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

Xianfu Zhang1*, Kaixing Qu2, Peng Jia3, Jicai Zhang2, Jianyong Liu2, Chuzhao Lei3 and Bizhi Huang2*

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

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

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

INTRODUCTION

Domestic cattle generally refers to two subspecies, Bos taurus and Bos taurus indicus. They were domesticated in the Fertile Crescent (∼10,000 years ago) and the Indus Valley (∼8,000 years ago), respectively (Utsunomiya et al., 2019). These two subspecies can be interbred without barriers, unlike other bovine subspecies (buffalo, American bison, etc.) that have reproductive isolation (Wu et al., 2018). According to previous genomic study, the domestic cattle in the world could be divided into six major groups: European taurine, Eurasian taurine, East Asian taurine, African taurine, Indian indicine, and Chinese indicine (Chen et al., 2018). Besides that,
as one of the important routes for the migration of Indian indicine into China, there are many hybrid cattle with different lineages in the Yunnan region of China.

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

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

**MATERIALS AND METHODS**

**Sample Collection and Sequencing**

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

**Alignments and Variant Identification**

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

**Population Genetic Diversity**

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

The LD decay for each group was measured using PopLDdecay (Zhang et al., 2019) with default parameters. Runs of homozygosity (ROHs) were identified using the Runs of Homozygosity program implemented in PLINK, which slides a window of 50 SNPs (-homozyg-window-snp 50) across the genome in estimating homozygosity (Xia et al., 2021). The following settings were performed for ROH identification: -homozyg-density 50 -homozyg-window-het 1 -homozyg-window-missing 2. The number and length of ROH for each group were estimated, and the length of ROH was divided into
understanding of the gene functions and signaling pathways of the step by using VCFtools for each candidate gene. To gain a better derived from recent ancestors and short haplotype segments genomic loci formed ROH, with long haplotype segments historical, and recent inbreeding, respectively (Bhati et al., 2020). Available genomes from three representative groups (10 Indian zebu, new and six downloaded) were jointly genotyped with 29 publicly reveal the diversity of Dianzhong cattle, 10 Dianzhong cattle (four the average alignment rate and sequencing depth of the average of

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

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

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

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

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

In the Dianzhong cattle vs. Yanbian cattle comparison, the top 1% of candidate genes in four selection methods were extracted at the intersection (Supplementary Figure S2): MARS2 (Webb et al., 2015) and ARL6IP6 (Abumansour et al., 2015) (growth and development), RFTN2 (Wei et al., 2018) (DNA damage response), and N6AMT1 (Wang et al., 2020) (feed efficiency). GO and KEGG analyses were performed on candidate genes scanned twice or more by four methods. From DAVID gene ontology, 12 significant (\(p < 0.05\)) GO biological process (BP) terms were enriched (Supplementary Table S16). Among them, “oxidation–reduction process, GO:0055114” (\(n = 18\)) contained more genes than other GO terms. The other candidate genes were mainly enriched in hot resistance (“cardiac muscle cell differentiation, GO:0055007,” “lipid catabolic process, GO: 0016042,” “positive regulation of ATPase activity, GO: 0032781,” “mitochondrial DNA repair, GO:0043504,” and “regulation of oxidative stress-induced intrinsic apoptotic signaling pathway, GO:1902175”). Moreover, we noticed a region of about 0.06 Mb scanned by XP-CLR on chromosome 17 (including TRNAG-CCC, SDSL, SDS, and PLBD2), showing a strong positive selection signal. Thereinto, a missense mutation (c.944C > A, p.Ala315Glu) was found at the SDS gene. This mutation presented a huge divergence between Dianzhong cattle (allele C frequency = 1) and Yanbian cattle (allele A frequency = 1).
A total of 42 candidate genes were overlapped in four selection methods by comparing Dianzhong cattle with Indian zebu (Supplementary Figure S2), and three of them were significantly enriched in “kinase binding, GO: 0019900” (p-value = 0.02). In the functional prediction of all candidate genes, eight GO BP terms were significantly enriched (p < 0.05) (Supplementary Table S17). The candidate genes were mainly enriched in growth (“post-embryonic development, GO:0007519”, “skeletal muscle tissue development, GO:0006935”, and “embryonic skeletal system morphogenesis, GO:0048704”) and heat stress resistance (“DNA recombination, GO:0006310,” “inflammatory response, GO:0006954,” “response to hydrogen peroxide, GO:0042542,” and “respiratory gaseous exchange, GO:0007585”). Most genes (n = 11) were included in “inflammatory response, GO:0006954.” Furthermore, we detected a missense mutation (c.473A > G, p.Lys158Arg) at the fat deposition-related gene PDGFD. Allele A displayed a rare distribution (frequency 0.2) in Dianzhong cattle, whereas it showed an opposite pattern (frequency 0.9) in Indian zebu.

