus* sequences from GenBank. Both maximum-likelihood (ML) and Bayesian methods strongly support that all *E. asinus* individuals form a separate clade within non-caballine horses (**Figures 2**, **3**). The *E. asinus* branch is further divided into three clades, i.e., one Somali wild ass clade, which diverges from the *E. asinus* branch first, and two domestic donkey clades (clade I and clade II), containing modern domesticated donkeys, Nubian wild asses and our ancient individuals (**Figures 2**, **3**). Our results are consistent with previous studies (Kimura et al., 2010; Kimura et al., 2013; Han et al., 2014; Ma et al., 2020; Wang et al., 2020). Interestingly, the three samples investigated in this study fall into different clades of domestic donkey, i.e., specimens SG1 and SG3 cluster within clade I (Nubian lineage), while LXH1 groups into clade II (with no extant wild representatives). Both of these two donkey clades are distant from the Asiatic wild asses (*E. kiang* and *E. hemionus*), which reveals that the maternal origin of Chinese domestic donkeys is most likely from African wild asses instead of Asian wild asses, as suggested by previous analyses (Han et al., 2014; Ma et al., 2020; Wang et al., 2020).

According to the fossil record (Epstein, 1971; Clutton-Brock, 1992; Rossel et al., 2008) and molecular data (Ivankovic et al., 2002; Beja-Pereira et al., 2004; Sun et al., 2007; Kimura et al., 2010, 2013; Han et al., 2014; Ma et al., 2020; Wang et al., 2020), African wild asses are the most likely ancestor of the domestic donkey. It is commonly believed that donkeys first dispersed from Africa to Northwest China through Central Asia about 4,000 years ago (Xie, 1987; Lu et al., 2008). If correct, this means that domestic donkeys had spread into Northwestern China before the establishment of the Han Dynasty (about the second century BC) (Han et al., 2014). After the Southern and Northern Dynasties (420-589 AD), people from Central China also started raising and breeding donkeys, and its population size gradually increased since then (Yang and Hong, 1989).

Two out of three ancient samples in this study have been dated at similar ages (i.e., 2,349-2,301 cal. BP for SG1 and 2,160-2,004 cal. BP for LXH1), yet they fall into different donkey clades (**Figures 2**, **3**). The results demonstrate that the two donkey maternal lineages had been introduced into China at least at the beginning of Han Dynasty, i.e. around the opening of the Silk Road (about the first century BC). Unfortunately, due to a lack of earlier samples, our knowledge about when and how the two donkey maternal lineages were introduced to China is very limited so far, and further research is needed to explore these questions.

Divergence Time of Different *E. asinus* Lineages

We carried out a mitogenome relaxed molecular clock analysis to investigate the coalescence times among *E. asinus* lineages, using

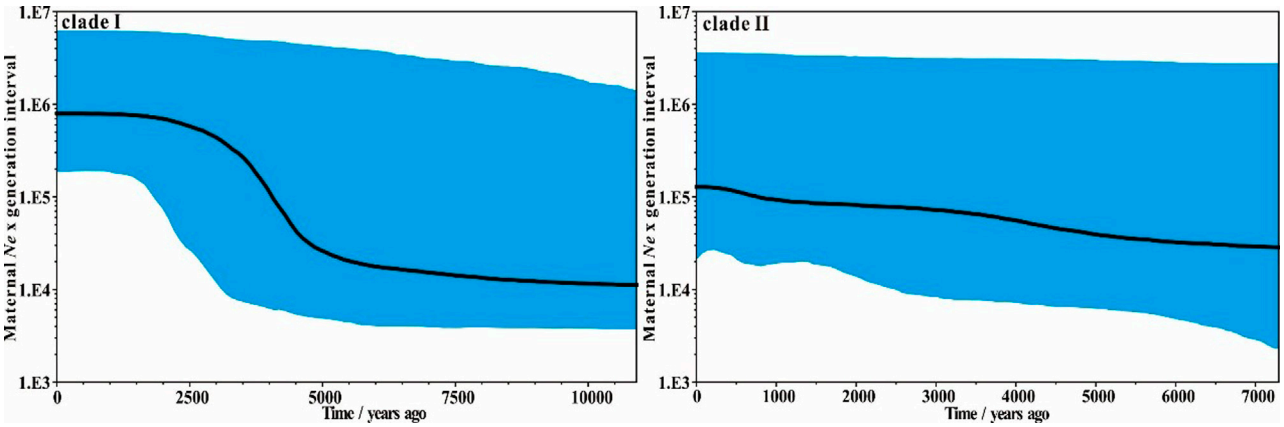

FIGURE 4 | Bayesian skyline plot of the two different domestic donkey lineages based on near-complete mitochondrial genomes. Black line indicates median female N_e change over time, while the shaded blue area indicates the 95% credibility interval.

root-and-tip dating calibrations in BEAST (**Figure 3**). Our analysis reveals that the divergence time between Somali wild ass and domestic donkey is at 0.715 Ma (95% CI: 1.169–0.305 Ma), and the split of the two domestic donkey maternal lineages is dated at 0.323 Ma (95% CI: 0.583–0.191 Ma). The times of the most recent common ancestor (TMRCA) of clade Ⅰ and clade Ⅱ are 0.185 Ma and 0.145 Ma, respectively.

Beja-Pereira et al. (2004) also estimated the divergence time of the two donkey maternal clades, and suggested a somewhat more ancient divergence in the time range of 0.910–0.303 Ma. Our point estimate (0.323 Ma) is close to the lower limit of that predicted by Beja-Pereira et al. (2004), while the confidence intervals of the two estimates overlap widely. The difference may at least partially be due to the fact that we use near-complete mitochondrial genomes to calculate the divergence time, while only cyt *b* gene sequences were included in Beja-Pereira et al. (2004). Another possible reason is that different calibration methods are implemented. Beja-Pereira *et al.* (2004) chose the previously estimated divergence time between horse and donkey (10–8 Ma) as the calibration node (Xu et al., 1996), whereas we considered the 4.5–4.0 Ma from Orlando et al. (2013) for the TMRCA of all extant *Equus* representatives and the median radiocarbon date or strata age of specimens as calibration points. Although our estimate is younger, our results together with the previous study (Beja-Pereira et al., 2004) suggest that the split of the two donkey lineages dates much earlier than its first known domestication date.

Our estimation of the TMRCA of the donkey clade Ⅰ and clade Ⅱ maternal lineages are 0.185 Ma and 0.145 Ma, respectively. This is much younger than the estimates of Kimura et al. (2010), who analyzed mitochondrial D-loop sequences of historic Nubian wild ass, Somali wild ass and ancient donkey, arriving at ages for clade I of 0.406 Ma, clade II of 0.334 Ma and the Somali wild ass clade of 0.360 Ma, respectively. However, even our younger estimates predate the domestication of donkeys by a large margin, suggesting that in both clades multiple wild lineages were incorporated into the domestic gene pool.

Demographic Dynamics of Domestic Donkeys

The Bayesian skyline plot analyses indicate an apparent population expansion between 5,000 and 2,500 years ago for clade I, following by a relatively stable population to the present day (**Figure 4**). However, compared to clade Ⅰ, clade II keeps a relatively stable population size overall, only showing a slight population increase during the past 5,000 years, which is similar to the result obtained by Ma et al. (2020).

Domestication of animals is generally accompanied by population expansion, as seen e.g. in horse (Fages et al., 2019), goat (Al-Araimi et al., 2017) and camel (Chen et al., 2019). The donkey population expansions of clade I and clade II may also relate to their domestication. If the Nubian lineage (clade I) and the lineage of unknown origin (clade II) were domesticated simultaneously, a similar demographic history may be expected from them. Ma et al. (2020) also assessed the population dynamics of the two domestic donkey lineages based on modern donkey mitogenomes. Their analyses suggested that clade II had a constant effective population size during most of its history, while clade I experienced a rapid population expansion starting approximately 8,000 years ago. Our estimates are overall similar to the estimates by Ma et al. (2020). In addition, Wang et al. (2020) found that there were no obvious differences in effective population size of Tropical African donkeys and North African and Eurasian donkeys, proposing that these donkeys were probably derived from the domestication of one common ancestral group. However, they noted that their analyses did not allow determining whether donkeys were domesticated at a single or multiple locations. Thus, currently, the history of donkey domestication remains at least partially unresolved. Therefore, ancient DNA is key to explore this essential question, as shown for other domesticated species such as goat (Daly et al., 2018) or cattle (Verdugo et al., 2019). Our estimates confirm that the two donkey lineages experienced somewhat different past demographic expansion histories. Together with the split time of the two clades, our

results at least suggest that donkeys might have undergone at least two independent domestication events.

METHODS

DNA Extraction

We performed DNA extraction in a dedicated ancient DNA laboratory at the University of Potsdam, following the protocol of Dabney et al. (2013) with several modifications as described in Yuan et al. (2020). The tooth samples were ground into fine powder with mortar and pestle, and for each sample about 50 mg powder was added to 1 ml extraction buffer containing 0.45 M EDTA and 0.25 mg/ ml proteinase K. The tooth powder was resuspended by vortexing and incubated overnight at 37°C under constant rotation. Next, we centrifuged the samples for 2 min at 13,300 rpm to pellet the powder, followed by adding the supernatant to 13 ml binding buffer. Then the mixtures were poured into the binding apparatus reservoirs, followed by centrifugation for 4 min at 1,500 rpm. We added 650 μL PE buffer to the silica membrane in the washing step and then carried out centrifugation again at 1,500 rpm for 4 min. DNA was eluted twice by adding 12.5 μL TET buffer each time to the silica membrane, incubating for 10 min at room temperature and centrifugation at 13,300 rpm for 30 s each time. In total, we obtained 25 μL DNA extract. In addition, an extraction blank was included alongside the samples.

Library Construction and Hybridization Capture

Single-stranded DNA libraries were prepared by using 20 μL DNA extract for each sample, following the protocol described by Gansauge and Meyer (2013) with the modifications in Yuan et al. (2020). The amount of Circligase II was reduced to 2 μL (100 U/ μL) in the ligation step of the first adapter; accordingly, incubation time was increased to overnight at 60°C. Hybridization capture of the complete mitochondrial genome was carried out following previously published procedures (González-Fortes and Paijmans, 2019). Baits were prepared as in the following protocol. First, total DNA was extracted from a modern horse sample and the mitochondrial genome was amplified using four overlapping long range PCR (LR-PCR) primer pairs (Vilstrup et al., 2013; Yuan et al., 2019). Second, LR-PCR products were sheared, blunt-end repaired and ligated to biotinylated adapters. Subsequently, two rounds of hybridization capture were carried out to improve the enrichment rate as described in Yuan et al. (2019). The enriched libraries were purified using Minelute columns (Qiagen) and DNA was eluted twice by adding 10 μL EB buffer each time. Concentration and fragment size of the DNA were measured on a Qubit 2.0 and a TapeStation 2200 (Agilent). Finally, the enriched libraries were pooled and sequenced on 75 cycle single-end runs on the Illumina NextSeq 500 sequencing platform, following the procedures described in Paijmans et al. (2017). Blanks were also included in single-stranded library preparation and hybridization capture procedures to monitor potential contamination.

Data Analysis

Sequencing reads were processed as follows: 3' adapter sequences were removed from raw reads by using cutadapt v1.4.2 (Martin, 2011), and reads shorter than 30 bp were discarded. The trimmed reads were mapped to a complete mitochondrial genome of *E. asinus* (GenBank No. X97337) using the "aln" algorithm in Burrows-Wheeler aligner (BWA) (Li and Durbin, 2010) with default parameters, and converted to bam format using the "samse" algorithm in BWA. Next, reads with a MapQuality score less than 30 and PCR duplicates were removed by using "view" and "rmdup" in samtools v0.1.9 (Li et al., 2009). Finally, a mitochondrial consensus sequence was generated in Geneious (https://www.geneious.com/), called with a minimum coverage of 2 and a base agreement greater than 75%.

Bioinformatics Analysis

To reconstruct the phylogenetic relationships and investigate the phylogenetic status of the analyzed samples among the *E. asinus*, the three newly obtained near-complete mitochondrial genomes were aligned with 95 *Equus* mitochondrial genomes from GenBank, including *E. asinus*, *E. kiang*, *E. hemionus*, *E. hydruntinus*, *E. ovodovi*, *E. zebra*, *E. grevyi*, *E. burchellii*, and *E. caballus* using MAFFT v7.471 (Katoh et al., 2002) on the CIPRES portal (Miller et al., 2010). The ambiguous section of the D-loop was discarded and the length of the final alignment was 16,621 bp. The substitution model GTR + G for each section was selected and the data set was divided into seven partitions (**Supplementary File S1**) using PartitionFinder v2.1.1 (Lanfear et al., 2016). We conducted a maximum-likelihood analysis using 1,000 bootstrap replicates in RAxML-HPC v8.2.12 (Stamatakis, 2014) with *E. caballus* as outgroup.

In addition, in order to estimate the divergence time of the two main donkey maternal lineages, we also performed a Bayesian analysis in BEAST v1.8.2 (Drummond et al., 2012) using the same partitioning as above. The phylogenetic tree was calibrated by root-tip dating, using the median calibrated radiocarbon ages or stratigraphic ages for all sequences (**Supplementary Table S2**), and assuming a most recent common ancestor (TMRCA) of all equids of 4.0–4.5 million years ago (Ma) (Orlando et al., 2013) as calibration points. We selected constant coalescent to provide the prior distribution for the branch lengths. The GTR + G substitution model was used, running 100,000,000 generations and sampling every 10,000 steps. The first 50,000,000 samples for each chain were discarded as burn-in. The result was analyzed with Tracer v1.7 (Rambaut et al., 2018) to check effective sample size for each model parameter. A Maximum Clade Credibility tree was calculated using TreeAnnotator v1.5.4 (Drummond et al., 2012) and viewed in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree). Moreover, the donkey female effective population size changes through time (**Supplementary Table S2**) were estimated using the Bayesian skyline plot (BSP) analysis in Tracer v1.7 (Rambaut et al., 2018).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI using accession numbers MZ823384, MZ823385, and MZ823386. Our data can be freely downloaded from NCBI after October 11th 2021.

ETHICS STATEMENT

The animal study was reviewed and approved by Shaanxi Provincial Institute of Archaeology, Xi'an 710054, China; Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Beijing 100044, China.

AUTHOR CONTRIBUTIONS

JY, XL, and MH conceived the project. SH and TD collected the samples and performed morphological analyses. GSH and AB guided the experimental and bioinformatics analyses. JY, MP, and UT performed laboratory work. LW, JH, BX, GSU, SS, and XH supported data analyses. LW, GSH, MH, and JY wrote the paper. All authors read and gave comments to the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 41472014; 42172027) and the "PPP" project jointly funded by China Scholarship Council and DAAD (No. 2016–6041).

ACKNOWLEDGMENTS

We are grateful to Dr. Stefanie Hartmann at the University of Potsdam, Dr. Johanna Paijmans at the University of Leicester, and Dr. Michael V. Westbury at the University of Copenhagen for their help in the process of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.759831/full#supplementary-material

REFERENCES

Al-Araimi, N. A., Gaafar, O. M., Costa, V., Neira, A. L., Al-Atiyat, R. M., and Beja-Pereira, A. (2017). Genetic Origin of Goat Populations in Oman Revealed by Mitochondrial DNA Analysis. *PLoS ONE* 12, e0190235. doi:10.1371/journal.pone.0190235

Bai, D., Zhao, Y., Bei, L., Gerelchimeg, B., Zhang, X., and Dugar-Javiin, M. (2017). Progress in the Whole Genome of *Equus* by Using High-Throughput Sequencing Technologies. *Hereditas* 39, 974–983. doi:10.16288/j.yczz.17-122

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, 1781. doi:10.1126/science.1096008

Blench, R. M. (2000). *A History of Donkeys, Wild Asses and Mules in Africa*. London (UK): University College London Press.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prufer, K., et al. (2007). Patterns of Damage in Genomic DNA Sequences from a Neandertal. *Proc. Natl. Acad. Sci.* 104, 14616–14621. doi:10.1073/pnas.0704665104

Chen, J., Sun, Y., Manglai, D., Min, L., and Pan, Q. (2010). Maternal Genetic Diversity and Population Structure of Four Chinese Donkey Breeds. *Livestock Sci.* 131, 272–280. doi:10.1016/j.livsci.2010.04.012

Chen, S., Li, J., Zhang, F., Xiao, B., Hu, J., Cui, Y., et al. (2019). Different Maternal Lineages Revealed by Ancient Mitochondrial Genome of *Camelus bactrianus* from China. *Mitochondrial DNA A* 30, 786–793. doi:10.1080/24701394.2019.1659250

Clutton-Brock, J. (1992). *Horse Power: A History of the Horse and the Donkey in Human Societies*. London (UK): Harvard University Press and the Natural History Museum.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., et al. (2013). Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave bear Reconstructed from Ultrashort DNA Fragments. *Proc. Natl. Acad. Sci.* 110, 15758–15763. doi:10.1073/pnas.1314445110

Daly, K. G., Maisano Delser, P., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., et al. (2018). Ancient Goat Genomes Reveal Mosaic Domestication in the Fertile Crescent. *Science* 361, 85–88. doi:10.1126/science.aas9411

Denbow, J., Shaw, T., Sinclair, P., Andah, B., and Okpoko, A. (1993). The Archaeology of Africa: Food, Metals and Towns. *J. Field Archaeol.* 23, 256. doi:10.2307/530508

Denham, T., and Iriarte, J. (2007). *Rethinking Agriculture: Archaeological and Ethnoarchaeological Perspectives*. Walnut Creek, CA: Left Coast Press

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi:10.1093/molbev/mss075

Epstein, H. (1971). *The Origin of the Domestic Animals of Africa*. New York (USA): African Publishing Corporation.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., et al. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell* 177, 1419–1435. doi:10.1016/j.cell.2019.03.049

Gansauge, M., and Meyer, M. (2013). Single-stranded DNA Library Preparation for the Sequencing of Ancient or Damaged DNA. *Nat. Protoc.* 8, 737–748. doi:10.1038/nprot.2013.038

González-Fortes, G., and Paijmans, J. L. A. (2019). "Whole-Genome Capture of Ancient DNA Using Homemade Baits," in *Ancient DNA: Methods and Protocols* (New York (USA): Humana Press), 93–105. doi:10.1007/978-1-4939-9176-1_11

Han, L., Zhu, S., Ning, C., Cai, D., Wang, K., Chen, Q., et al. (2014). Ancient DNA Provides New Insight into the Maternal Lineages and Domestication of Chinese Donkeys. *BMC Evol. Biol.* 14, 246. doi:10.1186/s12862-014-0246-4

Ivankovic, A., Kavar, T., Caput, P., Mioc, B., Pavic, V., and Dovc, P. (2002). Genetic Diversity of Three Donkey Populations in the Croatian Coastal Region. *Anim. Genet.* 33, 169–177. doi:10.1046/j.1365-2052.2002.00879.x

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* 30, 3059–3066. doi:10.1093/nar/gkf436

Kimura, B., Marshall, F. B., Chen, S., Rosenbom, S., Moehlman, P. D., Tuross, N., et al. (2010). Ancient DNA from Nubian and Somali Wild Ass Provides Insights into Donkey Ancestry and Domestication. *Proc. R. Soc. B.* 278, 50–57. doi:10.1098/rspb.2010.0708

Kimura, B., Marshall, F., Beja-Pereira, A., and Mulligan, C. (2013). Donkey Domestication. *Afr. Archaeol Rev.* 30, 83–95. doi:10.1007/s10437-012-9126-8

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* 34, msw260–773. doi:10.1093/molbev/msw260

Li, H., and Durbin, R. (2010). Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698

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

Lindgren, G., Backström, N., Swinburne, J., Hellborg, L., Einarsson, A., Sandberg, K., et al. (2004). Limited Number of Patrilines in Horse Domestication. *Nat. Genet.* 36, 335–336. doi:10.1038/ng1326

Lippold, S., Knapp, M., Kuznetsova, T., Leonard, J. A., Benecke, N., Ludwig, A., et al. (2011). Discovery of Lost Diversity of Paternal Horse Lineages Using Ancient DNA. *Nat. Commun.* 2, 450. doi:10.1038/ncomms1447

Liu, J., Yang, B., Xia, L., Sun, X., Guo, J., and Cheng, S. (2010). Study on Some Fragment Analysis of mtDNA D-loop and Phylogenesis for Nine Domestic Donkey Breeds in China. *Chinese J. Anim. Sci. (Chinese Version)* 46, 1–6.

Lu, C., Xie, W., Su, R., Ge, Q., Chen, H., Shen, S., et al. (2008). African Origin of Chinese Domestic Donkeys. *Hereditas (Beijing)* 30, 324–328. doi:10.3724/sp.j.1005.2008.00324

Ma, X. Y., Ning, T., Adeola, A. C., Li, J., Esmailizadeh, A., Lichoti, J. K., et al. (2020). Potential Dual Expansion of Domesticated Donkeys Revealed by Worldwide Analysis on Mitochondrial Sequences. *Zool. Res.* 41, 51–60. doi:10.24272/j.issn.2095-8137.2020.007

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17, 10–12. doi:10.14806/ej.17.1.200

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees," in 2010 Gateway Computing Environments Workshop (GCE), December 14, 2010. doi:10.1109/gce.2010.5676129

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., et al. (2013). Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse. *Nature* 499, 74–78. doi:10.1038/nature12323

Paijmans, J. L. A., Baleka, S., Henneberger, K., Taron, U. H., Trinks, A., Westbury, M. V., et al. (2017). *Sequencing Single-Stranded Libraries on the Illumina NextSeq 500 Platform*. arXiv. arxiv: 1711.11004.

Peters, J., and Driesch, A. v. d. (1997). The Two-Humped Camel (*Camelus bactrianus*): New Light on its Distribution, Management and Medical Treatment in the Past. *J. Zool.* 242, 651–679. doi:10.1111/j.1469-7998.1997.tb05819.x

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi:10.1093/sysbio/syy032

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., et al. (2013). IntCal13 and marine13 Radiocarbon Age Calibration Curves 0-50,000 Years Cal BP. *Radiocarbon* 55, 1869–1887. doi:10.2458/azu_js_rc.55.16947

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.* 105, 3715–3720. doi:10.1073/pnas.0709692105

Smith, D. G., and Pearson, R. A. (2005). A Review of the Factors Affecting the Survival of Donkeys in Semi-arid Regions of Sub-saharan Africa. *Trop. Anim. Health Prod.* 37, 1–19. doi:10.1007/s11250-005-9002-5

Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Sun, W., Yang, B., and Liang, C. (2007). Phylogenetic Relationship and Genetic Diversity of Chinese Four Domestic Donkeys Using mtDNA D-Loop. *China Herbivores (Chinese Version)* 27, 7–10.

Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Maisano Delser, P., et al. (2019). Ancient Cattle Genomics, Origins, and Rapid Turnover in the Fertile Crescent. *Science* 365, 173–176. doi:10.1126/science.aav1002

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., et al. (2013). Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLoS ONE* 8, e55950. doi:10.1371/journal.pone.0055950

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, 6014. doi:10.1038/s41467-020-19813-7

Xie, C. (1987). *Horse and Ass Breeds in China*. Shanghai (China): Shanghai Scientific and Technical Publishing House.

