Insights into adaption and growth evolution: a comparative genomics study on two distinct cattle breeds from Northern and Southern China

Chugang Mei,1,4,8 Linsheng Gui,1,6,8 Jieyun Hong,1,5,8 Sayed Haidar Abbas Raza,1 Chen Aorigele,2 Wanqiang Tian,3 Matthew Garcia,7 Yaping Xin,1 Wucai Yang,1 Song Zhang,1 and Linsen Zan1,4

1Northwest A&F University, Yangling 712100, Shaanxi, China; 2Inner Mongolia Agricultural University, Hohhot 010000, China; 3Yangling Vocational & Technical College, Yangling 712100, China; 4National Beef Cattle Improvement Center, Yangling 712100, China; 5Yunnan Agricultural University, Kunming 650031, China; 6Qinghai University, Xining 810008, China; 7Utah State University, Logan, UT 84322, USA

Mongolian cattle (MG, Bos taurus) and Minnan cattle (MN, Bos indicus) are two different breeds of Chinese indigenous cattle, representing North type and South type, respectively. However, their value and potential have not yet been discovered at the genomic level. In this study, 26 individuals of MN and MG were sequenced for the first time at an average of 13.9- and 12.8-fold, respectively. Large numbers of different variations were identified. In addition, the analyses of phylogenetic and population structure showed that these two cattle breeds are distinct from each other, and results of linkage disequilibrium analysis revealed that these two cattle breeds have undergone various degrees of intense natural or artificial selection. Subsequently, 496 and 306 potential selected genes (PSRs) were obtained in MN and MG, containing 1,096 and 529 potential selected genes (PSGs), respectively. These PSGs, together with the analyzed copy number variation (CNV)-related genes, showed potential relations with their phenotypic characteristics, including environmental adaptability (e.g., DVL2, HSPA4, CDHR4), feed efficiency (e.g., R3HD1, PLAG1, XKR4), and meat/milk production (e.g., PDHB, LEMD3, APOF). The results of this study help to gain new insights into the genetic characteristics of two distinct cattle breeds and will contribute to future cattle breeding.

INTRODUCTION
More than 1,000 cattle breeds are living in the world,1 making up a vital part of economic livestock by offering major sources (e.g., milk, meat, leather, and power) to humans. Modern cattle are stratified into two types according to the common usage, Bos taurus and Bos indicus, which have been differentiated from each other for >250,000 years.5,6 and independently domesticated in the Fertile Crescent ~8,000–10,000 years ago and the Indus Valley ~6,000–8,000 years ago.7,8 China has broad and diverse bovine genetic resources, with 53 indigenous cattle breeds. These Chinese breeds have various intrinsic characteristics, considered as important genetic resources for cattle around the world. It is generally known that for thousands of years, Chinese indigenous cattle breeds served as a major labor force in agricultural production and are well known for their endurance and adaption.7 Chinese cattle have long been used as draft animals and valued for their resistance to parasites, roughage-based diets, and the great tolerance to environmental challenges.10,11 Among the mentioned breeds, Mongolian cattle (MG) and Minnan cattle (MN) are two distinct types representing Bos taurus and Bos indicus. The former, living in Northern China, is herded primarily in the Inner Mongolian region and seems well adapted to the cold environment and grazing. The latter originates from South China and is known for its tolerance to muggy weather (Figure S1). These two cattle breeds are varying in their intrinsic characteristics, whereas their value and potential have not yet been discovered. In the course of scientific and technological innovations, two projects, namely HapMap and bovine genome, have been completed.12,13 In the meantime, the project of multiple bovine species has made rapid progress, and whole-genome resequencing includes Bos taurus,13–15 Bos indicus,15,16 Bos mutus, and Gayal.17,18 These projects explored the depth of evolution science of these large ungulates and better clarified the complex process of domestication and adaptation in these species, which suggested that the whole-genome sequencing is an efficient and effective method to explore genetic information of multiple species. In the present study, the whole genomes of 26 MG and MN cattle were sequenced to determine their genetic diversity. In addition, the genes with distinct characteristics that are positively selected through subtle combinations of human and natural selection were explored. The results of this study provided novel insights into their genetic difference in the whole genome and further identified genomic loci that might be highly important for cattle breeding programs.