In the selection signal analysis between Dianzhong cattle and Chinese zebu, 36 common candidate genes were scanned by four selection methods. There were two genes enriched in the “response to retinoic acid, GO:0032526” (p-value = 0.04). The strongest signal was found in KIT gene, which was associated with sperm differentiation [33]. In KEGG analysis, four digestion-related pathways were enriched (Supplementary Table S18): “gastric acid secretion, bta04971,” “pancreatic secretion, bta04972,” “salivary secretion, bta04970,” and “carbohydrate digestion and absorption, bta04973.” In addition, three missense mutations (rs460251486, rs722912933, and rs517668236) with strong linkage effect were detected at candidate gene DDX4. According to the analyses for LD and haplotype construction of the three mutations, they exhibited a low recombination rate with r^2 values ranging from 0.31 to 0.61 (Supplementary Figure S3) [Ardlie et al., 2002]. DDX4 gene was the candidate gene selected by Fst (Figure 2A), and through the calculation of Fst and π with a smaller window (5 kb), a significant differentiation was observed between Dianzhong cattle and Chinese zebu (Figure 2B). Simultaneously, according to the
calculation of the combined haplotype frequencies at the three loci, the haplotype with the most distribution in Dianzhong cattle was the homozygous mutational haplotype (frequency 0.601), while the homozygous wild haplotype (frequency 0.700) was the commonest one in Chinese zebu (Figure 2C).

**DISCUSSION**

Genomic features may reflect a variety of multifarious historical events, including climate change, species introduction, and artificial selection (Chen et al., 2021). The characteristics of population genetic diversity are essential for assessing the genetic potential of breeds as well as for the utilization and protection of cattle breed resources (Xia et al., 2021). In this study, three underlying ancestors of Dianzhong cattle were selected as reference groups to explore the genetic information of Dianzhong cattle. The nucleotide diversity ranking calculated in this study (Chinese zebu > Dianzhong cattle > Indian zebu > Yanbian cattle) was basically consistent with the earlier results (Chen et al., 2018). Dianzhong cattle, belonging to hybrid cattle, obtained more abundant genetic information from multiple genetic sources, while Chinese zebu had high nucleotide diversity which may be due to previous reception of introgression from Banteng (Bos javanicus).
The monotonous environment and high degree of breeding may be the reasons why the nucleotide diversity of Yanbian cattle is lower than that of indicine and hybrid cattle (Singh et al., 2021; Xia et al., 2021). In addition, more 2–4 MB and 1–2 MB ROHS and more rapid LD decayed in Yanbian cattle confirmed this once again. The nucleotide diversity in our Indian zebu was higher than those in other groups, and the LD level was lower than others, suggesting that Indian zebu had a higher effective population size (Chen et al., 2021). In summary, the LD decay pattern and ROH distribution of each group were roughly consistent with the results of nucleotide diversity.

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

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

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

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

**DATA AVAILABILITY STATEMENT**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

**ETHICS STATEMENT**

The animal study was reviewed and approved by the Administration of Affairs Concerning Experimental Animals of China (Protocol number, WAFAC1008).
**AUTHOR CONTRIBUTIONS**

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

**FUNDING**

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

**ACKNOWLEDGMENTS**

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

**SUPPLEMENTARY MATERIAL**

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