Xu, X., Gullberg, A., and Arnason, U. (1996). The Complete Mitochondrial DNA (mtDNA) of the Donkey and mtDNA Comparisons among Four Closely Related Mammalian Species-pairs. *J. Mol. Evol.* 43, 438–446. doi:10.1007/BF02337515

Yang, Z., and Hong, Z. (1989). Geographic and Population Ecology of Donkey in China. *J. Ecol. (Chinese Version)* 8, 40–42.

Yuan, J., Hou, X., Barlow, A., Preick, M., Taron, U. H., Alberti, F., et al. (2019). Molecular Identification of Late and Terminal Pleistocene *Equus Ovodovi* from Northeastern China. *PLoS ONE* 14, e0216883. doi:10.1371/journal.pone.0216883

Yuan, J., Sheng, G., Preick, M., Sun, B., Hou, X., Chen, S., et al. (2020). Mitochondrial Genomes of Late Pleistocene Caballine Horses from China Belong to a Separate Clade. *Quat. Sci. Rev.* 250, 106691. doi:10.1016/j.quascirev.2020.106691

Zheng, P. (1980). *Livestock Breeds in China and Their Ecological Characteristics*. 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 in Genetics

ORIGINAL RESEARCH
published: 16 November 2021
doi: 10.3389/fgene.2021.761252

Trait-specific Selection Signature Detection Reveals Novel Loci of Meat Quality in Large White Pigs

*Yu Shen†, Haiyan Wang†, Jiahao Xie, Zixuan Wang and Yunlong Ma**

Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction of the Ministry of Education and Key Laboratory of Swine Genetics and Breeding of the Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China

In past decades, meat quality traits have been shaped by human-driven selection in the process of genetic improvement programs. Exploring the potential genetic basis of artificial selection and mapping functional candidate genes for economic traits are of great significance in genetic improvement of pigs. In this study, we focus on investigating the genetic basis of five meat quality traits, including intramuscular fat content (IMF), drip loss, water binding capacity, pH at 45 min (pH45min), and ultimate pH (pH24h). Through making phenotypic gradient differential population pairs, Wright's fixation index (F_{ST}) and the cross-population extended haplotype homozogysity (XPEHH) were applied to detect selection signatures for these five traits. Finally, a total of 427 and 307 trait-specific selection signatures were revealed by F_{ST} and XPEHH, respectively. Further bioinformatics analysis indicates that some genes, such as *USF1*, *NDUFS2*, *PIGM*, *IGSF8*, *CASQ1*, and *ACBD6*, overlapping with the trait-specific selection signatures are responsible for the phenotypes including fat metabolism and muscle development. Among them, a series of promising trait-specific selection signatures that were detected in the high IMF subpopulation are located in the region of 93544042-95179724bp on SSC4, and the genes harboring in this region are all related to lipids and muscle development. Overall, these candidate genes of meat quality traits identified in this analysis may provide some fundamental information for further exploring the genetic basis of this complex trait.

Keywords: selective sweeps, phenotypic gradient differential population pairs, population differentiation-based methods, functional annotation, meat quality

OPEN ACCESS

Edited by:
Martijn Derks,
Wageningen University and Research, Netherlands

Reviewed by:
Tatsuhiko Goto,
Obihiro University of Agriculture and Veterinary Medicine, Japan
Muhammed Walugembe,
Iowa State University, United States

****Correspondence:***
Yunlong Ma
Yunlong.Ma@mail.hzau.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: *19 August 2021*
Accepted: *18 October 2021*
Published: *16 November 2021*

Citation:
Shen Y, Wang H, Xie J, Wang Z and Ma Y (2021) Trait-specific Selection Signature Detection Reveals Novel Loci of Meat Quality in Large White Pigs. Front. Genet. 12:761252. doi: 10.3389/fgene.2021.761252

INTRODUCTION

The most important purpose of pig breeding is the genetic improvement of important economic traits (Price 1999; Zhang et al., 2020). In past decades, the aim of genetic improvement of pig breeding has mainly focused on improving meat production through growth rate and feed efficiency, lean percentage, and decreasing backfat thickness. As expected, the genetic gain of these traits is successful in most selection programs. Simultaneously, human-driven selection has also indirectly shaped the meat quality traits, such as intramuscular fat content (IMF), pH values, drip loss (DL), and meat color (Herault et al., 2018). From the perspective of population genetics, the effect of human-driven selection as well as natural selection would leave detectable signatures in the genome. Therefore, detecting the selection signatures of these important economic traits can provide an insight into molecular mechanisms by which genomic fragments shape phenotypic diversity (Qanbari and Simianer 2014).

Although genomic selection has been widely applied in animal breeding in recent years, it is still difficult to carry out genetic improvement of meat quality traits in pigs (Zhan et al., 2021). The most critical factor is the high cost of measuring meat quality traits, which makes it difficult to build a large enough reference population. Therefore, marker-assisted selection (MAS) based on functional candidate genes is still an important choice for genetic improvement of meat quality traits. So far, genome-wide association analysis (GWAS) has been conducted to reveal some functional candidate genes, such as *PRKAG3*, *MC4R*, and *PIT1*, which have been identified in different populations to be related to meat quality traits (Yu et al., 1999; Milan et al., 2000; Lopez et al., 2015; Zhang et al., 2019). In addition, there are a lot of quantitative trait loci (QTL) associated with meat quality in pig QTLdb, more specifically, 1092 QTLs are associated with drip loss, 851 QTLs are associated with intramuscular fat content, and 667 QTLs are associated with meat color (http://www.animalgenome.org/QTLdb/SS/index) (Hu et al., 2019). However, it is still a challenge to reveal the exact genetic mechanism of meat quality traits.

To further explore the genetic mechanism of meat quality traits, we put forward a hypothesis: although the allele frequency has increased underlying human-driven selection, the loci related to meat quality traits are still polymorphic (Ma et al., 2019). Here, we construct three phenotypic gradient differential population pairs based on phenotype and use population differentiation–based methods to detect selection signatures associated with meat quality traits. If successful, the results can provide more information for understanding the genetic mechanism of meat quality traits and facilitate the genetic improvement of meat quality traits through marker-assisted selection.

MATERIALS AND METHODS

Animals and Phenotypes

In this study, a total of 233 castrated large white pigs were used. The experimental pigs were raised in the same farm, had a common diet, and drank water freely. The same standard management conditions were applied in the whole process of the experiment. Antibiotics are banned in the 3 months before slaughter. Finally, healthy individuals were chosen and slaughtered at around 90 Kg weight. The meat samples of longissimus lumborum from all pigs were collected for measuring meat quality traits, including IMF, DL, water binding capacity (WBC), pH45min, and pH24h (Horwitz and Latimer, 1995; Cherel et al., 2011; Zhan et al., 2021; Zhang et al., 2019).

Here, IMF was measured as percentage of lipid (lipid weight per 100 g of muscle tissue). Correspondingly, the Soxhlet extraction method was used following the standard AOAC Official method in foods (Horwitz and Latimer, 1995). The pH values of each sample were measured by a waterproof meat pH meter (Hanna, Romania). The electrode of the pH meter was calibrated in buffers at pH 7.00 and 4.00 before pH measurement. To calculate DL, we first measured the weight of the meat sample with 2.5 cm thickness at the 12th rib. Second, the final weight was measured, and after that, the meat sample was suspended in a sealed tube for 48 h at 4°C. Finally, the formula of (original weight—final weight)/original weight × 100 was applied to predict the DL. WBC was evaluated according to the Graua–Hamma method (Hamm, 1986). In this study, all five traits were measured in triplicate to reduce random error, and mean values were applied to the following analysis.

Genotyping and Quality Control

In this study, genomic DNA was extracted from ear tissue using a standard phenol-chloroform method. All 233 castrated large white pigs were genotyped using Illumina PorcineSNP60 BeadChips, which includes 62,163 single nucleotide polymorphisms (SNPs). Then, quality control was performed using the following criteria: i) SNP missing rate <0.05, ii) individual call rate >0.90, iii) SNPs in Hardy–Weinberg equilibrium (p > 10e–6), iv) SNP minor allele frequency ˃0.05, v) autosomal SNPs with known positions extracted. After quality control, 11,624 SNPs with minimum allele frequency less than 0.05 were deleted, and 42 markers were deleted after the Hardy–Weinberg test ($p \leq$ 10e-6). Finally, the data set contained 37,061 autosome SNPs with an average inter-marker spacing of 62.06 kb. The genotype data can be downloaded from Figshare (https://figshare.com/s/cd815d8930c75561392c). The BEAGLE software was then applied to impute the missing genotypes and infer haplotypes (Browning and Browning 2016). The PLINK (Version 1.90) software was used to measure the linkage disequilibrium and allele frequency in large white pigs (Purcell et al., 2007). Principal component analysis (PCA) was further performed using the PLINK (Version 1.90) software. To visualize the LD decay, the r2 values for 1 kb distance bins were averaged and drawn using the R program.

Detection of Trait-Specific Selection Signatures

To reveal trait-specific selection signatures, we construct three phenotypic gradient differential population pairs based on phenotype first and then identify selection signatures using population differentiation–based methods in this study. The detailed analysis flow applied the following steps: i) ranking by phenotypic values of each trait; ii) based on the rank of phenotypic values, we equally divided the source population into high and low phenotypic subpopulations and recorded them as first population pair; iii) based on step ii, we chose 75 individuals with a higher phenotype from the high phenotype subpopulation and 75 individuals with a lower phenotype from the low phenotype subpopulation to create the second population pair; iv) similarly, we further chose 45 individuals with a higher phenotype and 45 individuals with a lower phenotype from the second population pair subpopulations to create the third population pair; v) population differentiation–based methods XPEHH and F_{ST} were separately applied to identify selection signatures in three phenotypic gradient differential population pairs (Sabeti et al., 2007; Weir and Cockerham 1984).

In this analysis, the XPEHH scores do not need to be standardized. The empirical p-values were generated by

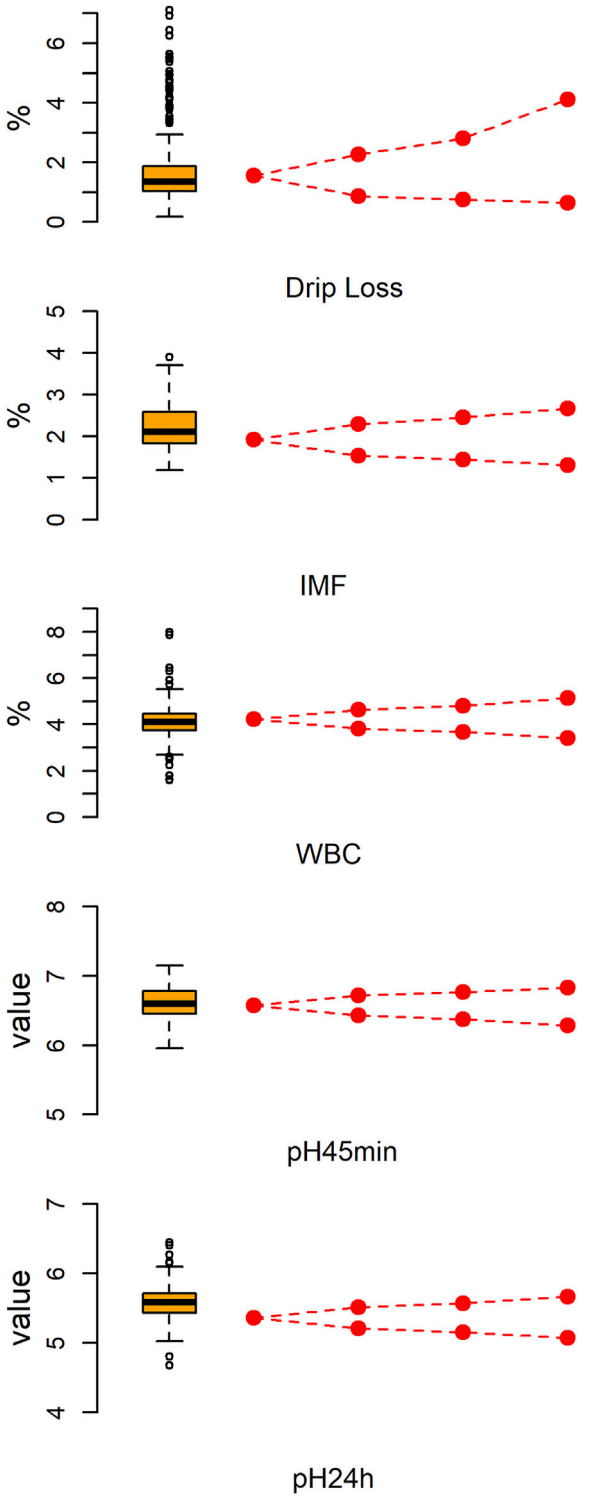

FIGURE 1 | The phenotypic distribution of five meat quality traits and the tendency of phenotypic values in phenotypic gradient differential population pairs. The left is a box-whiskers plot. The right line graph displays a gradual change tendency of phenotypic mean value from the source population to first through third population pairs.

genome-wide ranking of F_{ST} and XPEHH scores, respectively. The trait-specific selection signatures were defined using the following two criteria: i) the SNPs with p-value < .01 were considered as significant loci; ii) from first to second to third population pair, the F_{ST} or XPEHH scores of each significant locus displayed a gradient change.

Functional Annotation for Trait-Specific Selection Signatures

To reveal the potential biological functions of trait-specific selection signatures, we defined the trait-specific selection region as the genomic region within a distance of 200 kb around the trait-specific selection signatures. The BioMart program in ensembl (https://www.ensembl.org/index.html) was employed to search the genes and their orthologous genes of mouse harboring the trait-specific selection signature regions. Then, the database of Mouse Genome Informatics (MGI) was used to perform functional annotation (Dickinson et al., 2016). The trait-specific selection signature regions were also annotated using pigQTLdb in this analysis. Based on the genes harboring in trait-specific selection signature regions, GO and pathway analysis was used for the functional annotation and classification using DAVID 6.8 (https://david.ncifcrf.gov/) (Huang et al., 2009). The GO terms and pathways with p-value < .05 were considered as significant after Bonferroni correction.

RESULTS

Phenotypes Among Phenotypic Gradient Differential Population Pairs

Figure 1 shows the descriptive statistics of five meat quality traits of large white pigs. As shown in **Supplementary Table S1**, there was a significantly positive association between DL and the other traits. However, IMF only has a significant positive correlation with WBC. For each trait, all 233 pigs were divided into two subpopulations: high and low phenotype value groups. Then, the extreme individuals were chosen to construct three phenotypic gradient differential population pairs based on the ranking of phenotype values. As expected, the average phenotype value of the high phenotypic subpopulation increased sequentially from the first to the third population pair. Correspondingly, the average phenotype value of low phenotypic subpopulation decreased sequentially. As expected, the average phenotypic value of population pairs for all five meat quality traits are significant differences (p < .01).

Genomic Characters Among Phenotypic Gradient Differential Population Pairs

To explore the influence of population division in this analysis, minor allele frequency, and linkage disequilibrium were investigated among different populations. As expected, the distribution of the minor allele frequency (MAF) in each

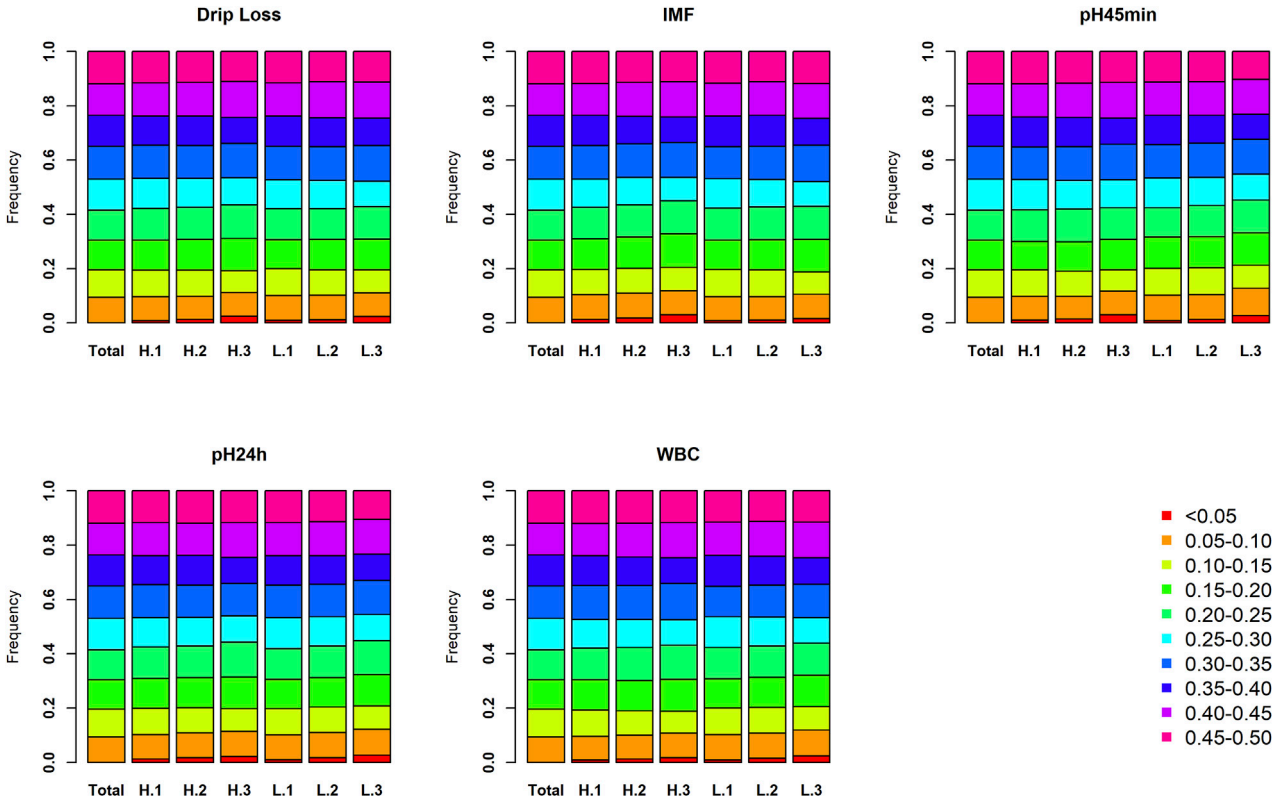

FIGURE 2 | The proportion of each MAF bin in each subpopulation for five meat quality traits. The total represents the proportion of each MAF bin in source population. Similarly, H.3 (H.2, H.1) and L.3 (L.2, L.1) represent the proportion of each MAF bin in the two subpopulations of third (second and first) population pair.

subpopulation is similar to the results using all 233 large white pigs (**Figure 2**). The proportion of MAF between 0 and 0.1 increases slightly as the sample size of each subpopulation decreases. Simultaneously, the trend of linkage disequilibrium decay is similar in all subpopulations (**Supplementary Figure S1**). The results show that the genomic characters of each subpopulation in the three phenotypic gradient population pairs have little difference. Therefore, the population division would not affect the identification of selection signatures. Further PCA analysis indicated that the population division for each trait will not cause population stratification (**Supplementary Figures S2–S6**). Overall, the results can prove that the trait-specific selection signatures revealed in this study are caused by phenotypic differences rather than an accidental genomic structural difference because of population division.

Trait-Specific Selection Signatures

Based on the constructed population pairs with extreme differences in phenotypes, XPEHH and F_{ST} were employed to identify positive selection signatures. In this analysis, about 370 positive selection signatures were detected by each method in each population pair for each trait (**Supplementary Table S2**). Then, these positive selection signatures with gradient changes in three phenotypic gradient differential population pairs were defined as trait-specific selection signatures. Finally, 55, 49, 43, 111, and 49 trait-specific selection signatures were detected in DL, IMF, pH45min, pH24h, and WBC using F_{ST} test, respectively. Similarly, the XPEHH test revealed 59, 102, 159, 43, and 64 trait-specific selection signatures in DL, IMF, pH45min, pH24h, and WBC, respectively (**Figure 3**, **Supplementary Figures S7–S10**).

Candidate Genes Overlapping With Trait-Specific Selection Signatures

Based on the trait-specific selection signatures, the genes overlapping with the potential selection regions were determined using the pig reference genome (Sscrofa11.1). Enrichment analysis showed that no significant biological terms may be associated with five meat quality traits after multiple correction in this study (**Supplementary Table S3**). Nevertheless, we note that genes harboring trait-specific selection signatures are related with muscle morphology by functional annotations based on the MGI database (**Table 1**). For example, the *SCYL3* gene that overlaps with the DL-specific selection signatures is related to abnormal morphology of mouse skeletal muscle fibers through orthologous alignment and MGI annotation (Dickinson et al., 2016). *ENSSSCG00000027613* genes were found overlapping with the IMF-specific selection signature. Note that the ortholog gene of *ENSSSCG00000027613* is *Trdn* in the mouse, which is related to abnormal skeletal muscle fiber triad morphology (Oddoux et al., 2009). The previous study indicates that this gene plays an important role in skeletal muscle function

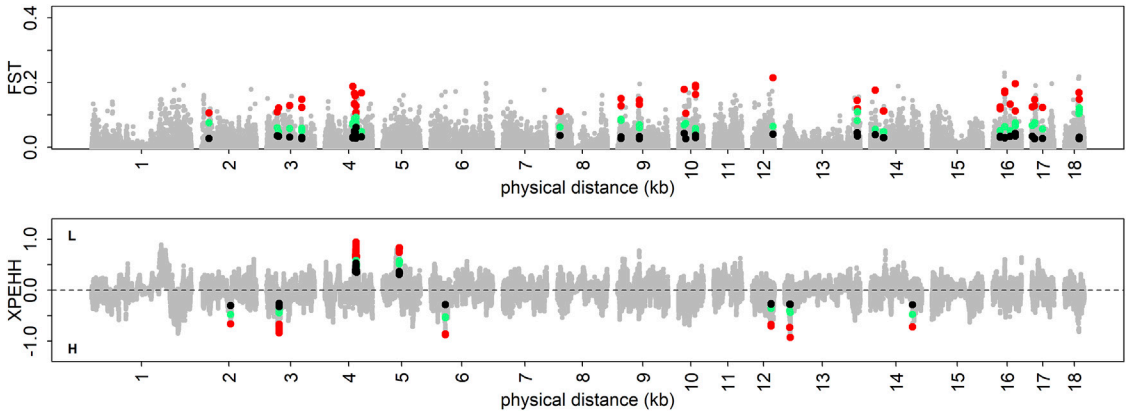

FIGURE 3 | Visualization of trait-specific selection signatures for DL. The colored dots represent trait-specific selection signatures.

TABLE 1 | The summary of functional annotation of the trait-specific selection signatures.