RESULTS
Whole-genome sequencing and mapping
Whole-genome sequencing of the genomic DNA extracted from MG (n = 13) and MN (n = 13) was performed on an Illumina HiSeq X Ten

Received 15 July 2020; accepted 31 December 2020; https://doi.org/10.1016/j.omtn.2020.12.028.
*These authors contributed equally
Correspondence: Linsen Zan, No. 3 Taicheng Road, Northwest A&F University, Yangling 712100, Shaanxi, China.
E-mail: zanlinsen@163.com
fact that cattle in South China have higher genetic diversity. In this study, the TS/TV ratios (MN: 2.40, MG: 2.38) were close to the empirical human TS/TV ratio > 2.1, suggesting the high quality of the identified SNPs in an oblique manner. In addition, the heterozygous/homozygous ratios of MN and MG were 1.47 and 2.54, respectively. The ratio of heterozygous/homozygous SNPs in MN was found to be lower than that of MG, which is noteworthy, since MN have been considered to be indigenous. In addition, the heterozygous/homozygous ratios of MN and MG were 1.47 and 2.54, respectively. Moreover, it is believed that MG have experienced more robust artificial selection than MN.

### SNP/insertion or deletion (indel) detection and annotation

In total of 25,501,400 single-nucleotide polymorphisms (SNPs; MN: 23,644,213, MG: 15,309,216) were identified, of which 9.78% (MG: 9.73%, MN: 9.76%) were novel in comparison with the latest cattle SNP database (dbSNP Build 140; ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/archive/cow_9913/chr_rpts/) (Table 2). Transition-to-transversion (TS/TV) ratios were calculated as indicators of potential random sequence errors.13 In this study, the TS/TV ratios (MN: 2.40, MG: 2.38) were close to the empirical human TS/TV ratio > 2.1, suggesting the high quality of the identified SNPs in an oblique manner.19

| Table 1. Summary of sequence read alignments to the reference genome |
|-----------------------------------------------|-----------------|-----------------|
| Summary                  | MN (n = 13)     | MG (n = 13)     |
| Mean depth               | 13.9±          | 12.8±           |
| GC content rate          | 44.84%         | 44.78%          |
| Coverage rate            |                 |                 |
| Coverage rate (≥1x)      | 97.62%         | 97.98%          |
| Coverage rate (≥4x)      | 92.41%         | 92.08%          |
| Coverage rate (≥10x)     | 41.62%         | 34.59%          |
| Total reads (bp)         | 36,297,349,362 | 33,546,860,746  |
| Clean reads (bp)         | 35,533,395,809 | 32,492,362,240  |
| Q30                      | 92.45%         | 91.43%          |
| Mapped reads             | 98.37%         | 99.21%          |

In total of 25,501,400 single-nucleotide polymorphisms (SNPs; MN: 23,644,213, MG: 15,309,216) were identified, of which 9.78% (MG: 9.73%, MN: 9.76%) were novel in comparison with the latest cattle SNP database (dbSNP Build 140; ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/archive/cow_9913/chr_rpts/) (Table 2). Transition-to-transversion (TS/TV) ratios were calculated as indicators of potential random sequence errors.13 In this study, the TS/TV ratios (MN: 2.40, MG: 2.38) were close to the empirical human TS/TV ratio > 2.1, suggesting the high quality of the identified SNPs in an oblique manner.19