Chr	Position (bp)[a]	XPEHH (F_{ST}) Scores[b]	Genes[c]	Trait	MGI phenotype
10	20392326–20707474	(0.03 < 0.07 < 0.11)	*DENND1B*	-,DL	MP:0001257_increased body length
12	51773420–51794447	−0.26 > −0.35 > −0.67	*SCIMP*	High, DL	MP:0001260_increased body weight
4	81133591–81176676	(0.05 < 0.07 < 0.13)	*SCYL3*	-,DL	MP:0003084_abnormal skeletal muscle fiber morphology
5	46244283–46274049	0.31 < 0.52 < 0.75	*ENSSSCG000000005*(*Smco2*)	Low, DL	MP:0003960_increased lean body mass
1	39180526–39339224	(0.06 < 0.09 < 0.12)	*ENSSSCG0000002761*(*TRDN*)	-,IMF	MP:0009411_abnormal skeletal muscle fiber triad morphology
10	25822287–25826726	0.30 < 0.41 < 0.55	*ENSSSCG00000010936*(*Fabp3*)	Low, IMF	MP:0002118_abnormal lipid homeostasis
10	25707711–25735067	0.30 < 0.42 < 0.55	*ZNF367*(*Zfp367*)	Low, IMF	MP:0005553_increased circulating creatinine level
4	15895092–15930934	(0.04 > 0.09 > 0.13)	*FBXO32*	-, IMF	MP:0004064_decreased susceptibility to induced muscular atrophy
4	90275213–90294819	−0.38>−0.46>−0.66	*CASQ1*	High, IMF	MP:0002106_abnormal muscle physiology
7	25751185–25808637	0.31 < 0.44 < 0.64	*GFRAL*	Low, IMF	MP:0001259_abnormal body weight
11	23237692–23894699	(0.05 < 0.08 < 0.14)	*ENOX1*	-,pH45m	MP:0000062_increased bone mineral density
11	68920906–69301332	0.53 < 0.62 < 0.96	*PCCA*	Low, pH45m	MP:0001429_dehydration
7	114035383–114541592	−0.41 < −0.48 < −0.55 (0.06 < 0.11 < 0.18)	*RIN3*	High, pH45m	MP:0005560_decreased circulating glucose level
3	24815010–24835836	0.33 < 0.36 < 0.55	*CRYM*	Low, pH24h	MP:0005472_abnormal triiodothyronine level
4	123674697–123779615	(0.05 < 0.08 < 0.12)	*FNBP1L*	-,pH24h	MP:0003961_decreased lean body mass
1	268581835–268590836	−0.35 > −0.46 > −0.55 (0.04 < 0.07 < 0.11)	*PTGES2*	High, WBC	MP:0001785_edema
8	132035777–132200420	(0.03 < 0.06 < 0.08)	*PTPN13*	-,WBC	MP:0001261_obese

[a]*This column is the position of candidate genes.*
[b]*This column is the XPEHH(F_{ST}) scores of three phenotypic gradient change population pairs.*
[c]*The gene in brackets is the mouse ortholog gene.*

and structure. In addition, the ortholog gene *FBX O 32* and *CASQ1* are separately related to reducing the susceptibility to induced muscle atrophy and abnormal muscle physiology (Mosca et al., 2016; Singh et al., 2017). Both of them were identified as a potential selection signatures in the IMF content trait.

A Highlighted IMF-specific Selection Signature Region in Chromosome 4

MF is one of the most important meat quality traits in pig breeding programs. Here, a promising genomic region (SSC4 93544042–95179724bp) that was detected by XPEHH and F_{ST} simultaneously is associated with high IMF content. Based on the pigQTLdb, we found that 10 QTLs related to meat quality and fat metabolism are harbored in this genomic region (**Supplementary Table S4**). Further bioinformatics annotation in this region found that *USF1*, *NDUFS2*, *PIGM*, *IGSF8*, and *CASQ1* gene can be considered as functional candidate genes for meat quality. Among them, the *USF1* gene plays an important role in lipid homeostasis; *PIGM* is related to lipid metabolism and catabolism; *IGSF8* and *CASQ1* are related to muscle development; *NDUFS2* is related to the synthesis of energy metabolism (**Table 2**) (Kim et al., 2015).

TABLE 2 | The annotation of seven potential candidate genes in SSC4 (Kim et al., 2015).

Chr	Position (bp)[a]	XPEHH (F_{ST}) Scores[b]	*Trait*	Genes	Function
4	89395156–89401258	−0.42 > −0.50 > −0.64 (0.04 < 0.06 < 0.15)	High, IMF	*USF1*	Lipid homeostasis
4	89247150–89262159	−0.42 > −0.50 > −0.64 (0.04 < 0.06 < 0.15)	High, IMF	*NDUFS2*	Integration of energy metabolism
4	90461734–90467492	−0.34 > −0.43 > −0.61	High, IMF	*PIGM*	Lipid metabolism and catabolism `
4	90370235–90405480	−0.34 > −0.44 > −0.64	High, IMF	*IGSF8*	Muscle development
4	90275213–90294819	−0.38 > −0.46 > −0.66	High, IMF	*CASQ1*	Muscle development

[a]*This column is the position of candidate genes.*
[b]*This column is the XPEHH(F_{ST}) scores of three phenotypic gradient change population pairs.*

DISCUSSION

In this study, we planned to map some candidate genes of meat quality using population differentiation–based selection signature methods through constructing phenotypic gradient difference population pairs. In theory, meat quality traits are quantitative traits, controlled by minor-effect polygenes, and their genetic mechanisms are complex (Falconer and Mackay, 1960; Hill et al., 2007). Although these loci are related to the complex traits, they are not easy to detect through genome-wide association analysis because of their small effect, especially after multiple correlations. However, these loci have been shaped by human-driven selection; they can be detected through sweep analysis from the perspective of population genetics (Qanbari and Simianer 2014). Therefore, it is theoretically feasible to use population differentiation–based selection signature methods to identify trait-specific selection signatures through constructing phenotypic gradient population pairs. The results of the functional annotation also support this hypothesis. A series of genes and QTLs harboring trait-specific selection signatures are related with meat quality traits (**Table 1**, **Supplementary Table S4**). There is no doubt that this strategy also has a shortcoming that the fixed loci caused by selection will not be identified after MAF quality control. In addition, both the sample size of the third population pair and the gradient across different population pairs are at least 40 unrelated individuals according to our previous study (Ma et al., 2015). Although small sample sizes appear to be sufficient in the detection of selection signatures according to simulation research, there is no doubt a larger sample size will contribute to decrease the risk of detection bias.

The genetic mechanism of meat quality traits is complex. The candidate genes related to meat quality traits discovered in this study may have pleiotropism effects. As shown in **Table 1**, some orthologous genes, such as *SCIMP*, *ENSSSCG00000000550*, *GFRAL*, and *FNBP1L*, are related to the body weight and lean body weight of mice (George et al., 2014; Tsai et al., 2019). We found that some genes overlapping with meat quality traits–specific selection signatures are also associated with lipid traits, such as *ENSSSCG00000010936* and *PTPN13*. Simultaneously, the orthologous gene of *ZNF367* that was considered an IMF-specific selection signature is related to the increase in circulating levels of creatinine in mice, which is consistent with the fact that creatinine has flavor properties. After the annotation of pigQTLdb, we note that trait-specific selection signals are mainly associated with QTLs for meat quality traits, but there are also QTLs related to backfat thickness, body weight, and body length (**Supplementary Table S4**). In general, there is a negative correlation between backfat thickness and meat quality traits, especially DL and IMF. This result indicates that the artificial selection of backfat thickness and body size in recent pig breeding programs should have a significant impact on meat quality traits.

Because the XPEHH method can identify the selected population, this study can further study the complex genetic mechanism of meat quality traits (Sabeti et al., 2007). In general, the artificial selection of economic traits shows a single direction of phenotypic changes, but the corresponding genetic basis changes are in two directions: increasing and decreasing effects. Taking IMF as an example, 48 and 54 selection signatures were separately detected in low and high IMF subpopulations (**Supplementary Table S5**). Note that the biological phenotypes of the *CASQ1* and *ENSSSCG00000010936* genes identified in the high and low IMF subpopulations are abnormal muscle physiology and abnormal lipid homeostasis based on MGI annotation (Dickinson et al., 2016). In addition, 115 trait-specific selection signatures detected in the low pH45min subpopulation are more than 44 trait-specific selection signatures detected in the high pH45min subpopulation. This is consistent with the fact that the breeding direction of lean pigs under artificial selection leads to a decrease in water holding capacity and a rapid decrease in muscle pH after slaughter in recent years. However, 37 trait-specific selection signatures detected in the low DL subpopulation are more than 22 trait-specific selection signatures detected in the high DL subpopulation (**Supplementary Table S5**). This further indicates the complexity of the genetic mechanism of meat quality traits. Meat quality traits are affected by many factors, including fat metabolism, transportation, muscle fiber formation, and physiological conditions. Therefore, it can be inferred that further multi-omics integrated analysis is an important way to analyze the genetic mechanism of meat quality traits in the future.

CONCLUSION

In this study, we propose a new strategy to identify trait-specific selection signatures. The application in five meat quality traits in large white pigs indicate that this strategy is promising in gene

mapping. Furthermore, we detected a series of genomic selection signatures and identified some genes related to meat quality traits, such as *USF1*, *NDUFS2*, *PIGM*, *IGSF8*, and *CASQ1*, which provide a reference for future pig breeding.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by All research involving animals was conducted under protocols (No. 5 proclaim of the Standing Committee of Hubei People's Congress) approved by the Standing Committee of Hubei People's Congress and the ethics committee of Huazhong Agricultural University in China. All experiments were performed in accordance with approved relevant guidelines and regulations.

AUTHOR CONTRIBUTIONS

YM conceived the study, participated in the design, performed the statistical analyses, and drafted the manuscript. YS, HW, JX and ZW drafted the paper. YM revised the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (Nos. 32072696 and 31790414), China Agriculture Research System of MOF and MARA (No. CARS-35) and SRF (No. 2021033), HW was supported by China Scholarship Council (201906765023) and Hubei Chenguang Talented Youth Development Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.761252/full#supplementary-material

REFERENCES

Browning, B. L., and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. doi:10.1016/j.ajhg.2015.11.020

Cherel, P., Pires, J., Glénisson, J., Milan, D., Iannuccelli, N., Hérault, F., et al. (2011). Joint Analysis of Quantitative Trait Loci and Major-Effect Causative Mutations Affecting Meat Quality and Carcass Composition Traits in Pigs. *BMC Genet.* 12, 76. doi:10.1186/1471-2156-12-76

Dickinson, M. E., Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., White, J. K., et al. (2016). High-throughput Discovery of Novel Developmental Phenotypes. *Nature* 537, 508–514. doi:10.1038/nature19356

Falconer, D. S., and Mackay, T. F. (1960). *Introduction to Quantitative Genetics*. London: Longman.

George, B., Jain, N., Fen Chong, P., Hou Tan, J., and Thanabalu, T. (2014). Myogenesis Defect Due to Toca-1 Knockdown Can Be Suppressed by Expression of N-WASP. *Biochim. Biophys. Acta (Bba) - Mol. Cell Res.* 1843, 1930–1941. doi:10.1016/j.bbamcr.2014.05.008

Hérault, F., Damon, M., Cherel, P., and Le Roy, P. (2018). Combined GWAS and LDLA Approaches to Improve Genome-wide Quantitative Trait Loci Detection Affecting Carcass and Meat Quality Traits in Pig. *Meat Sci.* 135, 148–158. doi:10.1016/j.meatsci.2017.09.015

Hill, W. G., Mulder, H. A., and Zhang, X.-S. (2007). The Quantitative Genetics of Phenotypic Variation in Animals. *Acta Agriculturae Scand. Section A - Anim. Sci.* 57, 175–182. doi:10.1080/09064700801959353

Hirwitz, W., and Latimer, G. (1995). Official Methods of Analysis of Aoac International. *Trends Food Sci. Technol.* 6 (11), 382.

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

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, 44–57. doi:10.1038/nprot.2008.211

Kim, E.-S., Ros-Freixedes, R., Pena, R. N., Baas, T. J., Estany, J., and Rothschild, M. F. (2015). Identification of Signatures of Selection for Intramuscular Fat and Backfat Thickness in Two Duroc Populations1. *J. Anim. Sci.* 93, 3292–3302. doi:10.2527/jas.2015-8879

López, M. E., Neira, R., and Yáñez, J. M. (2015). Applications in the Search for Genomic Selection Signatures in Fish. *Front. Genet.* 5, 458. doi:10.3389/fgene.2014.00458

Ma, H., Zhang, S., Zhang, K., Zhan, H., Peng, X., Xie, S., et al. (2019). Identifying Selection Signatures for Backfat Thickness in Yorkshire Pigs Highlights New Regions Affecting Fat Metabolism. *Genes (Basel)* 10. doi:10.3390/genes10040254

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, 426–436. doi:10.1038/hdy.2015.42

Milan, D., Jeon, J.-T., Looft, C., Amarger, V., Robic, A., Thelander, M., et al. (2000). A Mutation in PRKAG3 Associated with Excess Glycogen Content in Pig Skeletal Muscle. *Science* 288, 1248–1251. doi:10.1126/science.288.5469.1248

Mosca, B., Eckhardt, J., Bergamelli, L., Treves, S., Bongianino, R., De Negri, M., et al. (2016). Role of the JP45-Calsequestrin Complex on Calcium Entry in Slow Twitch Skeletal Muscles. *J. Biol. Chem.* 291, 14555–14565. doi:10.1074/jbc.m115.709071

Oddoux, S., Brocard, J., Schweitzer, A., Szentesi, P., Giannesini, B., Brocard, J., et al. (2009). Triadin Deletion Induces Impaired Skeletal Muscle Function. *J. Biol. Chem.* 284, 34918–34929. doi:10.1074/jbc.m109.022442

Price, E. O. (1999). Behavioral Development in Animals Undergoing Domestication. *Appl. Anim. Behav. Sci.* 65, 245–271. doi:10.1016/s0168-1591(99)00087-8

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

Qanbari, S., and Simianer, H. (2014). Mapping Signatures of Positive Selection in the Genome of Livestock. *Livestock Sci.* 166, 133–143. doi:10.1016/j.livsci.2014.05.003

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

Singh, P., Li, D., Gui, Y., and Zheng, X.-L. (2017). Atrogin-1 Increases Smooth Muscle Contractility through Myocardin Degradation. *J. Cell. Physiol.* 232, 806–817. doi:10.1002/jcp.25485

Tsai, V. W.-W., Zhang, H. P., Manandhar, R., Schofield, P., Christ, D., Lee-Ng, K. K. M., et al. (2019). GDF15 Mediates Adiposity Resistance through Actions on

GFRAL Neurons in the Hindbrain AP/NTS. *Int. J. Obes.* 43, 2370–2380. doi:10.1038/s41366-019-0365-5

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

Yu, B. T. P., Wang, L., Tuggle, C. K., and Rothschild, M. F. (1999). Mapping Genes for Fatness and Growth on Pig Chromosome 13: a Search in the Region Close to the Pig PIT1 Gene. *J. Anim. Breed. Genet.* 116, 269–280. doi:10.1046/j.1439-0388.1999.00198.x

Zhan, H., Xiong, Y., Wang, Z., Dong, W., Zhou, Q., Xie, S., et al. (2022). Integrative Analysis of Transcriptomic and Metabolomic Profiles Reveal the Complex Molecular Regulatory Network of Meat Quality in Enshi Black Pigs. *Meat Sci.* 183, 108642. doi:10.1016/j.meatsci.2021.108642

Zhang, S., Zhang, K., Peng, X., Zhan, H., Lu, J., Xie, S., et al. (2020). Selective Sweep Analysis Reveals Extensive Parallel Selection Traits between Large white and Duroc Pigs. *Evol. Appl.* 13, 2807–2820. doi:10.1111/eva.13085

Zhang, Y., Zhang, J., Gong, H., Cui, L., Zhang, W., Ma, J., et al. (2019). Genetic Correlation of Fatty Acid Composition with Growth, Carcass, Fat Deposition and Meat Quality Traits Based on GWAS Data in Six Pig Populations. *Meat Sci.* 150, 47–55. doi:10.1016/j.meatsci.2018.12.008

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: 23 November 2021
doi: 10.3389/fgene.2021.759827

Ancient Mitogenomes Reveal the Domestication and Distribution of Cattle During the Longshan Culture Period in North China

Xing Zhang[1,2], Liu Yang[1,2], Lingyun Hou[1], Hua Li[2], Hai Xiang[2] and Xingbo Zhao[1]**

[1]National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing, China, [2]Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, School of Life Science and Engineering, Foshan University, Foshan, China

Cattle, as an important tool for agricultural production in ancient China, have a complex history of domestication and distribution in China. Although it is generally accepted that ancient Chinese taurine cattle originated from the Near East, the explanation regarding their spread through China and whether or not this spread was associated with native aurochs during ancient times are still unclear. In this study, we obtained three nearly complete mitochondrial genomes (mitogenomes) from bovine remains dating back ca. 4,000 years at the Taosi and Guchengzhai sites in North China. For the first time at the mitogenome level, phylogenetic analyses confirmed the approximately 4,000-year-old bovines from North China as taurine cattle. All ancient cattle from both sites belonged to the T3 haplogroup, suggesting their origin from the Near East. The high affinity between ancient samples and southern Chinese taurine cattle indicated that ancient Chinese cattle had a genetic contribution to the taurine cattle of South China. A rapid decrease in the female effective population size ca. 4.65 thousand years ago (kya) and a steep increase ca. 1.99 kya occurred in Chinese taurine cattle. Overall, these results provide increasing evidence of the origin of cattle in the middle Yellow River region of China.

Keywords: ***Bos taurus*, Ancient DNA, mitochondrial genome, domestication, longshan culture**

OPEN ACCESS

Edited by:
Ino Curik,
University of Zagreb, Croatia

Reviewed by:
Takahiro Yonezawa,
Tokyo University of Agriculture, Japan
Eileen Armstrong,
Universidad de la República, Uruguay

Correspondence:
Xingbo Zhao
zhxb@cau.edu.cn
Hai Xiang
xh@fosu.edu.cn

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

Received: *17 August 2021*
Accepted: *29 October 2021*
Published: *23 November 2021*

Citation:
Zhang X, Yang L, Hou L, Li H, Xiang H and Zhao X (2021) Ancient Mitogenomes Reveal the Domestication and Distribution of Cattle During the Longshan Culture Period in North China. Front. Genet. 12:759827. doi: 10.3389/fgene.2021.759827

INTRODUCTION

As one of the earliest domesticated animals (Diamond, 2002), cattle not only provide meat, dairy products, and leather but also played an important role in cultivation and other agricultural activities in the past 13,000 years (Ebersbach, 2002). According to their morphological characteristics and living habits, modern cattle are divided into humpless group (taurine, *Bos taurus*) and humped group (zebu, *Bos indicus*). Both archeological and genetic evidence have demonstrated that taurine cattle and zebu cattle were independently domesticated from aurochs in the Near East approximately 10,500 years before present (YBP) (Ajmone-Marsan et al., 2010) and in the Indus Valley approximately 8,500 YBP (Loftus et al., 1994), respectively, and then separately spread to the rest of the world following human migrations (Patel, 2009; Felius et al., 2014).

The abundant cattle resources in China, including the aurochs (*Bos primigenius*) existing in ancient China, as well as the zebu and taurine cattle in modern China, imply a complicated history of formation of Chinese taurine cattle (Qiu, 2006; Zhang et al., 2013; Cai et al., 2014; Brunson et al., 2016b). Analyses of

partial mitochondrial DNA (mtDNA) sequences and complete mitochondrial genomes (mitogenomes) of various modern cattle breeds have revealed multiple region-specific maternal lineages [e.g., the T4 haplogroup was found mainly in East Asia (Felius et al., 2014)]. Analyses of mtDNA D-loop sequences revealed that the northern Chinese breeds originated from taurine cattle, the southern breeds originated from zebu cattle, whereas the central Chinese cattle breeds were of hybrid origin (Zhang et al., 2015). Whole-genome sequencing has revealed three different ancestors for the East Asian cattle, including East Asian taurine ancestry, Eurasian taurine ancestry, and Chinese indicine ancestry (Chen et al., 2018). Further genome analyses suggested that at least two historical migration events occurred in Chinese cattle in northern China (Chen et al., 2018).

Although the analyses of these modern specimens based on partial and complete genomes provide reasonable assumption into the study of cattle history, a direct ancient DNA (aDNA) analysis focusing on the ancient time has been valued more highly as it can provide a wealth of direct and solid evidence on the origin and dispersal of cattle species. In recent decades, advances in high throughput sequencing makes breakthroughs in aDNA research, which still performed inadequately. Specifically, analyses of ancient agricultural animal samples including pig, cattle, sheep and dog etc. especially from China, where exists numerous archeological sites, provide significant value to represent the history of animals and their companions, human society.

The earliest known taurine cattle remains in China were found in the Houtaomuga site in Jilin Province 5,500–5,300 YBP (Cai et al., 2018) and supported a significant population expansion of taurine cattle that spread from Northeast Asia into China. Through the discontinuous ancient DNA analysis on ancient cattle remains in China till now, it is widely accepted that Chinese taurine cattle were originated from the Near East (Brunson et al., 2016a; Cai et al., 2018) with two routes: Northwest entrance route and Northeast entrance route. In addition, genomic study of ancient cattle samples from the Shimao site has confirmed that the domestic cattle present at the site are pure East Asian taurine cattle (Chen et al., 2018). Although there is growing evidence that the domestication of Chinese cattle was a non-native process, the increasing excavation of auroch remains has made the origin of Chinese cattle contentious. Particularly, the exist of haplogroup C auroch 10,000 years ago in Northeast China (Zhang et al., 2013) and the evidence of haplogroup (P1a) incorporated into domestic cattle of northeastern Asia (Mannen et al., 2020) make it necessary to determine the species and phylogenetic relationships among suspected bovine remains from different sites in China, as well as to define the role that aurochs play in the development of taurine cattle in China.

MATERIALS AND METHODS

Samples

Bovine bone remains were collected from the Taosi and Guchengzhai sites. Both sites are located at the lower and middle reaches of the Yellow River drainage basin (**Figure 1**) and date back to the Longshan Culture period (approximately 4,000 YBP) (**Supplementary Table S1**). Specifically, the Taosi site is a large urban site located between the Ta'er Mountain and the Fen River, approximately 7.5 km northeast of Xiangfen County, Shanxi Province. The Taosi site dates back 4,350–3,900 years (Brunson et al., 2016a). The paleoenvironmental conditions around the Taosi site were warmer and wetter than now (Wang et al., 2011). Many domestic animals remain, including cattle bones, were found at the Taosi site; these animals were kept primarily as animals of the wealthy inhabitants (Brunson et al., 2016a). The Guchengzhai site is located in Dafanzhuang Village, 35 km southeast of Xinmi City, Henan Province, 1.5 km northeast of the intersection of the Zhenshui River and the Yan River. The Guchengzhai site dates back 4,150–3,950 years. Similarly, the paleoenvironment of the Guchengzhai site was slightly warmer and more humid than now (Xu et al., 2013). Domestic animals, including dogs, pigs, cattle, and sheep, as well as wild animals such as bears, sika deer, and elks were extensively discovered at the Guchengzhai site (Guo, 2018). The detailed information on all samples were shown in **Supplementary Table S1**. Before aDNA analysis, all samples were morphologically determined to be cattle remains.

FIGURE 1 | Location of the Taosi and Guchengzhai archaeological sites (red dots).

Extraction of Ancient DNA

All aDNA experiments were conducted in the Ancient DNA Laboratory at China Agricultural University and repeated independently in the Animal Genetic and Evolutionary Laboratory at Foshan University, following the experimental guidelines to eliminate contamination with modern samples. First, the equipment involved in the experiment was soaked with 5% (v/v) sodium hypochlorite solution and washed with 75% (vol/vol) alcohol, followed by UV irradiation for 1 h. Thereafter, the samples were prepared by cautiously removing the adhering soils on the surface of samples with a drill, followed by washing with 5% (vol/vol) sodium hypochlorite solution and then with double-distilled water and subsequent UV irradiation. The bones were then ground to fine-

grained powder using an automatic sample quick grinding machine (Jingxin Industrial Development Co., Ltd., China), and 100–200 mg of powder was subjected to DNA extraction using a silica adsorption column system. At China Agricultural University, 50 μL of extracted DNA was obtained using the QIAamp DNA Investigator Kit (QIAGEN, Germany). At Foshan University, aDNA extraction was performed following the method described by Dabney (Dabney et al., 2013) with some adaptive modifications, including setting the lytic temperature to 45°C and extending the time of incubation with double distilled water (ddH_2O) to 10 min.

Mitogenome Capture and High-Throughput Sequencing

The mitogenomes of the genus *Bos*, including Holstein (*Bos taurus*), zebu (*Bos indicus*), gayal (*Bos frontalis*), and yak (*Bos grunniens*), were obtained using long-range PCR amplification. The PCR products were then used to prepare *Bos* hybrid baits. A total of 50 ng DNA per sample was used to construct a double-stranded DNA library using the KAPA Hyper Prep Kit for Illumina® platforms (KAPA Biosystems, San Francisco, USA) with minor modifications. The libraries were quantified using a Qubit 4.0 Fluorometer (Life Technologies Corporation, USA) and submitted for subsequent mitogenome enrichment and capture. Four libraries of Taosi samples were equally enriched to a total of 1 ng of DNA with unique adapter index sequences. Thereafter, mitogenome enrichment and hybrid capture were performed using the xGen Hybridization and Wash Kit (Integrated DNA Technologies, Inc., USA). The enriched library was then sequenced with 150 paired end reads (PE150) using the Illumina HiSeq X Ten platform. For the resequencing samples, up to 100 ng DNA per sample was used to prepare the library with the VAHTS Universal Plus DNA Library Prep Kit for MGI (Vazyme Biotech Co., Ltd, China). The qualified libraries were sequenced with PE100 using the MGI platform.

Data Alignment, Filtering, and Authenticity Estimation

The quality of the raw data was assessed using the FastQC v0.11.9 (Andrews, 2010). Adapter and low-quality reads were filtered using clip&merge v1.7.8 from EAGER pipeline (Peltzer et al., 2016) with the following parameters: *-minlength* 25, *--minquality* 30. Before mapping, the first 250 and 500 bp of the taurine cattle reference mitogenome (GenBank accession No.: V00654) were concatenated to the end using CircularGenerate from the EAGER pipeline (Peltzer et al., 2016). The filtered reads were mapped to the circularized reference V00654 using CircularMapper v1.0 (Peltzer et al., 2016), with parameters as *-n* 0.01, *-l* 16,500. The resulting file was converted to a bam file and sorted using Samtools v1.9 (Li, 2011). Unmapped reads and reads with mapping quality <25 were removed using Samtools v1.9. Duplicate reads were removed using MarkDuplicates of Picard-tools v2.22.9 (http://broadinstitute.github.io/picard/). The endogenous DNA, coverage rate, and mapping quality distribution were calculated using Qualimap v2.2.2-dev (Okonechnikov et al., 2016). The authenticity of the retrieved sequences was evaluated according to the characteristics of the

TABLE 1 | Sequencing information of three ancient samples.

Sample name	RawBaseNum	FilBaseNum	mappedBaseNum	Duplication rate (%)	Endogenous mtDNA (%)	Mitogenome Length >1X	mtDNA coverage >1X (%)	Mean coverage	Mean length for mapped reads	Data sources
GCZ4C	162,165,074,800	87,622,422,828	57,698	5.50	0.0001	14,764	90.21	3.13	57.27 bp	resequencing
TS1C	57,079,200	31,114,331	111,968	19.80	0.3599	16,299	99.43	5.42	140 bp	capture
TS2C	210,187,800	116,303,356	55,160	21.40	0.0474	15,054	92.13	2.61	138 bp	capture

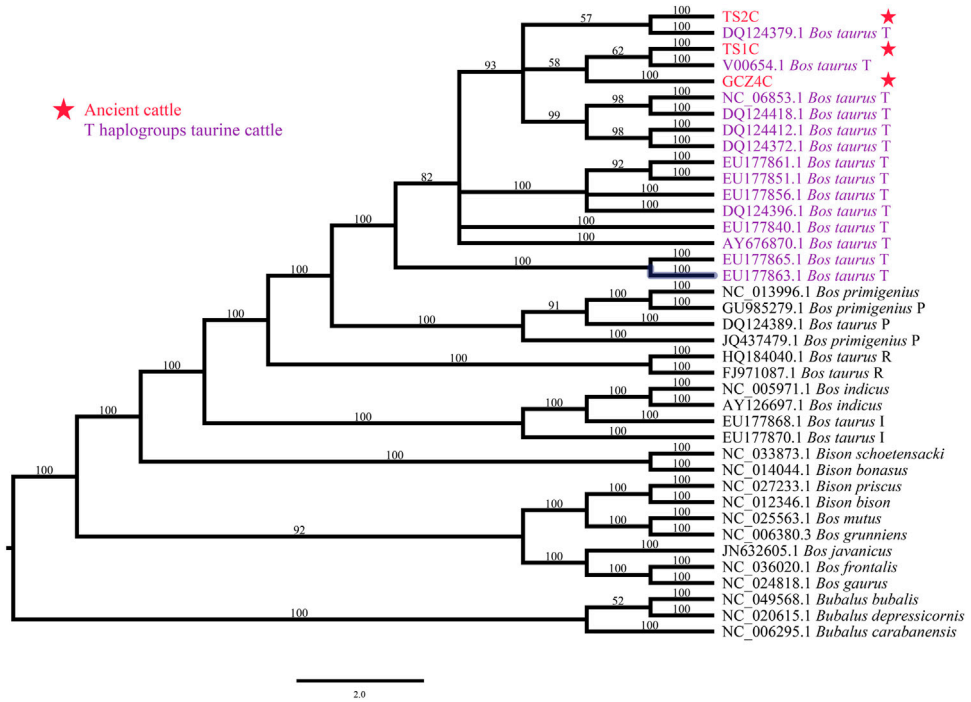

FIGURE 2 | Species identification though phylogenetic analysis. ID marked with red font and star represents the ancient sample, ID with purple font represents T-haplogroup taurine cattle.

aDNA (Hofreiter et al., 2001; Willerslev and Cooper, 2005) using the mapDamage software (Jonsson et al., 2013).

Assembly of Mitochondrial Genomes and Variant Calling

Three methods, including Mapping Iterative Assembler (MIAv.1.0) (Green et al., 2008), BCFtools (Li, 2011), and UnifiedGenotyper in the Genome Analysis Tool Kit (GATK, v3.5) (DePristo et al., 2011), were comparatively deployed to generate the mitogenome sequences. For MIA, the filtered bam files were converted to Fastq format using BEDtools (Quinlan and Hall, 2010). The program *ma* in MIA was used to generate complete sequences. The latter two software packages generate the complete sequence based on the mutation sites relative to the reference sequence using BCFtools consensus and manual generation, respectively. Finally, the complete mitogenome sequence of the ancient samples was obtained by comparing the differences in the complete sequences from different assembly methods and the filtered bam files.

Variant calling was carried out using BCFtools v1.8 mpileup (-d 5000 -Q 20 -a DP, AD, ADF, ADR, INFO/AD, INFO/ADF, INFO/ADR, SP, DV, DP4, DPR, INFO/DPR). Single nucleotide polymorphism (SNP) sites with <3× coverage depth were filtered out. UnifiedGenotyper in GATK3.5 was also used for the variation calling, whereas SNP sites with <3× coverage were discarded. Subsequently, the filtered SNP sites were manually checked and corrected according to the bam files visualized on the Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al., 2013). Eventually, common SNP loci were retained as the true mutation loci.

Sequence Alignment

All the captured and re-sequenced reads for the ancient samples were aligned to the reference mitogenome (GenBank accession No.: V00654) using the online MAFFT software with default parameters (Katoh et al., 2019). Positions embracing missing data or uncertain sites were removed using MEGA 7 (Kumar et al., 2016). Then the ratio of nonsynonymous and synonymous substitutions (dN/dS) was also calculated using MEGA 7. The 3 ancient consensus sequences were aligned to 36 others extant Bovinae mitogenomes from GenBank and called Dataset A (**Supplementary Table S2**), which included 1 *Bison,* 1 *Bison bonasus*, 1 *Bison schoetensacki*, 1 *Bison priscus*, 1 *Bubalus bubalis*, 1 *Bubalus depressicornis*, 1 *Bubalus carabanensis*, 1 *Bos frontalis*, 1 *Bos gaurus*, 1 *Bos grunniens*, 1 *Bos mutus*, 1 *Bos javanicus*, 3 *Bos primigenius,* 2 *Bos indicus*, and 19 *Bos taurus*. To further identify the haplogroup of the ancient samples, 510 cattle including *Bos taurus* and *Bos indicus* were aligned with ancient samples and called Dataset B (**Supplementary Table S3**).

Phylogenetic and Demographic Analyses

To construct phylogenetic trees and infer population structure, the nucleotide substitution model was selected using jModelTest 2.1.1 (Darriba et al., 2012), based on the model selection strategy of Bayesian information criteria. Bayesian phylogenetic trees were constructed using MrBayes 3.2.7 (Ronquist et al., 2012), with the following parameters: 20,000,000 generations of Markov Chain Monte Carlo (MCMC) chains, checking every 10,000, sampling every 2,000, and discarding the first 10% as burn in. The consensus tree was constructed using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

TABLE 2 | The haplotype and nucleotide diversity of T haplogroup cattle in dataset B.

Group	Region	No. of samples	Variable sites	H	K	Hd ± SD	π ± SD
Ancient	Ancient	3	25	3	16.667	1.000 ± 0.272	0.00122 ± 0.00037
East Asia	East Asia	58	152	53	9.845	0.996 ± 0.004	0.00072 ± 0.00004
Africa	Africa	68	201	66	13.874	0.999 ± 0.003	0.00101 ± 0.00005
America	America	39	96	29	10.227	0.974 ± 0.014	0.00075 ± 0.00005
Europe	Europe	105	319	99	11.204	0.999 ± 0.002	0.00082 ± 0.00003
West Asia	West Asia	16	78	16	12.992	1.000 ± 0.022	0.00095 ± 0.00009
North China	North China	9	38	8	9.944	0.972 ± 0.064	0.00073 ± 0.00012
South China	South China	64	114	33	10.209	0.958 ± 0.014	0.00075 ± 0.00005
China	China	73	141	41	10.295	0.967 ± 0.011	0.00075 ± 0.00005

TABLE 3 | Population pairwise *Fst* values (the below diagonal) and gene flow (Nm, the above diagonal) among T haplogroup of dataset B and ancient samples.

Region	Ancient	East Asia	Africa	America	Europe	West Asia	North China	South China	China
Ancient		3.96	1.66	3.44	3.80	4.30	4.92	7.70	7.69
East Asia	0.1120		1.69	4.95	8.49	7.98	16.20	7.65	9.02
Africa	0.2310	0.2279		4.28	3.33	2.88	1.85	2.85	2.75
America	0.1269	0.0918	0.1047		16.68	8.82	5.99	11.04	11.22
Europe	0.1164	0.0556	0.1307	0.0291		22.25	10.52	12.16	13.60
West Asia	0.1042	0.0590	0.1480	0.0537	0.0220		12.98	12.37	14.12
North china	0.0922	0.0299	0.2130	0.0770	0.0454	0.0371		10.23	20.97
South China	0.0610	0.0613	0.1495	0.0433	0.0395	0.0389	0.0466		−38.50
China	0.0611	0.0525	0.1540	0.0427	0.0355	0.0342	0.0233	−0.0132	

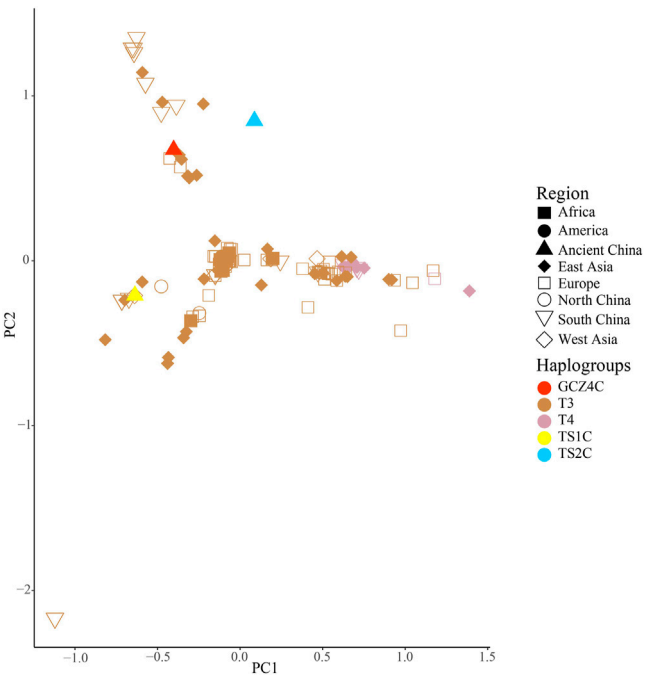

FIGURE 3 | PCA analysis of ancient samples and T3/T4 species haplogroup *Bos*.

Principal component analysis (PCA) was used to further verify the genetic affinities of the ancient samples among the genus *Bos* using the R Package "adegenet" (Jombart, 2008). Part of dataset A was interrogated to infer demographics using BEAST 2.6.3 (Bouckaert et al., 2014). The ages of the ancient cattle were set at 4,000 YBP, which was the time of the culture in the respective sites of origin. We used Strict Clock as the clock model and Coalescent Bayesian Skyline as a distribution of the tree priors. The MCMC chains were set to 1000,000,000 generations. The log file was checked using Tracer v1.7.1 (Rambaut et al., 2018), and the Bayesian Skyline Plot (BSP) was obtained using Tracer v1.7.1. The optimal nucleotide substitution model was derived using MEGA 7 estimation for the maximum likelihood (ML) tree. Thereafter, an ML tree was built using MEGA 7 with 1,000 bootstrapping replicates and a timetree inferred using the Reltime method (Tamura et al., 2012). The timetree was computed using two calibration constraints from TimeTree (Kumar et al., 2017), including the time of divergence between *Bos taurus*, *Bos primigenius*, *Bos taurus*, and *Bos indicus*.

T haplogroups of Dataset B was used to calculate population pairwise F_{ST} values, the gene flow (Nm) and the net genetic distance using DNAsp 5.10 (Librado and Rozas, 2009). Meanwhile, numbers of haplotypes and variable sites, haplotype diversity (Hd), nucleotide diversity (Pi), and the average number of nucleotide differences (K) were estimated using the same data and software. Besides, T haplogroups of Dataset B was used to conduct PCoA (Principal Coordinate Analysis) among populations. The neighbor net (Bryant and Moulton, 2004) was constructed using splitstree5 (Huson and Bryant, 2006) based on *Fst* value. PCoA analysis was performed using the OmicStudio tools at https://www.omicstudio.cn/tool. The final result was graphed using R Package "ggplot2" (Hadley, 2016).

RESULTS

aDNA Extraction and Sequencing

A total of eight ancient samples, including four bovines (4,350–3,850 YBP) from the Taosi site and four bovines (4,150–3,100 YBP) from

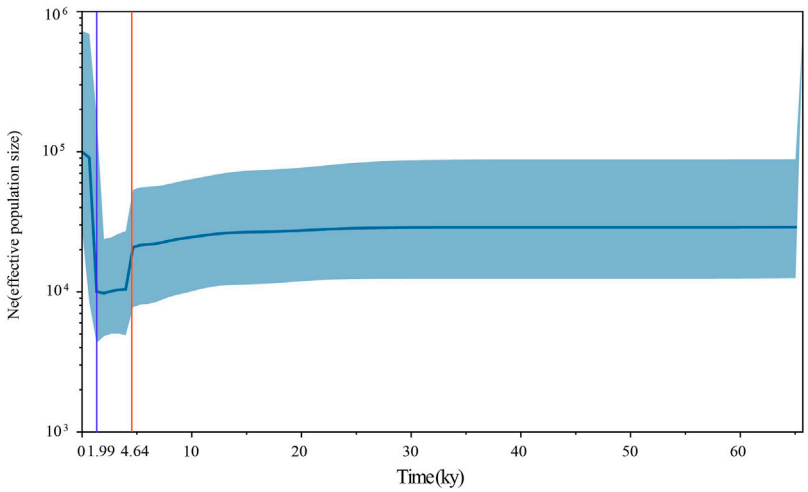

FIGURE 4 | The Bayesian Skyline Plot of ancient and modern *Bos*. The Y-axis represents the effective number of populations, and the *X*-axis represents the timeline. The orange line shows the timepoint of effective population decline and the blue line shows the timepoint of the rapid increase of effective population.

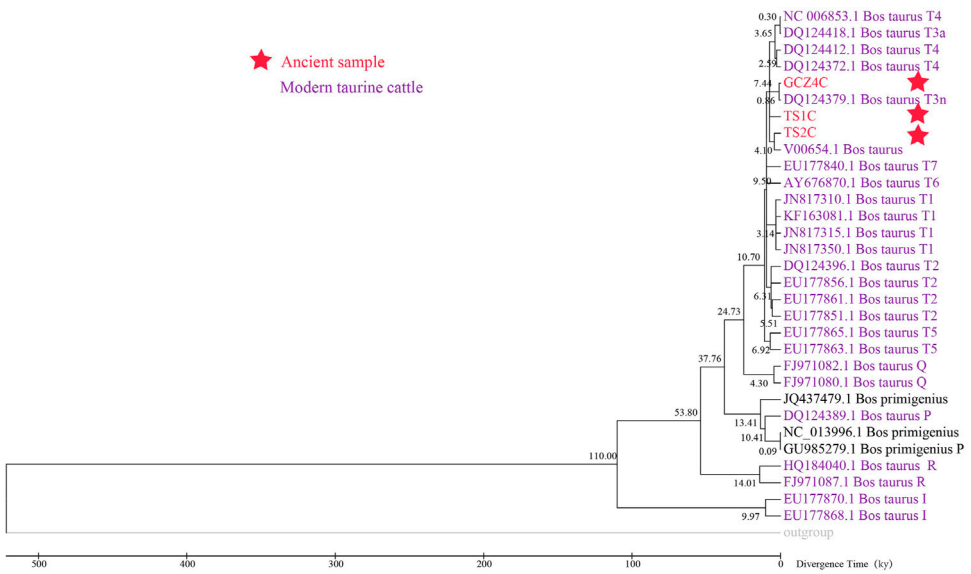

FIGURE 5 | The Maximum Likelihood timetree of ancient and modern *Bos*.

the Guchengzhai site, were subjected to DNA extraction. All qualified aDNA extractions were selected for shotgun sequencing. However, owing to the typically low levels of endogenous DNA in these samples, only one sample from the Guchengzhai site (GCZ4C) obtained sufficient endogenous reads (0.0001%) to yield a near-complete mitogenome. Therefore, the remaining samples were subjected to mtDNA enrichment and hybridization sequencing using *Bos* hybrid baits. Total sequences of the near-complete mitogenome for the two samples from the Taosi site (TS1C and TS2C) were retrieved after enrichment using baits. After capture sequencing, the endogenous mtDNA read ratio increased to 0.3599% and 0.0474% for TS1C and TS2C, respectively (**Table 1**), resulting in the recovery of three complete ancient mitogenomes for late Neolithic archeological sites in the Middle Yellow River region. The average length of hybrid capture reads was approximately 140 bp, and the average length of resequencing reads was approximately 57 bp (**Supplementary Figure S1**). Damage pattern analyses showed a high C–T mutation at the 5′ end and a high G–A mutation at the 3′ end (**Supplementary Figure S2**), showing typical damage and fragmentation patterns characteristic of endogenous aDNA.

Mitogenome Assembly of the Ancient Samples

By considering the effect on the integrity and accuracy of ancient mitogenome assembly, we combined three methods including

GATK, BCFtools and MIA to generate mitogenome sequences. The number of duplicated sequences varied from 5.50% to 21.6% (**Table 1**). As a result, we obtained a 16,299 bp (coverage 99.43%) mitogenome sequence for TS1C, a 15,054 bp (coverage 92.13%) sequence for TS2C, and a 14,764 bp (coverage 90.21%) for GCZ4C (**Table 1**). SNP sites with ≥3× coverage were used in subsequent analyses (**Supplementary Table S4**).

Species Identification

To determine the species of the ancient bovine samples, the near-complete mitogenome sequences were combined with 36 extant Bovinae mitogenomes, including the genus *Bison*, *Bubalus*, and *Bos* (**Supplementary Table S2**). The phylogenetic trees illustrated all three ancient samples belonging to the genus *Bos* rather than *Bison* or *Bubalus*, specifically close to the species *Bos taurus* (**Figure 2** and **Supplementary Figure S3**).

Phylogeny and Genetic Diversity of the Ancient Cattle

The mutation sites for each sample were discovered by alignment against the taurine cattle reference mitogenome (GenBank accession no.: V00654) (**Supplementary Table S4**). All ancient mitogenomes contained 13 coding genes, of which *COX3* had the most variable sites and the most non-synonymous SNPs. The values of dN/dS of all genes except *COX3* and *ND5* were lower than 1 (**Supplementary Table S5**), suggesting potential purifying selection on the mitochondrial functions in the ancient bovine in North China. Based on the key SNPs in the mitogenome, all three ancient samples were classified as haplogroup T3. Specifically, TS1C from the Taosi site belonged to haplotype T3k, while another sample, TS2C, from the Taosi site and GCZ4C from the Guchengzhai site belonged to haplotype T3n (**Supplementary Table S4**).

Relationship of Ancient Cattle With Modern *Bos* Cattle

To further investigate the phylogenetic relationships among the approximately 4,000-year-old cattle in North China, a Bayesian phylogenetic tree was inferred with 510 extant mitogenome sequences of *Bos taurus* and *Bos indicus* around the world (**Supplementary Table S3**). The Bayesian tree showed that all ancient cattle were on the T3 branch (**Supplementary Figure S4**). PCA (**Supplementary Figure S5A and S5B**) and PCoA (**Supplementary Figure S6A**) supported the phylogeny of the T3 haplogroup of the ancient samples. The population pairwise F_{ST} values, the gene flow, and the net genetic distance, illustrated that the ancient Chinese cattle had the closest genetic relationship to Chinese taurine cattle, followed by populations from East Asia, Europe, West Asia, America, and Africa (**Table 2**). Further subdivision of Chinese taurine cattle according to region showed that population in T haplogroups had high haplotype and nucleotide diversity (**Table 3**), and the ancient Chinese cattle were more closely related to the south Chinese taurine cattle than the north taurine Chinese cattle (**Table 2** and **Supplementary Table S6**). Moreover, PCA (**Figure 3**) and PCoA (**Supplementary Figure S6B**) analysis of ancient samples and T3/T4 cattle suggested that TS1C was genetically close to modern *Bos taurus* from Southern China, while TS2C and GCZ4C were closer to cattle from Europe and East Asia.