| Table 2. Functional classification of the detected single-nucleotide polymorphisms (SNPs) |
|-----------------------------------------------|-----------------|-----------------|
| SNP                        | MN (n = 13)     | MG (n = 13)     |
| Total numbers              | 23,644,213      | 15,309,216      |
| Heterozygous/homozygous    | 1.47            | 2.54            |
| Transition/transversion ratio | 2.40           | 2.38            |
| Novel rate (%)             | 9.73            | 9.76            |
| Intergenic                 | 13,605,091      | 8,827,342       |
| Up/downstream              | 323,597         | 214,385         |
| Gene                       | 9,715,525       | 6,267,489       |
| Intron                     | 8,521,597       | 5,477,158       |
| Splicing                   | 492             | 382             |
| Exonic                     | 191,076         | 129,293         |
| UTR                        | 1,002,360       | 660,656         |
| Non-synonymous             | 66,551          | 47,117          |
| Synonymous                 | 114,421         | 75,073          |
| Stop-gain                  | 573             | 399             |
| Stop-loss                  | 102             | 76              |
| Others                     | 9,429           | 6,628           | 10,402

Moreover, it is required by the beginning and developing stages of lactation in dairy cattle, and it is also associated with milk composition, reproduction, growth, carcass traits, and meat composition character of the dairy line and the intramuscular fat content of other varieties.22 It has been shown that the SPPI gene plays a critical part in regulating milk protein gene expression. The allelic variants of this gene have also been shown to have a relation to the variation in milk composition.23,24 It has been suggested that the expression of PPARGC1A is required by the beginning and developing stages of lactation in dairy cattle, and it is also associated with milk composition, reproduction, growth, carcass traits, and meat...
quality. SLC27A6 and ITFG1 were considered as candidate genes of milk fat mammary glands and daily production. PRKG1 is involved in feed efficiency and gap junction, acting as a candidate gene for intramuscular fat in pigs. It is well known that TG affects meat quality traits and carcass in beef cattle. CRH can significantly suppress glucose uptake and stimulate fat breakdown. This gene is critical for lipid metabolism, highly associated with ribiye area and marbling as well as the OMEGA 6:3 ratio. In addition, it was identified that GPC1 correlated with meat production and quality, which is also an economically significant trait, having been considered broadly during arti
cultural selection. It has been recently found that NEB is associated with both lipid and organoleptic traits. ATP8A1 is also correlated with weight gain, residual feed intake, feed conversion ratio, and feed intake. In addition, it was reported that some other genes are involved in body size, fertility, production, and milk fatty acid profile (e.g., FANCC, SRY, SLC5A4, and CTNNBLI). Moreover, several immune system activation genes respond to environmental stress, including CD59, CDH9, RLRBP1, ROLA, and STOM. These have close relevance to parasitic diseases, thereby clarifying the function of these genes and unraveling the underlying mechanisms of the innate immunity against several important tropical environments. Variants in SOX5 have effects on phenotype in colder climates. Furthermore, 1,164 and 596 breed-specific CNVs were included in MN and MG, covering 124 and 67 breed-specific CNV-related genes, respectively (Figures S5 and S6). According to enrichment analysis, genes in MN were significantly upregulated in adaption-related terms (e.g., GO:0042612—MHC class I protein complex, GO:0002474—antigen processing and presentation of peptide antigen via MHC class I, bta04360:Axon guidance, and GO:0006955—immune response, which might result from thousand-year natural selection in a particular environment) (Table S6). In addition to this, the MG-specific CNV-related genes were upregulated (e.g., GO:0003383—apical constriction, GO:0032525—somite rostral/caudal axis specification, GO:0006654—phosphatidic acid biosynthetic process, bta04726: Serotonergic synapse, etc.) (Table S7). This is consistent with the fact that MN (indicene cattle) are also more resistant to ticks, gastrointestinal parasites, and rinderpest than are MG (taurine cattle).

**Population structure**

The neighbor-joining tree and population structure indicated that MN and MG can be stratified into two groups (Figures 1A, 1B, and 1C). Also, according to principal-component analysis (PCA), MN and MG are significantly separated along the first principal component, accounting for the largest proportion of variability (Figures 1C and S8).