Demographic Inferences

To assess the population dynamics of taurine cattle in China, a BSP was produced using 25 mitogenomes. The BSP points to a rapid decrease in the female effective population size approximately 4.65 kya and a steep increase in population approximately 1.99 kya (**Figure 4**). To infer the divergence time between ancient and modern cattle, a ML tree was constructed with 3 *Bos primigenius* and 25 *Bos taurus* cattle, using 1 *Bison bonasus* as the outgroup (**Figure 5**). Concerning the differentiation of the major branches in the phylogenetic tree, the T3 and T4 sub-lineages possibly emerged approximately 7.44 kya, and at this time, the haplotype of sample TS1C appeared among the early domesticated cattle. TS2C, another sample from the Taosi site, was divergent to modern populations approximately 4.10 kya, in accordance with the period of rapid decrease in the female effective population size. GCZ4C was divergent to the modern T3n taurine population approximately 0.86 kya, suggesting possible long-term dilution of the ancient northern Chinese cattle.

DISCUSSION

Cattle genome research suggests that the Chinese taurine cattle originated from the Near East (Cai et al., 2014; Cai et al., 2018; Verdugo et al., 2019). According to the specific region where cattle remains were unearthed in China, it is believed that the earliest domesticated cattle in China occurred in northwestern China approximately 5,000–4,000 YBP (Flad et al., 2009). In addition, aDNA analysis of ancient Chinese cattle remains indicated that the cattle in northern China were taurine cattle since the Bronze Age (Cai et al., 2014). However, the abundant auroch remains prior to 5,000 YBP recovered from extensive regions in China suggest a potential early utilization and native domestication of Chinese bovines in the early Neolithic period (Zhang et al., 2013). Evidence from aDNA analysis of indigenous aurochs in central and northeast China during the Neolithic period adds to the uncertainty of the origins of the Chinese taurine cattle (Brunson et al., 2016b).

In this study, we obtained three nearly complete ancient mitogenomes from two ~4,000 YBP archeological sites in North China and, for the first time, reported the complete mitogenomes from ancient bovines in this region. By determining all ancient samples as taurine cattle, no zebu population was found in these sites, providing further evidence that zebu cattle did not spread to northern China 4,000 YBP (Yue et al., 2014).

As one of the representative Longshan culture sites in China, the Taosi site is unique in that it contains numerous animal remains, such as pig mandibles, which proved that the ancestors living at the Taosi site mastered advanced animal husbandry technologies and lived a long-term, settled agricultural life (Nu, 2013). Different analyses of animal remains at the Taosi site help to understand the relationship between humans and animals under agricultural development in China. Isotopic analysis of the bones of pigs, cattle, and dogs indicated that the staple food of cattle at the Taosi site was a by-product of millet (Chen et al., 2012). Meanwhile, strontium isotope analysis

indicated a mix of local and non-local animals at the Taosi site (Zhao et al., 2011). Additionally, the cattle were used for a variety of craft products at the Taosi site (Brunson et al., 2016a).

In our study, we analyzed the domestication status and genetic diversity of Chinese taurine cattle at the Taosi site at the level of the ancient mitogenome, which could be more accurate and direct than modern specimens. Both PCA, PCoA and phylogenetic analysis demonstrated that the ancient individuals of the Taosi site belonged to the T3 haplogroup, suggesting their origin from the Near East (Achilli et al., 2008), which was consistent with studies on mtDNA fragments (Cai et al., 2014; Brunson et al., 2016b). As shown, TS2C was localized to haplogroup T3n from East Asia, defined by 16119C, which probably had a founder effect (Cai et al., 2014). These data indicated that the appearance of ancient domestic cattle in northern China predated the Taosi site. A previous study based on a 294 bp mtDNA sequence showed that 15 ancient cattle in the Taosi site constituted 5 haplotypes (Cai et al., 2014). In our study, TS1C and TS2C originated from the same archeological site or even the same trench, but their haplotypes were different. This result reveals a high level of genetic diversity in the early cattle population at the Taosi site. TS1C diverged from the T3 haplogroup approximately ~7.44 kya, while TS2C separated from the T3 haplogroup ~4.10 kya, suggesting a rich diversity of cattle and the mixed presence of native and non-native cattle at the Taosi site.

The Guchengzhai site, dating back to the late period of the third Wangwan culture from Longshan culture in Henan Province, is one of the best-preserved ancient cities (Cai, 2002). A large number of wine vessels and animal remains, including pigs, cattle, and sheep have been found (Cai, 2008; Xu et al., 2013), indicating a locally developed agricultural level. However, genetic analysis of cattle remains at the Guchengzhai site has not been conducted. Through aDNA analysis, we generated the first complete mitogenome for Guchengzhai cattle and found that the bovine remains from this site belonged to taurine cattle. Interestingly, GCZ4C carries the same haplogroup (T3) as TS2C from the Taosi site, which has dominating founder effect in East Asian domestic cattle. Therefore, the similar genetic background shared by ancient cattle at Taosi and Guchengzhai site may indicate the early arrival of cattle at the northern China before the Longshan Culture period.

The F_{ST} values showed that ancient Chinese cattle have the closest affinity to modern Chinese cattle, indicating possible genetic contribution of ancient Chinese cattle populations to modern populations. Previously, whole-genome analysis of an ancient cattle from the Shimao site indicated that there were at least two migration events that occurred in Northern China (Chen et al., 2018). The closer genetic relationship of ancient Chinese cattle with the southern Chinese taurine than with the northwest China taurine may suggest that the ancient samples had genetic contributions to modern southern Chinese taurine cattle, and the northern taurine cattle may have been replaced by other taurine cattle. The BSP in our study also showed that the effective population size of cattle began to decline sharply approximately 4,500 YBP, in line with the arrival of taurine cattle into China. We also found that the ancient taurine cattle (TS2C) were separated approximately 4,000 YBP, which is a crucial period in the study of animal domestication in ancient China (Flad et al., 2009; Zhang et al., 2013; Brunson et al., 2016b; Cai et al., 2018). Meanwhile, the rapid increase in the effective cattle population approximately 1,900 YBP was also linked to the acceleration of communication between East and West at that time. During this period, the Han Empire in ancient China established the Silk Road to communicate with Western civilization. This information also suggests that cattle may have entered China *via* the Silk Road, increasing the effective population size of the Chinese cattle population.

CONCLUSION

In this study, we for the first time obtained three near-complete mitogenomes of cattle about 4,000 YBP from North China and confirmed the ~4,000-year-old bovine from North China as taurine cattle. All ancient cattle from Taosi and Guchengzhai sites belonged to the T3 haplogroup, suggesting their origin from Near East. The high affinity between ancient samples and southern Chinese taurine cattle indicated that the ancient Chinese cattle had a genetic contribution to taurine cattle of South China. A rapid decrease in the female effective population size ca. 4.65 kya and a steep increase ca. 1.99 kya occurred in Chinese taurine cattle.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in the China National GeneBank under accession number CNP0002141.

AUTHOR CONTRIBUTIONS

XBZ and HX designed this project; XZ, LY, and LH performed the experiments; XZ, LY, LH, HL, and HX analyzed the data; XZ, LY, HX, and XBZ interpreted the data and drafted the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

FUNDING

This work was funded by the Foundation for Distinguished Young Talents (2019KQNCX167) and Creative Team (2019KCXTD006) in Higher Education of Guangdong, the National Natural Science Foundation of China (31961133031), the Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding (2019B030301010), and the Scientific Research Fund for Postdoc of Foshan City (BKS209071). The funding bodies contributed nothing to the study design, data analyses, data interpretation, or manuscript preparation.

ACKNOWLEDGMENTS

We express sincere thanks to Prof. Jing Yuan (Institute of Archaeology, Chinese Academy of Social Science) for

generously providing the ancient samples, Prof. Zhenglong Gu, Prof. Yi Zhang and Ms. Yumei Wu (China Agricultural University) for great help with experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.759827/full#supplementary-material

Supplementary Figure S1 | Distribution of fragment lengths of ancient DNA from three ancient samples.

Supplementary Figure S2 | The DNA damage patterns of the ancient samples. Misincorporations frequencies are represented by the first and last 25 nucleotides of reads aligned to the bovine mitochondrial reference genome V00654.

Supplementary Figure S3 | Species identification by phylogenetic analysis of three ancient samples with 36 extant Bovinae mitogenomes, respectively.

Supplementary Figure S4 | Phylogenetic tree of three ancient mitogenomes and 510 cattle of Bos genus.

Supplementary Figure S5 | PCA of ancient samples and species from the *Bos* genus. **(A)** shows the relationship of three ancient samples to all *Bos* genus while **(B)** shows their relationship to the T haplogroup *Bos* genus.

Supplementary Figure S6 | PCoA of ancient samples and species from the *Bos* genus. **(A)** shows the relationship of three ancient samples to the T haplogroup *Bos* genus while **(B)** shows their relationship to the T3/T4 species haplogroup *Bos*.

REFERENCES

Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L., Al-Zahery, N., et al. (2008). Mitochondrial Genomes of Extinct Aurochs Survive in Domestic Cattle. *Curr. Biol.* 18 (4), R157–R158. doi:10.1016/j.cub.2008.01.019

Ajmone-Marsan, P., Garcia, J. F., and Lenstra, J. A. (2010). On the Origin of Cattle: How Aurochs Became Cattle and Colonized the World. *Evol. Anthropol.* 19 (4), 148–157. doi:10.1002/evan.20267

Andrews, S. (2010). *FASTQC. A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Institute. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a Software Platform for Bayesian Evolutionary Analysis. *Plos Comput. Biol.* 10 (4), e1003537. doi:10.1371/journal.pcbi.1003537

Brunson, K., He, N., and Dai, X. (2016a). Sheep, Cattle, and Specialization: New Zooarchaeological Perspectives on the Taosi Longshan. *Int. J. Osteoarchaeol.* 26 (3), 460–475. doi:10.1002/oa.2436

Brunson, K., Zhao, X., He, N., Dai, X., Rodrigues, A., and Yang, D. (2016b). New Insights into the Origins of oracle Bone Divination: Ancient DNA from Late Neolithic Chinese Bovines. *J. Archaeological Sci.* 74, 35–44. doi:10.1016/j.jas.2016.08.008

Bryant, D., and Moulton, V. (2004). Neighbor-Net: an Agglomerative Method for the Construction of Phylogenetic Networks. *Mol. Biol. Evol.* 21 (2), 255–265. doi:10.1093/molbev/msh018

Cai, D., Sun, Y., Tang, Z., Hu, S., Li, W., Zhao, X., et al. (2014). The Origins of Chinese Domestic Cattle as Revealed by Ancient DNA Analysis. *J. Archaeological Sci.* 41, 423–434. doi:10.1016/j.jas.2013.09.003

Cai, D., Zhang, N., Zhu, S., Chen, Q., Wang, L., Zhao, X., et al. (2018). Ancient DNA Reveals Evidence of Abundant Aurochs (Bos Primigenius) in Neolithic Northeast China. *J. Archaeological Sci.* 98, 72–80. doi:10.1016/j.jas.2018.08.003

Cai, Q. (2002). Guchengzhai Longshan chengzhi yu zhongyuan wenming de xingcheng. *Cultural Relics of Central China* 46 (6), 27–32.

Cai, Q. (2008). Investigation of the Pre-history Agriculture and the Structure of Man's Food in the Shuangji River basin. *J. Huanghe Sci Uni.* 10 (5), 46–49. doi:10.3969/j.issn.1008-5424.2008.05.012

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, X., Yuan, J., Hu, Y., He, N., and Wang, C. (2012). Study on Animal Husbandry in Taosi Site: Evidence Derived from Stable Isotope Analysis. *Archaeology* 75, 843–850.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., et al. (2013). Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave bear Reconstructed from Ultrashort DNA Fragments. *Proc. Natl. Acad. Sci. U. S. A.* 110 (39), 15758–15763. doi:10.1073/pnas.1314445110

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: More Models, New Heuristics and Parallel Computing. *Nat. Methods* 9 (8), 772. doi:10.1038/nmeth.2109

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data. *Nat. Genet.* 43 (5), 491–498. doi:10.1038/ng.806

Diamond, J. (2002). Evolution, Consequences and Future of Plant and Animal Domestication. *Nature* 418 (6898), 700–707. doi:10.1038/nature01019

Ebersbach, R. (2002). *Von Bauern und Rindern: Eine Ökosystemanalyse zur Bedeutung der Rinderhaltung in bäuerlichen Gesellschaften als Grundlage zur Modellbildung im Neolithikum*. Basel: Schwabe.

Felius, M., Beerling, M.-L., Buchanan, D., Theunissen, B., Koolmees, P., and Lenstra, J. (2014). On the History of Cattle Genetic Resources. *Diversity* 6 (4), 705–750. doi:10.3390/d6040705

Flad, R., Yuan, J., and Li, S. (2009). On the Source and Features of the Neolithic Domestic Animals in the Gansu and Qinghai Region, China. *Archaeology* 5 (8), 80–86.

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., et al. (2008). A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell* 134 (3), 416–426. doi:10.1016/j.cell.2008.06.021

Guo, R. (2018). Archaeological Observations on the Food Structure of the Prehistoric Ancestors in Xinmi of Henan. *Agri. Archaeol.* 3, 26–32.

Hadley, W. (2016). *ggplot2-Elegant Graphics for Data Analysis*. New York: Springer International Publishing.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. *Nat. Rev. Genet.* 2 (5), 353–359. doi:10.1038/35072071

Huson, D. H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* 23 (2), 254–267. doi:10.1093/molbev/msj030

Jombart, T. (2008). Adegenet: a R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24 (11), 1403–1405. doi:10.1093/bioinformatics/btn129

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., and Orlando, L. (2013). mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters. *Bioinformatics* 29 (13), 1682–1684. doi:10.1093/bioinformatics/btt193

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief Bioinform.* 20 (4), 1160–1166. doi:10.1093/bib/bbx108

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: a Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34 (7), 1812–1819. doi:10.1093/molbev/msx116

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

Li, H. (2011). A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data. *Bioinformatics* 27 (21), 2987–2993. doi:10.1093/bioinformatics/btr509

Librado, P., and Rozas, J. (2009). DnaSP V5: a Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics* 25 (11), 1451–1452. doi:10.1093/bioinformatics/btp187

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for Two Independent Domestications of Cattle. *Proc. Natl. Acad. Sci. U. S. A.* 91 (7), 2757–2761. doi:10.1073/pnas.91.7.2757

Mannen, H., Yonezawa, T., Murata, K., Noda, A., Kawaguchi, F., Sasazaki, S., et al. (2020). Cattle Mitogenome Variation Reveals a post-glacial Expansion of Haplogroup P and an Early Incorporation into Northeast Asian Domestic Herds. *Sci. Rep.* 10 (1), 20842. doi:10.1038/s41598-020-78040-8

Nu, H. (2013). "The Longshan Period Site of Taosi in Southern Shanxi Province," in *A Companion to Chinese Archaeology*. Oxford: Blackwell Publishing Ltd., 255–277. doi:10.1002/9781118325698.ch13

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* 32 (2), 292–294. doi:10.1093/bioinformatics/btv566

Patel, A. K. (2009). Occupational Histories, Settlements, and Subsistence in Western India: What Bones and Genes Can Tell Us about the Origins and Spread of Pastoralism. *Anthropozoologica* 44 (1), 173–188. doi:10.5252/az2009n1a8

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., et al. (2016). EAGER: Efficient Ancient Genome Reconstruction. *Genome Biol.* 17, 60. doi:10.1186/s13059-016-0918-z

Qiu, Z. (2006). Quaternary Environmental Changes and Evolution of Large Mammals in North China. *Vertebrata Palasiatica* 44 (2), 109–132. doi:10.3969/j.issn.1000-3118.2006.02.001

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

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67 (5), 901–904. doi:10.1093/sysbio/syy032

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Syst. Biol.* 61 (3), 539–542. doi:10.1093/sysbio/sys029

Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., and Kumar, S. (2012). Estimating Divergence Times in Large Molecular Phylogenies. *Proc. Natl. Acad. Sci. U. S. A.* 109 (47), 19333–19338. doi:10.1073/pnas.1213199109

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* 14 (2), 178–192. doi:10.1093/bib/bbs017

Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Maisano Delser, P., et al. (2019). Ancient Cattle Genomics, Origins, and Rapid Turnover in the Fertile Crescent. *Science* 365 (6449), 173–176. doi:10.1126/science.aav1002

Wang, S., Wang, Z., and He, N. (2011). Study on the Charcoal Samples in the Excavations of Taosi Site. *Archiaeology* 3, 91–96.

Willerslev, E., and Cooper, A. (2005). Review Paper. Ancient DNA. *Proc. R. Soc. B.* 272 (1558), 3–16. doi:10.1098/rspb.2004.2813

Xu, J., Mo, D., Wang, H., and Zhou, K. (2013). Preliminary Research of Environment Archaeologyin Zhenshui River, Xinmi City, Henan. *Quat. Sci.* 33 (5), 954–964. doi:10.3969/j.issn.1001-7410.2013.05.13

Yue, X., Li, R., Liu, L., Zhang, Y., Huang, J., Chang, Z., et al. (2014). When and How Did Bos indicus Introgress into Mongolian Cattle? *Gene* 537 (2), 214–219. doi:10.1016/j.gene.2013.12.066

Zhang, H., Paijmans, J. L. A., Chang, F., Wu, X., Chen, G., Lei, C., et al. (2013). Morphological and Genetic Evidence for Early Holocene Cattle Management in Northeastern China. *Nat. Commun.* 4, 2755. doi:10.1038/ncomms3755

Zhang, L., Jia, S., Plath, M., Huang, Y., Li, C., Lei, C., et al. (2015). Impact of ParentalBos taurusandBos indicusOrigins on Copy Number Variation in Traditional Chinese Cattle Breeds. *Genome Biol. Evol.* 7 (8), 2352–2361. doi:10.1093/gbe/evv151

Zhao, C., Yuan, J., and He, N. (2011). Strontium Isotope Analysis of Archaeological Fauna from the Taosi Site, Xiangfen County, Shanxi Province. *Quat. Sci.* 31 (1), 22–28. doi:10.3969/j.issn.1001-7410.2011.01.04

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

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

frontiers in Genetics

REVIEW
published: 03 December 2021
doi: 10.3389/fgene.2021.761728

Review: Balancing Selection for Deleterious Alleles in Livestock

Martijn F. L. Derks[1,2]* *and Marije Steensma*[1]

[1]*Animal Breeding and Genomics, Wageningen University and Research, Wageningen, Netherlands,* [2]*Topigs Norsvin Research Center, Beuningen, Netherlands*

Harmful alleles can be under balancing selection due to an interplay of artificial selection for the variant in heterozygotes and purifying selection against the variant in homozygotes. These pleiotropic variants can remain at moderate to high frequency expressing an advantage for favorable traits in heterozygotes, while harmful in homozygotes. The impact on the population and selection strength depends on the consequence of the variant both in heterozygotes and homozygotes. The deleterious phenotype expressed in homozygotes can range from early lethality to a slightly lower fitness in the population. In this review, we explore a range of causative variants under balancing selection including loss-of-function variation (i.e., frameshift, stop-gained variants) and regulatory variation (affecting gene expression). We report that harmful alleles often affect orthologous genes in different species, often influencing analogous traits. The recent discoveries are mainly driven by the increasing genomic and phenotypic resources in livestock populations. However, the low frequency and sometimes subtle effects in homozygotes prevent accurate mapping of such pleiotropic variants, which requires novel strategies to discover. After discovery, the selection strategy for deleterious variants under balancing selection is under debate, as variants can contribute to the heterosis effect in crossbred animals in various livestock species, compensating for the loss in purebred animals. Nevertheless, gene-assisted selection is a useful tool to decrease the frequency of the harmful allele in the population, if desired. Together, this review marks various deleterious variants under balancing selection and describing the functional consequences at the molecular, phenotypic, and population level, providing a resource for further study.

Keywords: balancing selection, overdominance, loss-of-function, artificial selection, animal breeding

OPEN ACCESS

Edited by:
Denis Milan,
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), France

Reviewed by:
Martin Johnsson,
Swedish University of Agricultural Sciences, Sweden
Julie Demars,
Institut National de Recherche pour l'Agriculture, l'aAimentation et l'Environnement (INRAE), France

***Correspondence:**
Martijn F. L. Derks
martijn.derks@wur.nl

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

Received: *20 August 2021*
Accepted: *19 November 2021*
Published: *03 December 2021*

Citation:
Derks MFL and Steensma M (2021) Review: Balancing Selection for Deleterious Alleles in Livestock. Front. Genet. 12:761728. doi: 10.3389/fgene.2021.761728

INTRODUCTION

Since the widespread use of artificial insemination in livestock, a small number of popular sires can 'dominate' the breeding population. Consequently, the effective populations size (Ne) decreases, causing deleterious alleles to rise in frequency. Hence, the inherited defects generally involve unique 'founder' variants (Fasquelle et al., 2009). Founder variants are variants observed at high frequency in a specific population that derive from a single influential ancestor. Generally, these type of deleterious alleles are purged from a population by (natural) selection. This purging is efficient for dominant deleterious alleles that lower the fitness of heterozygous animals. However, as a function of the low frequency, recessive deleterious alleles are generally masked from natural selection by a dominant non-deleterious allele, and will therefore be easily passed on into the next generation.

In population genetic theory, the amount of deleterious alleles arising from mutation is expected to be equal to the amount of deleterious alleles purged by (natural) selection, resulting in a mutation - selection balance (Charlesworth and Willis 2009). Hence, harmful alleles are kept in balance at a low frequency in the population, due to the elimination through purifying selection and the occurrence of new variants (Old, 1993). However, recent studies have revealed various harmful alleles present at a moderate to high frequency in the population (Georges et al., 2019). This high frequency of recessive harmful alleles can be a result of genetic drift, i.e. the random fluctuations in the numbers of gene variants in a population, especially in populations with low effective population size (Bosse et al., 2019). However, genetic drift alone cannot explain the high frequency of some relatively high frequency variants identified. Hence, a second explanation stems from variants exhibiting antagonistic pleiotropic effects, driving the frequency of such variants (Hedrick, 2015). The concept of antagonistic pleiotropy involves a trade-off between a beneficial effect on one trait and a detrimental effect on another trait, caused by a single variant (Hedrick, 1999). One common mechanism of antagonistic pleiotropy is heterozygous advantage, also called overdominance. Heterozygous advantage depends on an interplay between the strength of the advantage in heterozygotes and the negative consequences in homozygotes, leading to a balance between purifying selection against mutant homozygotes, and positive selection on heterozygotes (Hedrick, 2015). In addition, the advantage might be affected by numerous population genetic factors including the allele frequency, the sex, and the selection goal. Together, balancing selection refers to selective processes by which alleles are maintained in a population at frequencies larger than expected from genetic drift alone (Siewert and Voight, 2017).

Due to the wide increase of genomic and phenotypic data in livestock breeding, numerous examples of deleterious variants under balancing selection have been published. However, this mechanism is likely far more common than expected, as many deleterious variants remain "hidden" in the population (Charlesworth, 2006). The artificial selection for the variants is performed based on the advantage in heterozygotes, while the negative consequence in homozygotes is only expressed at higher frequencies. In addition, the molecular mechanisms underlying balancing selection are often poorly understood (Fijarczyk and Babik, 2015). Most examples refer to a single allele that exhibits antagonistic effects on distinct traits, usually resulting from a dominant effect in heterozygotes, while a complete loss of function of a gene leads to the deleterious phenotype in homozygous mutants. However, also multiple variants in close linkage disequilibrium (LD) have been described having separate effects on different traits (Charlesworth, 2006). Therefore, a similar effect can be caused by two variants that are closely linked on the same haplotype, each individually affecting a different trait. However, it remains a large challenge to disentangle the individual variant effects given that the variants are strongly linked (Li et al., 2015). Finally, providing an insight in the management of these variants in populations and selection schemes would be useful for a healthy population in the long term (Georges et al., 2019).

The aim of this literature review is to get an insight in known variants under balancing selection in livestock and possibly their underlying molecular mechanisms. In addition, this review aims at discussing several aspects including the management of these variants in populations and selection schemes.

HARMFUL ALLELES UNDER BALANCING SELECTION IN LIVESTOCK

This review focusses on harmful alleles under balancing selection that have been discovered in various livestock species including pig (*Sus scrofa*), cow (*Bos taurus*), sheep (*Ovis aries*), chicken (*Gallus gallus*) and the horse (*Equus caballus*). Deleterious alleles under balancing selection can have different modes of inheritance: dominant or recessive modes are most common. Additional types of inheritance include sex-linked (affecting one of the sex chromosomes), codominant (wherein the alleles of a gene pair in a heterozygote are fully expressed), and polar overdominance (where the phenotype is dependent on the parent of origin). In this section, we describe in detail harmful alleles under balancing selection, the affected traits, and the molecular background. The molecular background of the alleles under balancing selection includes a wide variety of types from single nucleotide polymorphism (mostly affecting a single amino acid) to complex large structural changes. Overall, most variants described lead to a loss of function of the affected gene, resulting in a functional change or the absence of the protein product. In this section we review the various types of molecular mechanism and the functional consequence of the alleles under balancing selection per species. An overview of the examples described in this review are presented in **Table 1**.

Pig (*Sus scrofa*)

In a Large White breed, a recessive lethal deletion within the *BBS9* gene is assumed to cause fetal lethality in mutant homozygotes (**Figure 3A**). The deletion suggests antagonistic pleiotropic effects on fertility and growth, as it indicates an increase infeed intake and growth in heterozygotes. It is likely that the *BBS9* gene is under balancing selection, due to the moderate carrier frequency of 10.8% and the appearance of positive phenotypic effects in carriers (Derks et al., 2018). The fetal lethality in pigs is caused by a 212-kb deletion in the Bardet-Biedl Syndrome 9 (*BBS9*) gene. The deletion causes skipping of 4 coding and 4 3′UTR exons, resulting in direct splicing from exon 19 to exon 28. The deletion leads to a frameshift that results in a truncated protein due to a premature stop codon. This truncated BBS9 protein has lost its function (causing increased growth rate in heterozygous pigs). The 212-kb deletion affects a region which contains *BMPER cis*-regulatory elements and therefore reduces the expression of the *BMPER* gene which seems to cause fetal mortality in mutant homozygous pigs (Derks et al., 2018).

A second deleterious allele in pigs has been described causing leg weakness syndrome (lameness) in mutant homozygotes. The allele increases muscle depth in heterozygotes explaining the high frequency (22%) in the Large White breed (Matika et al., 2019). Leg weakness in pigs is caused by a stop-gained variant replacing

TABLE 1 | An overview of the 18 examples of alleles with a heterozygote advantage and a homozygote disadvantage in livestock species.

Species (breed)	Trait (References)	Gene(s) involved	Type of variant	Inheritance	Heterozygote advantage	Homozygote disadvantage
Pig (Large White)	Fetal lethality Derks et al. (2018)	*BBS9, BMPER*	Frameshift (deletion)	Autosomal recessive	Increased feed intake and growth	Fetal death
Pig (Large White)	Leg weakness Matika et al. (2019)	*MSTN*	Stop-gained	Autosomal recessive	Increase in muscle depth and decrease in fat depth	Leg weakness syndrome
Pig (Finnish Yorkshire)	Immotile short-tail sperm Sironen et al. (2012)	*SPEF2*	Frameshift (insertion)	Autosomal recessive	High female litter size	Male infertility
Pig (Pietrain, Landrace)	Malignant hyperthermia Vogeli et al. (1994)	*RYR1*	Missense	Autosomal recessive	High lean meat content	Pale soft exudative meat
Cattle (Belgian Blue)	Crooked tail Fasquelle et al. (2009)	*MRC2*	Frameshift and missense	Autosomal recessive	Enhanced muscular development	Crooked tail syndrome
Cattle (Nordic Red)	Embryonic lethality Kadri et al. (2014)	*RNASEH2B*	Frameshift (deletion)	Autosomal recessive	High milk yield	Embryonic lethal
Cattle (Belgian Blue and Shorthorn)	Roan coat Seitz et al. (1999)	*KITLG*	Missense	Autosomal Co-dominant	Roan phenotype	White heifer disease
Sheep	Mastitis Rupp et al. (2015)	*SOCS2*	Missense	Autosomal recessive	Higher body weight and milk production (also in homozygotes)	Mastitis
Sheep	Fecundity Hanrahan et al. (2004)	*BMP15, GDF9*	Stop-gained and missense	BMP15: X-linked GDF9: Autosomal recessive	Increase ovulation rate and litter size	Female infertility
Sheep (Soay)	Polledness Wiedemar and Drögemüller (2015) Ho^{+}/Ho^{+} males: lower survival	*RXFP2*	Frameshift (insertion)	Autosomal recessive	Higher overall fitness	Ho^{P}/Ho^{P} males: lower reproductive success
Sheep	Chondrodysplasia Beever et al. (2006)	*FGFR3*	Missense	Autosomal recessive	Larger animals	Spider lamb syndrome
Sheep	Callipygous phenotype Kim et al. (2004)	*DLK1 - PEG11*	Intergenic	Polar overdominance	Muscular hypertrophy	-
Chicken (Wyandotte)	Rose comb Imsland et al. (2012)	*MNR2, CCDC108*	Inversion	Autosomal dominant	Rose comb	Male infertility
Chicken	Creeper Jin et al. (2016)	*IHH*	Deletion	autosomal dominant semi-lethal	Short legs	Lethal before hatch
Horse (Appaloosa, Knabstrupper)	Leopard complex spotting Bellone et al. (2010)	*TRPM1*	Frameshift (insertion)	Autosomal incompletely dominant	Leopard complex spotting	Congenital night blindness
Horse (American Paint)	Frame pattern Ayala-Valdovinos et al (2016)	*EDNRB*	Missense	Autosomal dominant	Frame overo	Lethal white foal syndrome
Rabbit	Dwarfism Carneiro et al. (2017)	*HMGA2*	Deletion	Autosomal recessive	Dwarfism	Lethal
Rabbit	English Spotting Fontanesi et al. (2014)	*KIT*	Regulatory	Autosomal recessive	English spotting	Dilated ("mega") cecum and ascending colon

The table also includes one example of polar-overdominance, of which the phenotype depends on the parental origin of the variant.

the glutamate with a stop codon (p.Glu274*). This results in a loss of function of the *MSTN* gene, that causes hypertrophy and leading to the so called 'double muscling' phenotype (Matika et al., 2019) (**Figure 1**).

Older examples in pig include a disruptive intronic 9-kb insertion variant in the *SPEF2* gene causing immotile short-tail sperm (ISTS) which leads to male infertility, but increased litter size in females (Sironen et al., 2012). A second example is a missense variant in the RYR1 gene (p.Arg615Cys) inducing malignant hyperthermia (MH) in homozygotes, but higher lean meat content in heterozygotes (Salmi et al., 2010).

Cow (*Bos taurus*)

Belgian Blue cattle are popular for their high muscularity. A variant in the *MRC2* gene causes enhanced muscular development in heterozygotes, but negative effects on fitness in homozygotes (Fasquelle et al., 2009). Affected individuals have the so called 'crooked tail syndrome'. All cases show, next to a deviation of the tail, growth retardation, abnormal shape of the skull and extreme muscular hypertrophy. Due to the heterozygote advantage, the variant was positively selected resulting in a carrier frequency of 25% in Belgian Blue cattle (Fasquelle et al., 2009). The crooked tail syndrome in cattle is caused by a c.1906T>C transition within exon 3 of the *MRC2* gene. This variant causes an amino acid substitution (p.Cys636Arg) within the third C type lectin-like domain (CTLD3). This leads to oligomerization of the Endo180 protein, subsequently leading to a loss of function. (Sartelet et al., 2012). A second variant caused by a 2-bp deletion (c.2904-2905delAG) results in a frameshift that causes a premature stop codon in the *MRC2* gene (p.Gly934*). The deletion in the *MRC2* gene causes a truncated Endo180 protein that results in the loss of function of *MRC2* (Fasquelle et al., 2009).

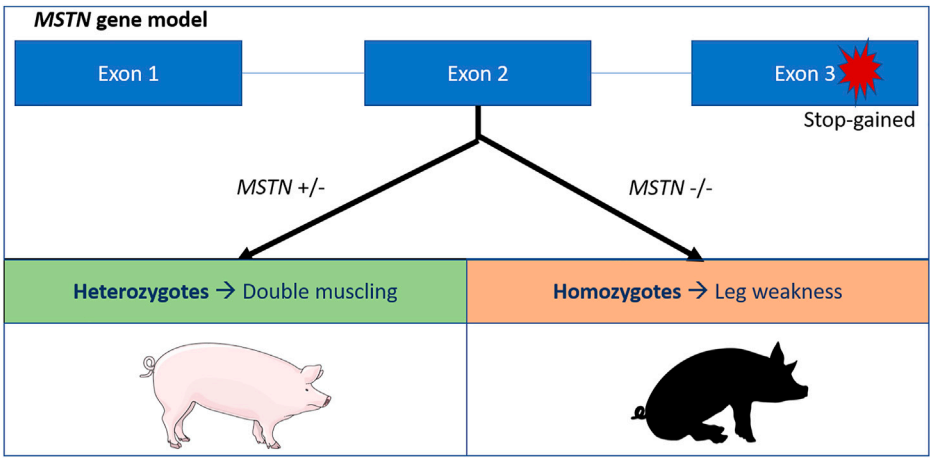

FIGURE 1 | Schematic overview of the *MSTN* stop gained variant in pigs (Matika et al., 2019). The stop codon affects the third exon of the *MSTN* gene, leading to a premature stop codon and a loss of function of the myostatin protein. Heterozygotes exhibit the double muscling phenotypes, while homozygotes suffer from leg weakness.

Next to that, Kadri et al. (2014) found a deletion comprising the *RNASEH2B* gene that results in embryonic lethality in mutant homozygous Nordic Red cattle. In carriers, positive effects on milk yield and composition were discovered. This resulted in the high carrier frequency of 13, 23 and 32% in Danish, Swedish and Finnish Red cattle, respectively. This high carrier frequency contributed to the decreased fertility in these breeds over the last years (Kadri et al., 2014). Embryonic lethality in cattle is caused by a 660-kb deletion encompassing 4 genes, including the ribonuclease H2 subunit B (*RNASEH2B*) gene. This deletion results in loss of the whole *RNASEH2B* gene, causing embryonic lethality. The deletion shows positive effects on milk yield in heterozygous cattle, but the molecular mechanism behind this advantage remains unknown (Kadri et al., 2014).

In addition, in Belgian Blue and Shorthorn cattle, a missense variant in the *KITLG* gene, resulted in females lacking the Müllerian ducts, causing sterility. The disease is also known as the White Heifer disease, as 90% of the cases show a completely white phenotype (**Figure 3D**) (Reissmann and Ludwig, 2013). Heterozygotes have the so called 'roan phenotype', an intermingled color with some white spotting, resulting in the blue and red phenotype in Belgian Blue and Shorthorn cattle, respectively. After elimination of 'White Heifer animals' and progeny testing of sires, the harmful allele frequency was reduced (Charlier et al., 1996). Roan coat in cattle is caused by an amino acid substitution of alanine to asparagine at amino acid 193 (p.Ala193Asp) (Seitz et al., 1999).

Sheep (*Ovis aries*)

In different sheep breeds, variants are found in the *BMP15* and *GDF9* genes underlying the "Fecundity" phenotype that cause increased ovulation rate and litter size in female heterozygotes but impaired oocyte development and maturation in homozygous females, resulting in female infertility (Javanmard et al., 2011). Due to the heterozygote advantage, the carrier frequency for variants in the *GDF9* and *BPM15* gene reached high frequencies in many sheep breeds (Demars et al., 2013). Fecundity in sheep is caused by different types of variants (mostly missense) in the growth differentiation factor 9 (*GDF9*) gene and/or *BMP15* gene. Both genes belong to the transforming growth factor beta (TGFβ) superfamily and the *GDF9* is an autosomal gene, while *BMP15* is X-linked. Five missense variants have been described in *GDF9*. In *BMP15*, however, six missense, two stop-gained, a small deletion variant, and a complex rearrangement have all been described (Nicholas, 2021).

A second unique form of pleiotropy is the 'polledness' trait found in Soay sheep, as both the recessive and dominant allele show negative effects on fitness in homozygous males. Homozygous males for the Ho+ allele, which confers larger horns, have lower survival while homozygous males for HoP allele, which confers smaller horns, have lower reproductive success (Johnston et al., 2013). In contrast, heterozygous males show a higher overall fitness due to increased survival and reproductive success. However, in females, both alleles do not have any effect on fitness. The allele frequency of the mutant HoP allele stabilized close to the equilibrium frequency, because of opposite selection on reproduction and survival (Johnston et al., 2013). The 'polledness' trait in sheep is caused by an 1833-bp insertion in the 3′UTR region of the relaxin/insulin-like family peptide receptor 2 (*RXFP2*) gene. This insertion adds an RNA antisense sequence of *EEF1A1* to the 3′end of *RFXP2* transcripts. Likely, *EEF1A1* transcripts bind to the 3′ end of the *RXFP2* mRNA, resulting in double stranded mRNA that will be degraded, reducing levels of the *RXFP2* protein product (Wiedemar and Drögemüller, 2015) (**Figure 2**).

Furthermore, a relatively old autosomal recessive disease, described by Cockett et al. (1999) causes hereditary chondrodysplasia in the Suffolk breed. Affected lambs exhibit the Spider Lamb Syndrome (SLS) (Cockett et al., 1999), a severe syndrome affecting growth of cartilage and bone (**Figure 3B**). In contrast, heterozygotes for the mutant allele are larger-framed

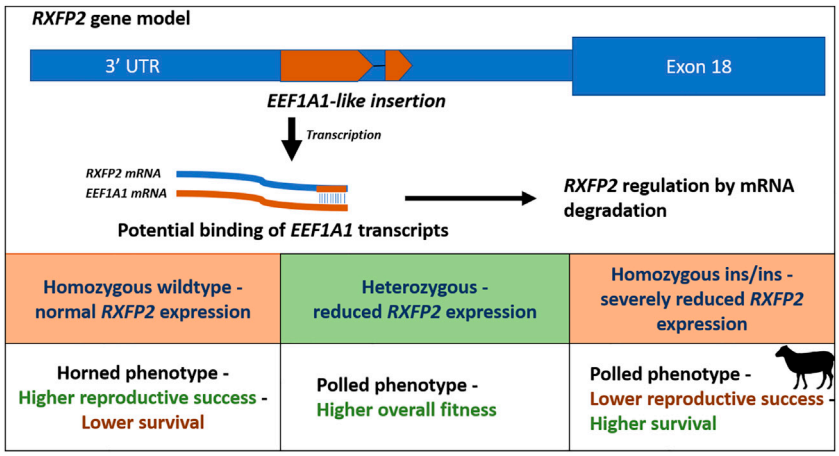

FIGURE 2 | Schematic overview of the polledness trait in Soay sheep (Johnston et al., 2013; Wiedemar and Drögemüller, 2015). The polledness phenotype is associated with a *EEF1A1*-like insertion in the 3′UTR of the *RXFP2* gene. The insertion potentially leads to *RXFP2* post transcriptional regulation by binding *EEF1A1* transcripts (caused by double stranded RNA degradation). Homozygous wildtype sheep have horns and higher reproductive success but lower survival, whereas homozygous ins/ins sheep exhibit lower reproductive success but higher survival. Heterozygotes exhibit the highest overall fitness underlying balancing selection.

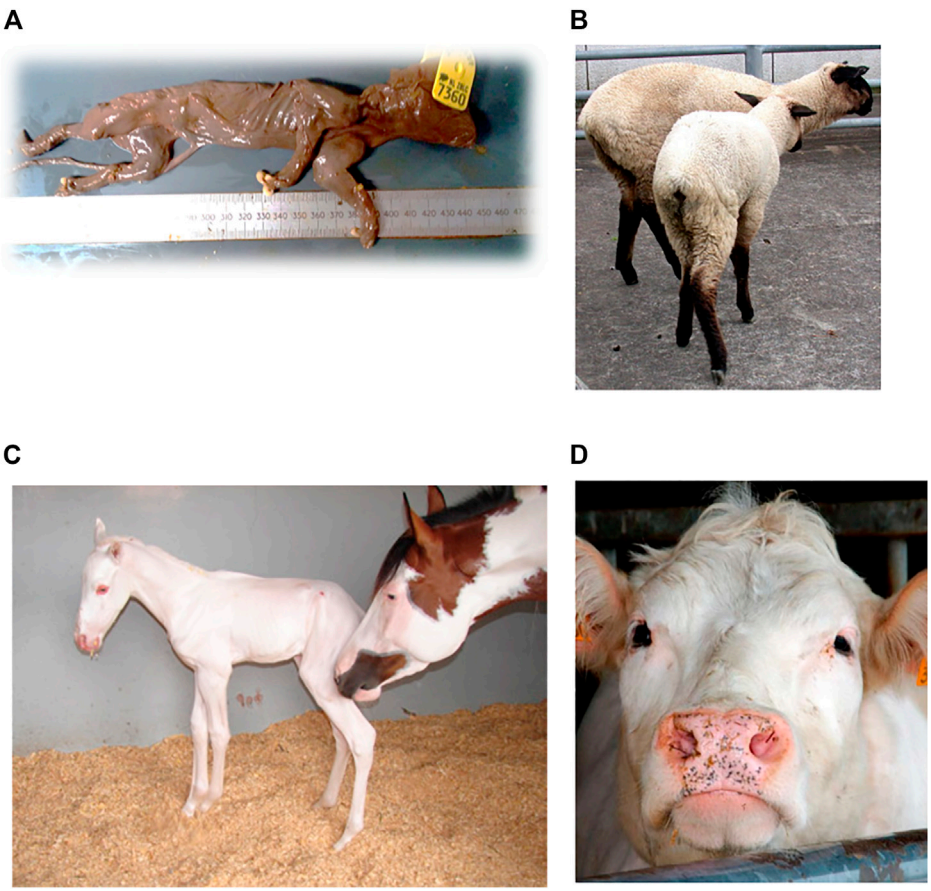

FIGURE 3 | Examples of **(A)** Mummified piglet resulting from a large deletion in pigs, figure from (Derks et al., 2018). **(B)** Ovine hereditary chondrodysplasia, also known as the Spider Lamb Syndrome, figure from (Thompson et al., 2008) **(C)** Lethal white foal syndrome, figure from (Ayala-Valdovinos et al., 2016) **(D)** White heifer disease, figure from (de Meuter, 2010).

sheep, with increased bone length and meat yield, possibly explaining the high frequency in this breed (Smith et al., 2006). Chondrodysplasia in sheep is caused by a p.Val700Glu substitution affecting the highly conserved tyrosine kinase domain ll, resulting in the loss of function of the *FGFR3* gene and thereby producing skeletal overgrowth (Beever et al., 2006).

Lastly, a single missense variant in the *SOCS2* gene (p.R96C) was found with positive effects on growth and milk production in sheep (Rupp et al., 2015). However, homozygous animals are also more susceptible to develop mastitis compared to wildtype and heterozygous animals. More specifically, heterozygous animals exhibit similar milk yield and fat content compared homozygotes, while growth is significantly higher in homozygotes compared to heterozygotes.

Chicken (*Gallus gallus*)

Strong artificial selection for the rose comb phenotype, affecting the *MNR2* gene, severely affects fertility in homozygous males (poor sperm motility). This results in male infertility when the sperm competes with that from heterozygous males (which is problematic if pools of semen are used) (Imsland et al., 2012). Rose comb in chicken is caused by a 7.38-Mb inversion with breakpoints located approximately at 16.50 Mb and 23.88 Mb within the homeodomain protein (*MNR2*) gene on chicken chromosome 7. This inversion causes a delocalization of the *MNR2* gene, leading to a misexpression of the gene during comb development. The rearrangement disrupts the gene *CDCC108* (located at breakpoint 23.88), responsible for male infertility in homozygous animals (Imsland et al., 2012).

Secondly, the creeper phenotype, caused by a deletion of the Indian hedgehog (*IHH*) gene, resulting in pronounced shortness of the extremities in heterozygotes was observed in chicken. Homozygous mutants generally die on the fourth day of embryonic development (Jin et al., 2016). The short legged phenotype was a popular characteristic in various breeds, explaining the high frequency of this allele in a wide range of breeds (Jin et al., 2016).

Horse (*Equus caballus*)

In the horse, various coat color phenotypes are associated with negative effects on fitness. In the Appaloosa and Knabstrupper, heterozygous animals for a variant in the *TRPM1* gene show a leopard complex spotting phenotype, a desired phenotype for breeders. However, the mutant homozygotes exhibit congenital stationary night blindness (Bellone et al., 2013). Leopard complex spotting in horses is caused by a 1378-bp insertion in intron 1 of the *TRPM1* gene. The insertion is a long terminal repeat (LTR) of an endogenous retrovirus (ERV) disrupting the expression of *TRMP1* by premature poly-adenylation preventing translation of the last amino acids (Bellone et al., 2013).

The "frame overo pattern" is another type of coat color in which horses exhibit white spots over their body. The frame overo pattern allele has a dominant mode of inheritance, but leads to the lethal white foal syndrome in homozygotes (**Figure 3C**). Affected foals have a near or completely white coat and lack enteric ganglion cells; intestinal agangliosis (Bellone, 2010), usually resulting in death within the first 24 h of life. The harmful allele is under balancing selection, as carriers have a phenotypic advantage, leading to a frequency up to 21.6% in the American Paint breed (Badial et al., 2018). The frame pattern in horses is caused by a di-nucleotide substitution c.353TC>AG within the first exon in the endothelin receptor type B (*EDNRB*) gene. This dinucleotide change causes an amino-acid substitution of isoleucine to lysine at amino acid 118 (p.Ile118Lys). This amino acid substitution to a charged amino acid is located within the first transmembrane domain of the 7-transmembrane domain G protein coupled receptor but how it affects function remains to be determined (Bellone, 2010).