Moreover, it was observed that linkage disequilibrium decays more slowly in MG than that in MN (Figure 1D). This may result from the population bottleneck event during domestication. Furthermore, the results of this study showed that the genetic diversity of MN cattle was higher than that of MG (Figure 2), which agrees with the different living environments and artificial selection directions of these two cattle types.

**Selective sweep analysis**

As a result, 496 and 306 PSRs were obtained in MN and MG, a total of 21.79 and 7.16 Mb, covering 1,096 and 529 potential selected genes (PSGs) (Figure 3A; Tables S8 and S9). Under domestication, extensive selection created different phenotypic features and biological characteristics for a variety of cattle breeds. Natural selection usually results in directional changes for adaptation traits, particularly acclimatization to harsh environments and resistance to parasites and other diseases for survival in a certain environment. Likewise, artificial selection exerts its impacts on the development of specific traits of economic importance (e.g., milk, meat, and fertility) through genetic improvement. According to further investigation of this study, a considerable number of the PSGs were correlated with shapping characteristics of the populations, including morphological and production traits (e.g., adaption, coat color, meat traits, etc.) (Table 3).

**DISCUSSION**

**Candidate genes correlated with environmental adaptability**

This study investigated whether the domestication and artificial selection could have shaped the genomes of MG and MN and how to adapt to local environmental challenges (e.g., parasite and viral challenges). It has been found that some of the PSGs identified in MN are associated with environmental adaptability (e.g., DVL2, VPS13A, GNA14, KLHL3, HSP4A, GPR50, and FGF9). Previous study suggested the regulation of the downstream of DVL2 in an endogenous Wnt pathway that can be operated in outer root sheath cells. DVL2 is a cell type that was previously shown to play a role in limiting hair
growth and the candidate for wool production because of its link to the hair follicle cycle. In this study, the frequencies of DVR2 mutations are obviously different between the MG and MN, especially the three mutations in CDS including Chr19:27574636A>G, Chr19:27574810G>A, and Chr19:27578136G>A (Figure 3B). These variations might affect the normal transcription and expression of the DVR2 gene and further have impacts on hair growth, which makes it adapt to the warm climate in the south and the cold climate in the north. Moreover, for the other PSGs in MN, VPS13A and GNA14 genes were located in significant sweep regions of the pig, which were correlated with blood coagulation and circulation and enabled temperature adaptation. KLHL3 was over-represented in biological processes correlated with kidney development, which can affect water reabsorption in extreme temperature conditions in several species. HSPA4 belongs to the family of heat shock 70 kDa protein, known for improving cell protection against heat damage and preventing protein denaturation. Genome-wide analysis of African cattle also showed selective sweeps for heat tolerance in the gene region. GPR50 plays a role in thermogenesis and might be directly correlated with wisent adaptation to colder climatic conditions. It was suggested that FGF9, the functions of which are correlated with the developments of the respiratory system, respiratory tube, and lung, is the adaptive gene in a test of genetic differentiation between domestic and argali sheep. Altogether, the selection on those genes agrees with a previous report that zebu breeds have a better ability to regulate body temperature to respond to heat stress. In addition, we further detected genes (e.g., CDHR4, IP6K1, and RNF123 under selection in the MG genome and MONIA, MST1R, DNAH2, LCT, and THPO under selection in the MN genome) that were reported as involved in internal and external parasite tolerance in different cattle breeds.