Rabbit (*Oryctolagus cuniculus*)

In rabbits the *KIT* gene was found to be responsible for the English spotting coat colour in heterozygous *En/en* individuals. However, homozygous *En/En* animals exhibit a dilated ("mega") cecum and ascending colon leading to problems with the digestive system especially during flare-ups of the disease (Fontanesi et al., 2014). Another example in rabbits includes a 12.1 kb deletion affecting the *HMGA2* gene leading to dwarfism in heterozygotes, but leading to a lethal phenotype in homozygotes (Carneiro et al., 2017).

Major Histocompatibility Complex

A gene family that is under balancing selection in livestock is the major histocompatibility complex (MHC) class genes. Heterozygotes usually exhibit increased disease resistance, while homozygosity in the MHC region will likely cause increased susceptibility to disease due to a lower diversity at the peptide-binding region, resulting in lower effectiveness in recognizing pathogens (Hedrick, 1998). Hence, genetic diversity in the MHC class gene family increases fitness, but scientists are only beginning to understand the molecular mechanisms driving this fitness advantage (Hammer et al., 2020).

COMMON TRAIT SELECTION AFFECTS SIMILAR GENES ACROSS LIVESTOCK SPECIES

Because of similar selection goals in livestock, genes can influence the same traits in different livestock species, leading to comparable phenotypes. However, variants affecting the same gene in different livestock species do not necessarily show antagonistic pleiotropic effects in either species.

Double Muscling

One key example is the 'double muscling' phenotype caused by variants in the *MSTN* gene described in pig, cattle, poultry, and sheep (Nicholas, 2021). The double-muscle trait is characterized by an increase in muscle mass, resulting in significantly higher meat yield (Aiello et al., 2018). The double muscling phenotype provides obvious advantages for animal breeders, but generally coincides with major drawbacks like the larger incidence of calving difficulties in cattle, and leg-weakness in pigs.

Coat Color

Genes affecting coat color give rise to comparable phenotypes in livestock. For example, mutations in the *EDNRB* gene, causing lethal white foals in horses (Bellone, 2010), have also been reported in sheep and chicken (Nicholas, 2021). More specifically, a 110-kb deletion in the *EDNRB* gene causes hypopigmentation and a megacolon, known as Waardenburg syndrome type 4A in sheep. Sheep homozygous for this deletion have a white coat and blue eyes and die shortly after birth, because of intestinal obstruction similar to the aganglionosis present in lethal white foals (Lühken et al., 2012). The Waardenburg syndrome is very similar to the lethal white foals in horses, as both homozygous animals express a white coat, intestinal obstruction or aganglionosis, and die shortly after birth. The *KIT* ligand is a second gene affecting coat color in a wide range of breeds and species (Reissmann and Ludwig, 2013). In addition to the roan coat color in cattle, also in horses an analogous phenotype associated with the *KIT* gene has been reported (Marklund et al., 1999).

Polledness

In sheep, homozygote males for both HoP and Ho+ have lower fitness and heterozygote males have an overall higher fitness (Wiedemar and Drögemüller, 2015). In goats, interestingly, the polled intersex syndrome (PIS) is described. This syndrome is a disorder of sexual development showing an association with the polled phenotype and intersexuality. PIS is caused by a 10-kb deletion and a 480-kb insertion containing the genes *ERG* and *KCNJ15* leading to a complex rearrangement (Simon et al., 2020). In cattle, polledness is caused by complex duplication variants on BTA1 that leads to differences in gene expression (Wiedemar and Drögemüller, 2015).

Other Examples

In sheep, variants in the *FGFR3* gene lead to the spider lamb syndrome, and larger animals in heterozygotes (Beever et al., 2006). In cattle, a stop-lost variant that affects the cattle *FGFR3* gene has been reported causing a dominant form of chondrodysplasia. The variant is a *de novo* mosaic variant as the parents of the calves with chondrodysplasia were not affected (Häfliger et al., 2020).

The gene responsible for malignant hyperthermia in pigs has also been reported in both cattle (Hill et al., 2000) and horse breeds (Aleman et al., 2004). In horses, more specifically, a harmful missense variant in the American Quarter horse leads to a phenotype analogous to the hyperthermia in pigs.

POLAR OVERDOMINANCE

Polar overdominance is a unique and rare type of overdominance. This phenomenon refers to heterozygous animals only expressing the phenotype depending on the parental origin of the variant.

Callipygous Phenotype

The first discovered and most well-known type of polar overdominance is the callipyge variant in sheep. Heterozygous sheep which inherited the variant from their sire express the phenotype. The callipygous (CLPG) phenotype is characterized by muscular hypertrophy ('double muscling') of the hind legs. Lambs which inherited the callipygous variant from their sire show a normal phenotype at birth, but develop the callipygous phenotype from 30 till 80 days after birth (Cockett et al., 1996). The callipygous phenotype is caused by an A to G transition in the intergenic region between the *DLK1* and *GTL2* genes. This results in an increased expression of the *DLK1* and *PEG11* genes on the paternal haplotype. Hence, the variant affects *cis*-acting regulatory sequences resulting in a gain of function of *DLK1* (Freking et al., 2002).

Furthermore, the callipygous phenotype also occurs in goats, but in this breed only individuals which receive the mutated allele from their dam express the phenotype. The callipygous phenotype in goats is caused by a C to T transition (Li et al., 2009). Next to that, Kim et al. (2004) found that the polar overdominance is present in the porcine *DLK1-GTL2* region that is homologous to that of the sheep CLPG locus. Pigs who inherited the paternal *DLK1* allele 2 and the maternal *DLK1* allele 1 exhibit higher lean muscle mass and decreased fat deposition. In contrast, pigs who inherited the maternal *DLK1* allele 2 and the paternal *DLK1* allele 1 exhibit lower pre- and postnatal growth (Kim et al., 2004).

POPULATION GENETICS AND MANAGEMENT OF ALLELES UNDER BALANCING SELECTION

The Balance Between Drift and Selection

Due to the overdominance effect, the frequency of the discussed alleles can become unexpectedly high in the population. Examples have been published describing lethal alleles at a carrier frequency >10% just by drift in various livestock species (Charlier et al., 2016; Efendić et al., 2018; Derks et al., 2019a). Hence, an unusual high frequency of deleterious variants does not constitute an advantage in heterozygotes per se. Therefore, not only pleiotropic effects can lead to a moderate frequency of deleterious alleles. The frequency can also be influenced by genetic drift and the power of genetic drift heavily depends on the effective population size (Derks et al., 2019a). Leroy et al. (2013) studied the effective population size in several dog, sheep, horse and cattle breeds and found that the effective population size is small (Ne < 100) in most commercial livestock breeds. As a consequence (mildly) deleterious alleles can reach a relatively high frequency, or even become fixed in a population due to genetic drift (Leroy et al., 2013). Charlier et al., 2016. specifically studied the number and frequency of lethal alleles (the most severe type of deleterious allele) (Charlier et al., 2016). They showed that the number and frequency of these type of variants depends on several population genetic parameters including the genomic target size for lethal mutations, the rate of recessive lethal mutations in this target space, and especially the present and past effective population size. They concluded that the number of recessive lethals carried on average per individual increases with effective population size from about one for an Ne

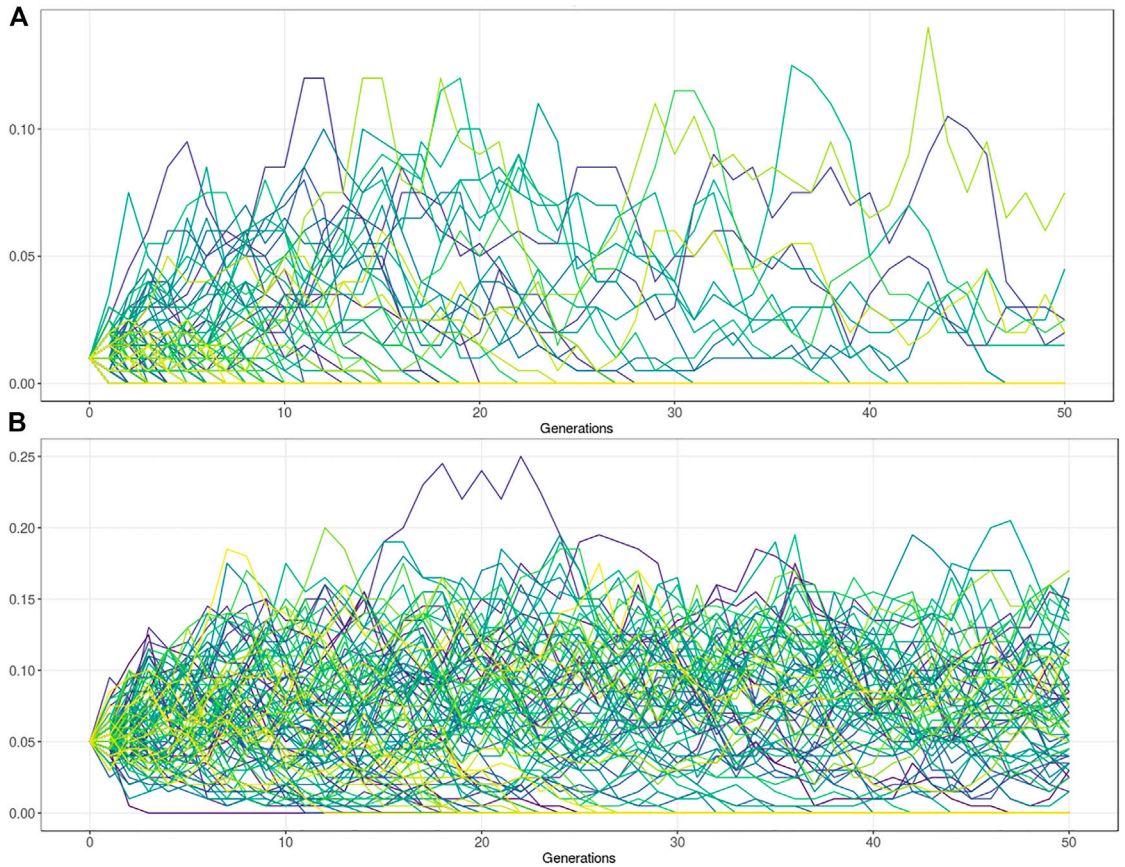

FIGURE 4 | Population genetic simulations for deleterious alleles with Ne = 100. **(A)** Frequency simulation of a deleterious allele at starting frequency 1%. The fitness of the homozygous BB genotype = 0, with equal fitness for the AB and AA genotype (fitness = 1). Figure shows that the allele is lost in the vast majority of the populations after 50 generations. The maximum frequency reached by drift is approximately 10%. **(B)** Frequency simulation of a deleterious allele with the fitness of the homozygous BB genotype = 0, and the AB genotype has a 10% fitness advantage over the homozygous AA genotype (AB = 1, AA = 0.9). Figure shows that the allele is lost in about one third of the populations after 50 generations. In two third of the populations the allele remains at a relatively stable equilibrium between 5–15% allele frequency.

of 100 to ~7.7 for Ne of 10.000. However, the frequency of deaths as a result of these lethal variants is nearly similar (between 1.54 and 1.73%). Together the results show that a small effective population size leads to high extinction rates of harmful alleles, but the few alleles that did not went extinct tend to rapidly spread in the population. Hence, the number of deleterious alleles that segregate in a population with low Ne is much smaller compared to high Ne populations. However, the deleterious variants that do segregate tend to segregate at higher frequency. Nevertheless, at a certain frequency, there will be a trade-off (the maximum allele frequency that can be reached by drift alone) between purifying selection and genetic drift. At this frequency, the homozygote loss will be larger than the heterozygote advantage.

The variants under balancing selection are subject to a complex interplay between the advantage in heterozygotes and the disadvantage in homozygotes. Artificial selection has had major consequences for these types of variants. By using influential sires in the population, novel variation can rise in frequency relatively rapidly. Strong selection and the use of influential sires (the influential sire effect) can cause undetected, undesirable deleterious alleles in the population to spread rapidly (Leroy and Baumung, 2011). It generally also reduces genetic diversity by the exclusion of other males. The influential sire effect reduces the effective population size which subsequently affects the power of drift effects in the population. In small Ne populations most deleterious variants that arise tend to disappear relatively quickly. However, a small subset of variant can spread rapidly throughout the population. To perform such population simulation studies several tools have been published (Peng et al., 2015). The tools are useful to simulate the expected frequency of deleterious alleles given several population genetic parameters. Users can assess the impact of heterozygous advantage on the allele frequency, and the possible trait-off value at which the deleterious phenotype in homozygotes prevents further frequency increase. The tool takes a starting frequency and relative fitness of three genotype classes (AA, AB, BB). Further population parameters include the mutation rate, migration rate, effective population size, and the number of generations. Given an Ne of 100, with no advantage in heterozygotes, the maximum frequency of a lethal allele (Fitness BB = 0) is about 10%, but in the vast majority of the

simulations the allele is lost after 50 generations (**Figure 4A**). Note, that a fitness of zero means that the animals cannot reproduce but could still result in a viable phenotype. However, if the heterozygotes (AB) have a 10% fitness advantage over homozygous wildtype (AA), the frequency stays at a relatively stable (10–15% AF) equilibrium (**Figure 4B**). Together this equilibrium frequency is affected by an interplay of the fitness advantage in heterozygotes, disadvantage in homozygous mutants, and the selection goal.

Purging Strategies

After discovery, the harmful allele can be eliminated from the population, if desired. However, the rate at which a deleterious allele can be removed from a population, partly depends on the stage at which the negative consequence is expressed. For example, defects that results in distinct birth defects will be noticed sooner compared to defects that lead to early lethality. The strategy of elimination depends on the heterozygote advantage, the homozygote disadvantage and the allele frequency in the population. One efficient way to purge the variant is by eliminating all carriers and affected individuals from the population. However, depending on the allele frequency, a large part of the population will be eliminated, increasing genetic drift, and lowering the genetic diversity. Consequently, the risk of inbreeding will increase and thereby the risk of developing new deleterious variants. Alternatively, gene-assisted selection can be used to prevent matings between parents carrying the same deleterious variant (Georges et al., 2019). This method reduced the carrier frequency of ISTS in Finnish Yorkshire pigs by 30%. Next to that, gene-assisted selection has been applied to Pietrain and Landrace breeds, resulting in elimination of the variant causing PSS in these breeds (Hedrick, 2015). Sometimes, the frequency of the harmful allele will be kept at a relatively stable level in the population (as for the rose comb trait in chickens). Additionally, the harmful allele causing polledness in sheep stabilized close to the equilibrium frequency, as both the wild-type and mutant homozygote state show negative fitness effects (Johnston et al., 2013). Nevertheless, we need to keep in mind that sometimes the carrier frequency is already greatly reduced before the causative variant was discovered. For example, PSS pigs were halothane sensitive and through halothane tests these pigs were excluded from the population before DNA tests were available (Vogeli et al., 1994). Moreover, breeding objectives can differ between countries and change over time, thereby influencing the allele frequency. For example, the mutant *PLAG1* gene increases stature but has a negative effect on the fertility in cattle (Utsunomiya et al., 2017). This allele increased in frequency after a period of selection on smaller cattle, while larger cows are now preferred in some breeds and countries. In addition, the frequency of the *DGAT1* variant, which decreases protein yield and increases fat yield, shifted due to selection goal differences over time (Georges et al., 2019). For the mutant alleles that only have a sex-limited negative effect, a mating strategy can be developed to increase the number of heterozygotes with positive phenotypic effects and decrease the number of lethal homozygotes. For example, the fecundity trait in sheep is lethal in homozygous females, and therefore mutant homozygous males (aa) could be crossed with homozygous females (AA).

DISCUSSION

In this review, we describe various harmful alleles under balanced selection exhibiting positive selection on heterozygotes and purifying selection on homozygotes. Despite the vast increase of genomic data, the exact molecular mechanisms underlying the alleles under balancing selection remain largely unknown.

The described examples likely only present the tip of the iceberg and many alleles under balanced selection remain hidden in the population due to the low frequency or by the lack of genomic resources. Next to that, rare birth defects in livestock species are often noted as 'weak animal' and no further investigation to possible underlying genetic syndromes is done (Derks et al., 2019b). However, only a small subset of birth defects will be caused by recessive deleterious variants, of which an even smaller subset could constitute an advantage in heterozygotes. In addition, variants with only limited advantage in heterozygotes, but severe disadvantages in homozygotes will likely not reach moderate to high frequency in the population. Additionally, the lack of genomic tools especially for natural and small-sizes domestic populations hamper the detection of alleles under balancing selection. Nevertheless, some examples of harmful alleles under balancing selection have been described in wild populations. For example, the black coat phenotype in the Yellowstone wolf population (affecting the *CBD103* gene) leads to a higher fitness compared to homozygotes that exhibit lower recruitment and survival (Coulson et al., 2011; Hedrick, 2012). Secondly, the leopard complex spotting in horses, described in this review, was also reported in ancient wild horse populations potentially providing camouflage for predators in the snow (Pruvost et al., 2011). Alternatively, an example of a balanced lethal system is found in crested newts, where heterozygotes are viable and both homozygotes (wild-type and mutant) are lethal (Grossen et al., 2012). Though, in most cases, high frequencies of harmful alleles in small threatened wild populations are likely driven by genetic drift (Trask et al., 2016).

Heterozygous Disadvantage

Interestingly, also variants that cause a heterozygote disadvantage have been described, like bovine tuberculosis. Cattle heterozygous for the locus on *BTA6* show increased susceptibility for bovine tuberculosis in comparison with homozygotes (Tsairidou et al., 2018). However, especially for most genes related to the immune system (e.g. the major histocompatibility complex), heterozygosity and high genetic variability is often beneficial for proper defense against various pathogens (Huchard et al., 2010). For example, heterozygous leopard frogs have a higher chance of survival when infected with a fungal pathogen compared to homozygotes (Savage and Zamudio, 2011).

Allelic Pleiotropy

Besides heterozygote advantage, variants can be under balancing selection because they affect multiple traits (allelic pleiotropy) that

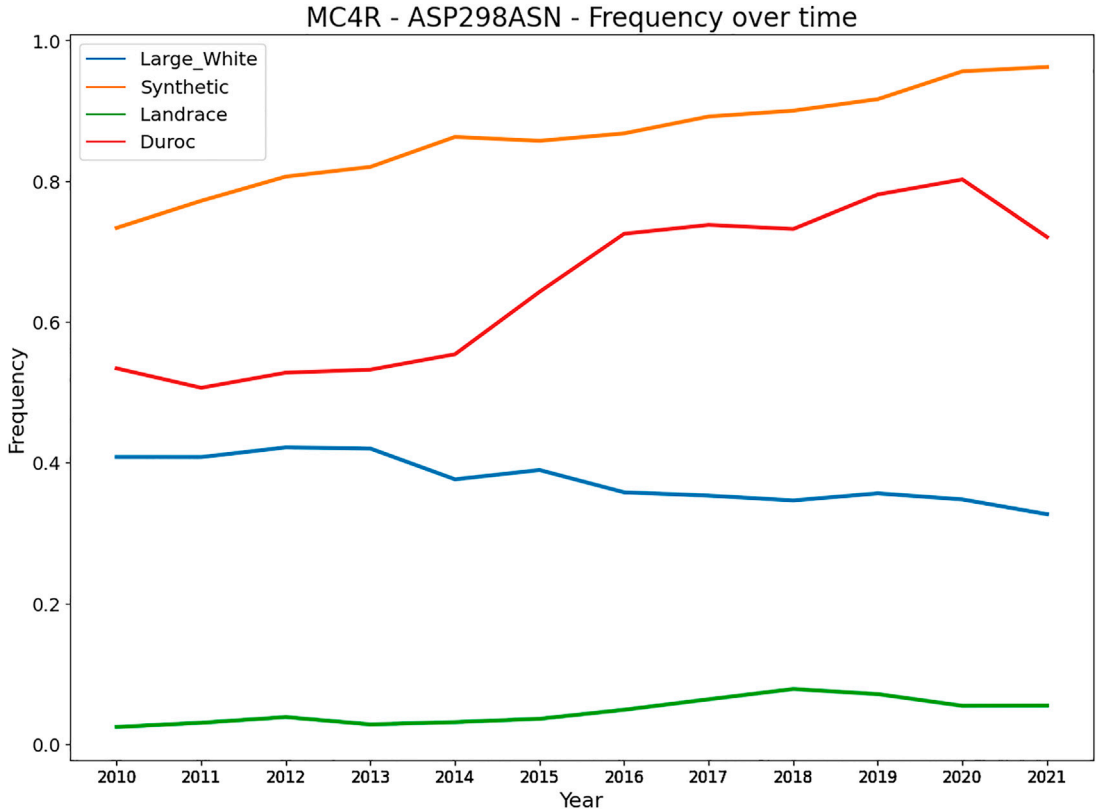

FIGURE 5 | Frequency of the MC4R - Asn298 allele in four commercial pig breeds from Topigs Norsvin. The Asn298 allele is associated with more fat, higher feed consumption and faster growth, while the Asp298 allele is associated with less backfat, slower growth, and lower feed intake (Kim et al., 2000).

are (negatively) correlated. Variants with a large effect on a trait likely also affect other traits, caused by (negative) trait correlations affecting similar pathways and genes that regulate multiple traits. A key example is a variant in the *MC4R* in pigs affecting growth rate, fat composition, and feed intake in different pig breeds (Kim et al., 2000). The Asp298 allele is associated with less backfat, slower growth, and lower feed intake, while the Asn298 allele is associated with more fat, higher-feed consumption, and faster growth. This variant can remain at a rather stable equilibrium in the population depending on the selection intensity of the associated traits. More specifically, we evaluated the frequency of the *MC4R* missense variant over the last decade in four commercial pig populations (**Figure 5**). The Large White and Landrace breed are sow breeds, with more emphasis on reproduction traits and mothering abilities, whereas the Duroc and Synthetic are boar breeds with more emphasis on growth. **Figure 4** shows the frequency of the Asn298 allele (associated with more fat, higher feed consumption and faster growth). Notably, the frequency of this allele is higher (and increasing) in both boar lines (with stronger selection on growth related traits). However, for the sow lines the frequency is much lower (especially in Landrace) and remains at a more stable equilibrium likely due to lower selection intensities on growth and other production traits. An example in cattle involves four clusters of SNPs close to the genes *GPAT4, MGST1, DGAT1 and PAEP*, of which each has an allele that lowers milk and protein yield and increases fat yield. (Xiang et al., 2017).

Most variants are first identified because they exhibit a deleterious effect on a specific trait. Once identified, its consequences on another traits can be tested and quantified. However, this antagonistic effect does not immediately proof the pleiotropic effect, despite the possibility that one gene affects several traits by affecting a single or multiple pathways. Testing the effect of the variant in an different genetic context would be required to further support the pleiotropic effect.

In this review, the described variants are pleiotropic and the positive selection in heterozygotes is caused by the direct effect of the variant itself. However, a similar effect can be caused by two variants that are closely linked on the same haplotype. For example, two different variants could reside on the same haplotype, one with beneficial effects and the other one with a deleterious effect. Examples of such variants are very rare and most described variants under balancing selection exhibit antagonistic pleotropic effects. The efficiency to separate the two variants by recombination depends on the local LD and haplotype structure in the population. In livestock, LD patterns can differ greatly between different loci in the genome which partly depends on the recombination frequency (Qanbari, 2019).