In addition, some PSGs found in MN (e.g., SAR1B, IFNAR2, HSPA9, CD5, VAMP7, IL15, FOXC2, and TPM2) were briefly reported to be associated with immune response and disease resistance in cattle. It is noteworthy that PSGs in MG also cover the RXFP2 and MC1R gene. RXFP2 serves as a receptor for the relaxin and insulin-like factor 3 proteins, of which the impacts on horn status and size are dependent on its biochemical interaction with testosterone. It has been suggested to be critical for developing the horns of goats and cattle and to determine the size and presence of the horns of wild and domestic sheep. Previous reports suggested that mutations in MC1R generate red (or chestnut) coat colors in many different species (e.g., dogs, mice, horses, and cattle). Selection on these genes associated with horn or coat color may be responses to adapt the free-range environment of most MG cattle. Some PSGs in MN (e.g., CNGB1 and R3HDM1) were also recommended as candidate genes for meat traits by whole-genome resequencing of the Japanese native cattle Kuchinoshima-Ushi. KLHL3 was over-represented in biological processes correlated with kidney development, which can affect water reabsorption in extreme temperature conditions in several species.
and ITPR3, have underlying correlations with economically significant traits (e.g., domestication, growth rate, and meat/milk production) in other cattle breeds. These results above might be responses to diet adaptation and artificial selection on meat and dairy traits.

MATERIALS AND METHODS

Ethics approval

All the experimental procedures with cattle used in the present study were approved by the Experimental Animal Manage Committee (EAMC) of Northwest A&F University (2011-31,101,684). All the operations and experimental procedures complied with the National Standard of Laboratory Animals Guidelines for Ethical Review of Animal Welfare (GB/T 35892-2018) and Guide for the Care and Use of Laboratory Animals: Eighth Edition.

Sample preparation and sequencing

Blood samples of MG (n = 13, female) and MN (n = 13, female) were randomly collected from Inner Mongolia and Fujian, China, respectively. The 5 μg of blood extracted genomic DNA underwent shearing into small fragments of 200–800 bp using the Covariates system (Life Technologies). The DNA library was built from these fragments through ligation of paired-end adapters and insertion with 500 bp PCR amplification. The amplicons underwent sequencing via the Illumina HiSeq X Ten platform, and 150 bp paired-end reads were produced.

Alignment and variation identification

The paired-end reads of 100 bp length reads were aligned to the reference Bos_taurus_UMD_3.1 by Burrows-Wheeler-Alignment (BWA) software with the parameters as follows: “bwa aln-m200000-o1-e30-15-L-I-t4-n0.04-R20-f” and “bwa sampe-a 650-n30-N30.” The output was aligned and converted to bam format from SAM format with the use of SAMtools. The resultant bam files underwent sorting, and the reads underwent filtering in duplicate using Picard package (http://broadinstitute.github.io/picard/, version 1.92). The indels and SNPs were identified using Genome Analysis Toolkit (GATK, version 2.4-9). Beagle was employed to refine the genotype calls and infer the haplotypes using genotype likelihoods from the GATK result. CNVs were identified using CNVnator v0.2.7 with default parameters. CNVs with fewer than 3 reads supported were removed.