Different Types of Causative Variants and the Coding Sequence Bias

We review a wide variety of causative variants underlying heterozygote advantage phenotypes. Some variants are due to

small changes (i.e., single nucleotide substitutions or small indels), while other variants include changes of structural origin. The majority of the described variants lead to a loss of function of a particular gene either by directly affecting the coding sequence of the gene, or due to changes in expression. Even if a single gene is affecting the same trait in multiple species (i.e., *MSTN* gene resulting in the double muscling phenotype) with different types of underlying variants, they overall lead to the same consequence; a loss of function of the myostatin protein.

Key examples of regulatory variants include the deletion in the *BBS9* gene in pigs and the *RXFP2* gene in sheep that reduce the expression of the downstream gene, while the inversion affecting *MNR2* in chickens leads to a misexpression of the causative gene (Hedrick, 2015; Wiedemar and Drögemüller, 2015; Derks et al., 2018). In contrast, the regulatory variant that causes the callipygous phenotype in sheep, results in increased expression of the *DLK1* gene leading to a gain of function (Freking et al., 2002). Most variants that are identified affect the coding sequence of the genes, this is partly because loss of function variants often affect the coding sequence of a gene. Nevertheless, we expect that regulatory variants are potentially equally common, but remain more challenging to identify.

Effects in Crossbreeding

The deleterious alleles described are mostly found in purebred breeds. However, the final production animals in pigs and chicken are crossbreds between purebred breeds. These crossbreds perform better on fertility and growth traits as a result of the heterosis effect. Therefore, deleterious alleles uniquely segregating in a purebred breed will never be homozygous in crossbreds. So, the final production animals only take advantage of the pleiotropic effect of the deleterious allele (Derks et al., 2019a). Nonetheless, crossbreds can become homozygous for the mutant allele when the variant is present in both the maternal and paternal purebred breeds. Despite the limited impact on crossbreds, given that most harmful alleles are population specific, purging of deleterious variants is desired within purebreds, because of the potential economic losses, and the effect on animal welfare.

In this review, we focused on variants that exhibit heterozygous advantage but have a deleterious effect in homozygotes. We believe that with current availability of genomic and phenotypic datasets a multitude of variants will likely be discovered in the near future, giving better insights into how common these types of variants are. We believe that livestock breeds provide a key framework to study these types of variants due to the high level of genomic resources, deviating selection intensities, and increase of molecular data.

AUTHOR CONTRIBUTIONS

MS drafted the initial manuscript. MD extensively drafted the manuscript further. Both authors approved the manuscript.

REFERENCES

Aiello, D., Patel, K., and Lasagna, E. (2018). Themyostatingene: an Overview of Mechanisms of Action and its Relevance to Livestock Animals. *Anim. Genet.* 49 (6), 505–519. doi:10.1111/age.12696

Aleman, M., Riehl, J., Aldridge, B. M., Lecouteur, R. A., Stott, J. L., and Pessah, I. N. (2004). Association of a Mutation in the Ryanodine Receptor 1 Gene with Equine Malignant Hyperthermia. *Muscle Nerve* 30 (3), 356–365. doi:10.1002/mus.20084

Ayala-Valdovinos, M. A., Galindo-García, J., Sánchez-Chiprés, D., and Duifhuis-Rivera, T. (2016). New Test for Endothelin Receptor Type B (EDNRB) Mutation Genotyping in Horses. *Mol. Cell Probes* 30 (3), 182–184. doi:10.1016/j.mcp.2016.03.005

Badial, P. R., Teixeira, R. B. C., Delfiol, D. J. Z., da Mota, L. S. L. S., and Borges, A. S. (2018). Validation of High-Resolution Melting Analysis as a Diagnostic Tool for Endothelin Receptor B Mutation in American Paint Horses and Allele Frequency Estimation. *Mol. Cell Probes* 41, 52–56. doi:10.1016/j.mcp.2018.08.002

Beever, J. E., Smit, M. A., Meyers, S. N., Hadfield, T. S., Bottema, C., Albretsen, J., et al. (2006). A Single-Base Change in the Tyrosine Kinase II Domain of Ovine FGFR3 Causes Hereditary Chondrodysplasia in Sheep. *Anim. Genet.* 37 (1), 66–71. doi:10.1111/j.1365-2052.2005.01398.x

Bellone, R. R., Forsyth, G., Leeb, T., Archer, S., Sigurdsson, S., Imsland, F., et al. (2010). Fine-mapping and Mutation Analysis of TRPM1: a Candidate Gene for Leopard Complex (LP) Spotting and Congenital Stationary Night Blindness in Horses. *Brief. Funct. Genomics* 9 (3), 193–207. doi:10.1093/bfgp/elq002

Bellone, R. R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S., et al. (2013). Evidence for a Retroviral Insertion in TRPM1 as the Cause of Congenital Stationary Night Blindness and Leopard Complex Spotting in the Horse. *PLoS One* 8 (10), e78280. doi:10.1371/journal.pone.0078280

Bellone, R. R. (2010). Pleiotropic Effects of Pigmentation Genes in Horses. *Anim. Genet.* 41 (Suppl. 2), 100–110. doi:10.1111/j.1365-2052.2010.02116.x

Bosse, M., Megens, H. J., Derks, M. F. L., Cara, Á. M. R., and Groenen, M. A. M. (2019). Deleterious Alleles in the Context of Domestication, Inbreeding, and Selection. *Evol. Appl.* 12 (1), 6–17. doi:10.1111/eva.12691

Carneiro, M., Hu, D., Archer, J., Feng, C., Afonso, S., Chen, C., et al. (2017). Dwarfism and Altered Craniofacial Development in Rabbits Is Caused by a 12.1 Kb Deletion at the HMGA2 Locus. *Genetics* 205 (2), 955–965. doi:10.1534/genetics.116.196667

Charlesworth, D. (2006). Balancing Selection and its Effects on Sequences in Nearby Genome Regions. *Plos Genet.* 2 (4), e64. doi:10.1371/journal.pgen.0020064

Charlesworth, D., and Willis, J. H. (2009). The Genetics of Inbreeding Depression. *Nat. Rev. Genet.* 10 (11), 783–796. doi:10.1038/nrg2664

Charlier, C., Denys, B., Belanche, J. I., Coppieters, W., Grobet, L., Mni, M., et al. (1996). Microsatellite Mapping of the Bovine Roan Locus: a Major Determinant of White Heifer Disease. *Mamm. Genome* 7 (2), 138–142. doi:10.1007/s003359900034

Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., et al. (2016). NGS-based Reverse Genetic Screen for Common Embryonic Lethal Mutations Compromising Fertility in Livestock. *Genome Res.* 26 (10), 1333–1341. doi:10.1101/gr.207076.116

Cockett, N. E., Jackson, S. P., Shay, T. L., Farnir, F., Berghmans, S., Snowder, G. D., et al. (1996). Polar Overdominance at the Ovine Callipyge Locus. *Science* 273 (5272), 236–238. doi:10.1126/science.273.5272.236

Cockett, N. E., Shay, T. L., Beever, J. E., Nielsen, D., Albretsen, J., Georges, M., et al. (1999). Localization of the Locus Causing Spider Lamb Syndrome to the Distal End of Ovine Chromosome 6. *Mamm. Genome* 10 (1), 35–38. doi:10.1007/s003359900938

Coulson, T., MacNulty, D. R., Stahler, D. R., vonHoldt, B., Wayne, R. K., and Smith, D. W. (2011). Modeling Effects of Environmental Change on Wolf Population Dynamics, Trait Evolution, and Life History. *Science* 334 (6060), 1275–1278. doi:10.1126/science.1209441

de Meuter, P. (2010). *Wittevaarzenziekte: Operatief Ingrijpen Geeft Een Reële Kans Op Een Normaal Afmesttraject*. Wageningen: VeeteeltVlees.

Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., et al. (2013). Genome-wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperprolificacy Phenotype in Sheep. *Plos Genet.* 9 (4), e1003482. doi:10.1371/journal.pgen.1003482

Derks, M. F. L., Gjuvsland, A. B., Bosse, M., Lopes, M. S., van Son, M., Harlizius, B., et al. (2019a). Loss of Function Mutations in Essential Genes Cause Embryonic Lethality in Pigs. *Plos Genet.* 15 (3), e1008055. doi:10.1371/journal.pgen.1008055

Derks, M. F. L., Harlizius, B., Lopes, M. S., Greijdanus-van der Putten, S. W. M., Dibbits, B., Laport, K., et al. (2019b). Detection of a Frameshift Deletion in the SPTBN4 Gene Leads to Prevention of Severe Myopathy and Postnatal Mortality in Pigs. *Front. Genet.* 10, 1226. doi:10.3389/fgene.2019.01226

Derks, M. F. L., Lopes, M. S., Bosse, M., Madsen, O., Dibbits, B., Harlizius, B., et al. (2018). Balancing Selection on a Recessive Lethal Deletion with Pleiotropic Effects on Two Neighboring Genes in the Porcine Genome. *Plos Genet.* 14 (9), e1007661. doi:10.1371/journal.pgen.1007661

Efendić, M., Maćešić, N., Samardžija, M., Vojta, A., Korabi, N., Capak, H., et al. (2018). Determination of Sublethal Mutation Causing Lavender Foal Syndrome in Arabian Horses from Croatia. *J. Equine Vet. Sci.* 61, 72–75. doi:10.1016/j.jevs.2017.11.014

Fasquelle, C., Sartelet, A., Li, W., Dive, M., Tamma, N., Michaux, C., et al. (2009). Balancing Selection of a Frame-Shift Mutation in the MRC2 Gene Accounts for the Outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *Plos Genet.* 5 (9), e1000666. doi:10.1371/journal.pgen.1000666

Fijarczyk, A., and Babik, W. (2015). Detecting Balancing Selection in Genomes: Limits and Prospects. *Mol. Ecol.* 24 (14), 3529–3545. doi:10.1111/mec.13226

Fontanesi, L., Vargiolu, M., Scotti, E., Latorre, R., Faussone Pellegrini, M. S., Mazzoni, M., et al. (2014). The KIT Gene Is Associated with the English Spotting Coat Color Locus and Congenital Megacolon in Checkered Giant Rabbits (*Oryctolagus cuniculus*). *Plos One* 9 (4), e93750. doi:10.1371/journal.pone.0093750

Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., et al. (2002). Identification of the Single Base Change Causing the Callipyge Muscle Hypertrophy Phenotype, the Only Known Example of Polar Overdominance in Mammals. *Genome Res.* 12 (10), 1496–1506. doi:10.1101/gr.571002

Georges, M., Charlier, C., and Hayes, B. (2019). Harnessing Genomic Information for Livestock Improvement. *Nat. Rev. Genet.* 20 (3), 135–156. doi:10.1038/s41576-018-0082-2

Grossen, C., Neuenschwander, S., and Perrin, N. (2012). The Balanced Lethal System of Crested Newts: a Ghost of Sex Chromosomes Past. *The Am. Naturalist* 180 (6), E174–E183. doi:10.1086/668076

Häfliger, I. M., Letko, A., Murgiano, L., and Drögemüller, C. (2020). De Novo stop-lost Germline Mutation in FGFR3 Causes Severe Chondrodysplasia in the Progeny of a Holstein Bull. *Anim. Genet.* 51 (3), 466–469. doi:10.1111/age.12934

Hammer, S. E., Ho, C.-S., Ando, A., Rogel-Gaillard, C., Charles, M., Tector, M., et al. (2020). Importance of the Major Histocompatibility Complex (Swine Leukocyte Antigen) in Swine Health and Biomedical Research. *Annu. Rev. Anim. Biosci.* 8, 171–198. doi:10.1146/annurev-animal-020518-115014

Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R., et al. (2004). Mutations in the Genes for Oocyte-Derived Growth Factors GDF9 and BMP15 Are Associated with Both Increased Ovulation Rate and Sterility in Cambridge and Belclare Sheep (Ovis aries)1. *Biol. Reprod.* 70 (4), 900–909. doi:10.1095/biolreprod.103.023093

Hedrick, P. W. (1999). Antagonistic Pleiotropy and Genetic Polymorphism: a Perspective. *Heredity* 82 (2), 126–133. doi:10.1038/sj.hdy.6884400

Hedrick, P. W. (1998). Balancing Selection and MHC. *Genetica* 104 (3), 207–214. doi:10.1023/a:1026494212540

Hedrick, P. W. (2015). Heterozygote Advantage: the Effect of Artificial Selection in Livestock and Pets. *J. Hered.* 106 (2), 141–154. doi:10.1093/jhered/esu070

Hedrick, P. W. (2012). What Is the Evidence for Heterozygote Advantage Selection. *Trends Ecol. Evol.* 27 (12), 698–704. doi:10.1016/j.tree.2012.08.012

Hill, B., McManus, A., Brown, N., Playford, C., and Noble, J. (2000). A Bovine Stress Syndrome Associated with Exercise-Induced Hyperthermia. *Aust. Vet J* 78 (1), 38–43. doi:10.1111/j.1751-0813.2000.tb10357.x

Huchard, E., Knapp, L. A., Wang, J., Raymond, M., and Cowlishaw, G. (2010). MHC, Mate Choice and Heterozygote Advantage in a Wild Social Primate. *Mol. Ecol.* 19 (12), 2545–2561. doi:10.1111/j.1365-294X.2010.04644.x

Imsland, F., Feng, C., Boije, H., Bed'hom, B., Bed'hom, V., Dorshorst, B., et al. (2012). The Rose-Comb Mutation in Chickens Constitutes a Structural Rearrangement Causing Both Altered Comb Morphology and Defective Sperm Motility. *Plos Genet.* 8 (6), e1002775. doi:10.1371/journal.pgen.1002775

Javanmard, A., Azadzadeh, N., and Esmailizadeh, A. K. (2011). Mutations in Bone Morphogenetic Protein 15 and Growth Differentiation Factor 9 Genes Are Associated with Increased Litter Size in Fat-Tailed Sheep Breeds. *Vet. Res. Commun.* 35 (3), 157–167. doi:10.1007/s11259-011-9467-9

Jin, S., Zhu, F., Wang, Y., Yi, G., Li, J., Lian, L., et al. (2016). Deletion of Indian Hedgehog Gene Causes Dominant Semi-lethal Creeper Trait in Chicken. *Sci. Rep.* 6, 30172. doi:10.1038/srep30172

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., et al. (2013). Life History Trade-Offs at a Single Locus Maintain Sexually Selected Genetic Variation. *Nature* 502 (7469), 93–95. doi:10.1038/nature12489

Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., et al. (2014). A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *Plos Genet.* 10 (1), e1004049. doi:10.1371/journal.pgen.1004049

Kim, K.-S., Kim, J.-J., Dekkers, J. M., and Rothschild, M. (2004). Polar Overdominant Inheritance of a DLK1 Polymorphism Is Associated with Growth and Fatness in Pigs. *Mamm. Genome* 15 (7), 552–559. doi:10.1007/s00335-004-2341-0

Kim, K. S., Larsen, N., Short, T., Plastow, G., and Rothschild, M. F. (2000). A Missense Variant of the Porcine Melanocortin-4 Receptor (MC4R) Gene Is Associated with Fatness, Growth, and Feed Intake Traits. *Mamm. Genome* 11 (2), 131–135. doi:10.1007/s003350010025

Leroy, G., and Baumung, R. (2011). Mating Practices and the Dissemination of Genetic Disorders in Domestic Animals, Based on the Example of Dog Breeding. *Anim. Genet.* 42 (1), 66–74. doi:10.1111/j.1365-2052.2010.02079.x

Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E., and Danchin-Burge, C. (2013). Methods to Estimate Effective Population Size Using Pedigree Data: Examples in Dog, Sheep, Cattle and Horse. *Genet. Sel Evol.* 45, 1. doi:10.1186/1297-9686-45-1

Li, X., Li, X., Fridman, E., Tesso, T. T., and Yu, J. (2015). Dissecting Repulsion Linkage in the Dwarfing Gene Dw3 Region for Sorghum Plant Height Provides Insights into Heterosis. *Proc. Natl. Acad. Sci. USA* 112 (38), 11823–11828. doi:10.1073/pnas.1509229112

Li, X., Wang, H., Zhou, R., Zheng, G., Li, L., and Shen, Z. (2009). Variation of 184C→T of Goat Callipyge Gene in Different Populations and its Effect on Body Weight. *Front. Agric. China* 3 (3), 319–324. doi:10.1007/s11703-009-0048-4

Lühken, G., Fleck, K., Pauciullo, A., Huisinga, M., and Erhardt, G. (2012). Familiar Hypopigmentation Syndrome in Sheep Associated with Homozygous Deletion of the Entire Endothelin Type-B Receptor Gene. *PLoS One* 7 (12), e53020. doi:10.1371/journal.pone.0053020

Marklund, S., Moller, M., Sandberg, K., and Andersson, L. (1999). Close Association between Sequence Polymorphism in the KIT Gene and the Roan Coat Color in Horses KIT Sequences: AJ224642–AJ224645.-->. *Mamm. Genome* 10 (3), 283–288. doi:10.1007/s003359900987

Matika, O., Robledo, D., Pong-Wong, R., Bishop, S. C., Riggio, V., Finlayson, H., et al. (2019). Balancing Selection at a Premature Stop Mutation in the Myostatin Gene Underlies a Recessive Leg Weakness Syndrome in Pigs. *Plos Genet.* 15 (1), e1007759. doi:10.1371/journal.pgen.1007759

Nicholas, F. W. (2021). Online Mendelian Inheritance in Animals (OMIA): a Record of Advances in Animal Genetics, Freely Available on the Internet for 25 Years. *Anim. Genet.* 52 (1), 3–9. doi:10.1111/age.13010

Old, R., Sewell, H., Norris, R., Steel, M., and Joyce, C. R. B. (1993). Heterozygote Advantage. *The Lancet* 341 (8839), 214. doi:10.1016/0140-6736(93)90076-s

Peng, B., Chen, H.-S., Mechanic, L. E., Racine, B., Clarke, J., Gillanders, E., et al. (2015). Genetic Data Simulators and Their Applications: an Overview. *Genet. Epidemiol.* 39 (1), 2–10. doi:10.1002/gepi.21876

Pruvost, M., Bellone, R., Benecke, N., Sandoval-Castellanos, E., Cieslak, M., Kuznetsova, T., et al. (2011). Genotypes of Predomestic Horses Match Phenotypes Painted in Paleolithic Works of Cave Art. *Proc. Natl. Acad. Sci.* 108 (46), 18626–18630. doi:10.1073/pnas.1108982108

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

Reissmann, M., and Ludwig, A. (2013). Pleiotropic Effects of Coat Colour-Associated Mutations in Humans, Mice and Other Mammals. *Semin. Cel Develop. Biol.* 24 (6-7), 576–586. doi:10.1016/j.semcdb.2013.03.014

Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., et al. (2015). A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *Plos Genet.* 11 (12), e1005629. doi:10.1371/journal.pgen.1005629

Salmi, B., Trefan, L., Bloom-Hansen, J., Bidanel, J. P., Doeschl-Wilson, A. B., and Larzul, C. (2010). Meta-analysis of the Effect of the Halothane Gene on 6 Variables of Pig Meat Quality and on Carcass Leanness1. *J. Anim. Sci.* 88 (9), 2841–2855. doi:10.2527/jas.2009-2508

Sartelet, A., Klingbeil, P., Franklin, C. K., Fasquelle, C., Géron, S., Isacke, C. M., et al. (2012). Allelic Heterogeneity of Crooked Tail Syndrome: Result of Balancing Selection. *Anim. Genet.* 43 (5), 604–607. doi:10.1111/j.1365-2052.2011.02311.x

Savage, A. E., and Zamudio, K. R. (2011). MHC Genotypes Associate with Resistance to a Frog-Killing Fungus. *Proc. Natl. Acad. Sci.* 108 (40), 16705–16710. doi:10.1073/pnas.1106893108

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 (7), 710–712. doi:10.1007/s003359901076

Siewert, K. M., and Voight, B. F. (2017). Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Mol. Biol. Evol.* 34 (11), 2996–3005. doi:10.1093/molbev/msx209

Simon, R., Lischer, H. E. L., Pieńkowska-Schelling, A., Keller, I., Häfliger, I. M., Letko, A., et al. (2020). New Genomic Features of the Polled Intersex Syndrome Variant in Goats Unraveled by Long-read Whole-genome Sequencing. *Anim. Genet.* 51 (3), 439–448. doi:10.1111/age.12918

Sironen, A., Uimari, P., Iso-Touru, T., and Vilkki, J. (2012). L1 Insertion within SPEF2 Gene Is Associated with Increased Litter Size in the Finnish Yorkshire Population. *J. Anim. Breed. Genet.* 129 (2), 92–97. doi:10.1111/j.1439-0388.2011.00977.x

Smith, L. B., Dally, M. R., Sainz, R. D., Rodrigue, K. L., and Oberbauer, A. M. (2006). Enhanced Skeletal Growth of Sheep Heterozygous for an Inactivated Fibroblast Growth Factor Receptor 31. *J. Anim. Sci.* 84 (11), 2942–2949. doi:10.2527/jas.2006-255

Thompson, K. G., Piripi, S. A., and Dittmer, K. E. (2008). Inherited Abnormalities of Skeletal Development in Sheep. *Vet. J.* 177 (3), 324–333. doi:10.1016/j.tvjl.2007.08.015

Trask, A. E., Bignal, E. M., McCracken, D. I., Monaghan, P., Piertney, S. B., and Reid, J. M. (2016). Evidence of the Phenotypic Expression of a Lethal Recessive Allele under Inbreeding in a Wild Population of Conservation Concern. *J. Anim. Ecol.* 85 (4), 879–891. doi:10.1111/1365-2656.12503

Tsairidou, S., Allen, A. R., Pong-Wong, R., McBride, S. H., Wright, D. M., Matika, O., et al. (2018). An Analysis of Effects of Heterozygosity in Dairy Cattle for Bovine Tuberculosis Resistance. *Anim. Genet.* 49 (2), 103–109. doi:10.1111/age.12637

Utsunomiya, Y. T., Milanesi, M., Utsunomiya, A. T. H., Torrecilha, R. B. P., Kim, E.-S., Costa, M. S., et al. (2017). A PLAG1 Mutation Contributed to Stature Recovery in Modern Cattle. *Sci. Rep.* 7 (1), 17140. doi:10.1038/s41598-017-17127-1

Vögeli, P., Bolt, R., Fries, R., and Stranzinger, G. (1994). Co-segregation of the Malignant Hyperthermia and the Arg615-Cys615mutation in the Skeletal Muscle Calcium Release Channel Protein in Five European Landrace and Pietrain Pig Breeds. *Anim. Genet.* 25 (Suppl. 1), 59–66. doi:10.1111/j.1365-2052.1994.tb00404.x

Wiedemar, N., and Drögemüller, C. (2015). A 1.8-kb Insertion in the 3′-UTR of RXFP2 Is Associated with Polledness in Sheep. *Anim. Genet.* 46 (4), 457–461. doi:10.1111/age.12309

Xiang, R., MacLeod, I. M., Bolormaa, S., and Goddard, M. E. (2017). Genome-wide Comparative Analyses of Correlated and Uncorrelated Phenotypes Identify Major Pleiotropic Variants in Dairy Cattle. *Sci. Rep.* 7 (1), 9248. doi:10.1038/s41598-017-09788-9

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

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

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