Variation annotation

Computer software ANNOVAR was applied for the identification of the variation resulting from these variants in protein coding regions and consequently on the amino acid sequences altered. The word “upstream” was used for the variants overlapping the 1,000 bp upstream region relative to transcription start site, while the term “downstream” annotates the variants overlapping 1,000 bp downstream region relative to the end site of the gene. Likewise, “upstream/downstream” reveals that the variant exists in both upstream and downstream positions (probably two different genes). The term
| Gene   | Description                                      | Chr. | FST   | Traits          | Reference |
|--------|--------------------------------------------------|------|-------|-----------------|-----------|
| VPS13A | vacuolar protein sorting 13 homolog A            | 8    | 0.6781| adaptability    | 41        |
| GNA14  | G protein subunit alpha 14                       | 8    | 0.6932| adaptability    | 41        |
| KLHL3  | kelch like family member 3                       | 7    | 0.7136| adaptability    | 41        |
| DVL2   | dishevelled segment polarity protein 2           | 19   | 0.8091| adaptability    | 41        |
| HSPA4  | heat shock 70kDa protein 4                       | 7    | 0.6965| adaptability    | 41        |
| GPR50  | G protein-coupled receptor 50                    | X    | 0.7483| adaptability    | 41        |
| FGFR9  | fibroblast growth factor 9                       | 12   | 0.6802| adaptability    | 41        |
| FGF2   | fibroblast growth factor 2                       | 19   | 0.8091| adaptability    | 41        |
| LCT    | lactase                                          | 2    | 0.6614| parasite resistance | 33    |
| THPO   | thrombopoietin                                   | 1    | 0.7626| parasite resistance | 33    |
| SARK1  | secretion associated Ras related GTPase 1B       | 7    | 0.7129| immune response  | 41        |
| IFNAR2 | interferon alpha and beta receptor subunit 2     | 1    | 0.6516| immune response  | 41        |
| HSPA9  | heat shock protein family A member 9             | 7    | 0.6390| immune response  | 41        |
| CD5    | CD5 molecule                                     | 29   | 0.6496| immune response  | 41        |
| VAMP7  | vesicle-associated membrane protein 7            | X    | 0.8344| immune response  | 41        |
| IL15   | interleukin 15                                   | 17   | 0.6597| immune response  | 41        |
| FOXO2  | forkhead box C2                                  | 18   | 0.7322| disease resistance | 42   |
| TPM2   | tropomyosin 2                                    | 8    | 0.6497| disease resistance | 42   |
| CNGB1  | cyclic nucleotide gated channel beta 1           | 18   | 0.6554| feed efficiency  | 41        |
| RHDM1  | R3H domain containing 1                          | 2    | 0.6497| feed efficiency  | 41        |
| FIK3CB | phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta | 1 | 0.7196| feed efficiency  | 41        |
| FGFR1  | fibroblast growth factor 1                       | 7    | 0.6860| reproduction     | 41        |
| ADRR2  | adrenoceptor beta 2                              | 7    | 0.7676| meat traits      | 41        |
| PDHB   | pyruvate dehydrogenase beta                      | 22   | 0.7017| meat traits      | 41        |
| SLC2A4 | solute carrier family 2 member 4                 | 19   | 0.7437| meat traits      | 41        |
| PCSK4  | proprotein convertase subtilisin/kexin type 4    | 7    | 0.7427| fertility        | 41        |
| SPDOCK1| SPARC/osteonectin, ccwv and kazal like domains proteoglycan 1 | 7 | 0.7334| fertility        | 41        |
| C4B1N1 | calcineurin binding protein 1                    | 17   | 0.6795| fertility        | 41        |
| SLC38A3| solute carrier family 38 member 3                | 22   | 0.8155| meat traits      | 41        |
| EOX3   | exocyst complex component 3                      | 20   | 0.8345| meat traits      | 41        |
| STAT5B | signal transducer and activator of transcription 5B | 19 | 0.6768| meat traits      | 41        |
| EMD    | emerin                                           | X    | 0.9695| meat traits      | 41        |

Mongolian cattle

| Gene   | Description                                      | Chr. | FST   | Traits          | Reference |
|--------|--------------------------------------------------|------|-------|-----------------|-----------|
| MC1R   | melancortin 1 receptor                           | 18   | 0.6532| coat color      | 37        |
| RXFP2  | relaxin/insulin like family peptide receptor 2   | 12   | 0.7337| horn development | 37,46     |
| CDH1R4 | cadherin related family member 4                 | 22   | 0.6483| parasite resistance | 35    |
| IP6K1  | inositol hexakisphosphate kinase 1               | 22   | 0.7258| parasite resistance | 35    |
| RNF123 | ring finger protein 123                          | 22   | 0.7129| parasite resistance | 35    |
| LEM3   | LEM domain containing 3                          | 5    | 0.6871| growth, meat traits | 42,43   |
| NCOA1  | nuclear receptor coactivator 1                   | 11   | 0.6699| muscular development, reproduction | 42,43 |
| APOF   | apolipoprotein F                                 | 5    | 0.7752| meat traits      | 41        |

(Continued on next page)
Table 3. Continued

| Gene      | Description                  |Chr. | $F_{ST}$ | Traits          | Reference |
|-----------|------------------------------|-----|----------|-----------------|-----------|
| HEYL      | hes related bHLH transcription factor with YRPW motif-like | 3   | 0.6418   | meat traits     | 45        |
| SLC29A1   | solute carrier family 29 member 1 | 23  | 0.7452   | meat traits     | 45        |
| TIFF1| tuftelin interacting protein 11 | 17  | 0.6634   | meat traits     | 46        |
| ASNS      | asparagine synthetase         | 4   | 0.6548   | meat traits     | 46        |
| RPL15     | ribosomal protein L15         | 27  | 0.7344   | meat traits     | 46        |
| RELA      | RELA proto-oncogene, NF-kB subunit | 29  | 0.6763   | dairy traits    | 46        |
| RDH5      | retinol dehydrogenase 5       | 5   | 0.7020   | meat traits     | 46        |
| ITPR3     | inositol 1,4,5-trisphosphate receptor type 3 | 23  | 0.6633   | meat traits     | 46        |

“stop gain” refers to the situation in which a stop codon was created by a nonsynonymous SNP in the variant sequence, while the term “stop loss” means loss of stop codon in the variant sequence site by the nonsynonymous SNP. The word “unknown” refers to the unknown function caused by certain errors within the genetic structure (defined in the database file). The term “splicing” refers to the variant’s placement position being inside the 2 bp splicing junction. The missense mutation functional impact was predicted by the SIFT algorithm.56 During annotation through ANNOVAR software, the source databases used included NCBI, dbSNP build 140 (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/cow_9913/chr_rpts/) and NCBI RNA seq. Ensembl release 78 provided by UCSC (ftp://hgdownload.cse.ucsc.edu/goldenPath/bosTau6/database).

**Phylogenetic and population structure analyses**

By using PHYLIP v3.695 program (https://evolution.genetics.washington.edu/phylip.html), this study built the neighbor-joining phylogenetic tree of the SNP data. By using ADMIXTURE,57 this study further inferred the structure of the population. The smartPCA program of the EIGENSOFT package was applied for the analysis of principal components.58

**Linkage disequilibrium**

For the estimation of the genome-wide linkage disequilibrium (LD) of each breed, the mean $r^2$ values for pairwise markers were calculated using Haplovieview59 software. This study only employed SNPs with a minor allele frequency above 0.05.

**Selective sweep analysis and enrichment analysis**

Domestication and artificial selection led to a reduction in nucleotide diversity and changed allele frequency. To identify the genes that have undergone positive selection and investigate the differences between these two cattle breeds, the differentiation ($F_{ST}$) was measured, and their genetic diversity ratio ($\theta_{pq}$) was compared with the use of a sliding window method (50 kb window and 10 kb step). The 5% of windows with the highest $F_{ST}$ and $\theta_{pq}$ ratio were considered the potential selected windows (PSWs), and the adjacent windows were merged into a single region, thereby becoming the potential selected region. The population variation was estimated by pairwise FST using an unbiased estimator.60 A sliding window approach was employed for measuring mean pairwise nucleotide diversity ($\theta_{pq}$) and $H_{p}$ value of each breed through VCFtools software with default parameters and 50 kb sliding windows in 10 kb steps.61 The $\theta_{pq}$ ratio between group MN and MG was calculated as $\ln(\theta_{pq,MN}/\theta_{pq,MG})$, reflecting the loss of nucleotide diversity in MG relative to MN. Tajima’s D statistic was calculated with the use of VCFtools for each candidate gene. Selective sweep regions were labeled with cattle QTLdb release 29 from the Animal Quantitative Trait Loci Database.62 Functional enrichment analyses of PSGs were conducted using DAVID tools.63

**SUPPLEMENTAL INFORMATION**

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2020.12.028.

**ACKNOWLEDGMENTS**

The research was supported by the National Key Research and Development Program of China (2018YFD0501700), the China National Natural Science Foundation of China (31972994), the China Postdoctoral Science Foundation (2020T130540), the Key Research and Development Program of Ningxia Province (2019BF02004), and the National Beef and Yak Industrial Technology System (CARS-37). We thank all the research assistants and laboratory technicians who contributed to this work.

**AUTHOR CONTRIBUTIONS**

L.Z. and C.M. conceived the study. C.M., L.G., J.H., and C.A. analyzed and interpreted data. C.M., L.G., and L.Z. wrote the manuscript. S.H.A.R., Y.X., W.T., W.Y., J.H., S.Z., M.G., and L.Z. contributed tools and materials. All authors read and approved the final manuscript.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**REFERENCES**

1. Scherf, B.D., and Pilling, D. (2015). The Second Report on the State of the World’s Animal Genetic Resources for Food and Agriculture (Food and Agricultural Organization of the United Nations).
2. Hiendleder, S., Lewalski, H., and Janke, A. (2008). Complete mitochondrial genomes of Bos taurus and Bos indicus provide new insights into intra-species variation, taxonomy and domestication. Cytogenet. Genome Res. 120, 150–156.
3. Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., et al.; Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528–532.

4. Canavez, F.C., Luche, D.D., Stothard, P., Leite, K.M.R.S., Sousa-Canavez, J.V., Plastow, G., Meidanis, J., Souza, M.A., Feijao, P., Moore, S.S., and Camara-Lopes, L.H. (2012). Genome sequence and assembly of Bos indicus. J. Hered. 103, 342–348.

5. Porto-Neto, L.R., Sonstegard, T.S., Liu, G.E., Bickhart, D.M., Da Silva, M.V., Machado, M.A., Usunumiyama, Y.T., Garcia, J.F., Gondo, C., and Van Tassell, C.P. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. BMC Genomics 14, 876.

6. Raza, S.H.A., Khan, S., Amjadi, M., Abdelnour, S.A., Ohran, H., Alannii, K.M., Abd El-Hack, M.E., Taha, A.E., Khan, R., Gong, C., et al. (2020). Genome-wide association studies reveal novel loci associated with carcass and body measures in beef cattle. Arch. Biochem. Biophys. 694, 108543.

7. Van Vuure, T. (2002). History, morphology and ecology of the Aurochs (Bos taurus primigenius). Lutra 44, 3–18.

8. Bickhart, D.M., Xu, L., Hutchison, J.L., Cole, J.B., Null, D.J., Schroeder, S.G., Song, J., Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., Zhao, C., Zhao, S., Song, K.-T., Chai, H.-H., Cho, Y.-M., and Lim, D. (2015). A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 16, R42.

9. Choi, J.-W., Choi, B.-H., Lee, S.-H., Lee, S.-S., Kim, H.-C., Yu, D., Chung, W.-H., Lee, K.-T., Chai, H.-H., Cho, Y.-M., and Lim, D. (2015). Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. Mol. Cells 38, 466–473.

10. Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., Zhao, C., Zhao, S., Song, J., Guang, X., et al. (2018). Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol. Biol. Evol. 35, 688–699.

11. Kim, J., Hanotte, O., Mwai, O.A., Dessie, T., Bashir, S., Diallo, B., Agaba, M., Kim, K., Kwak, W., Sung, S., et al. (2017). The genome landscape of indigenous African cattle. Genome Biol. 18, 34.

12. Taye, M., Kim, J., Yoon, S.H., Lee, W., Hanotte, O., Dessie, T., Kemp, S., Mwai, O.A., Caetano-Anolles, K., Cho, S., et al. (2017). Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. BMC Genet. 18, 11.

13. Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yak genome and adaptation to life at high altitude. Nat. Genet. 44, 946–949.

14. Wang, K., Hu, Q., Ma, H., Wang, L., Yang, Y., Luo, W., and Qiu, Q. (2014). Genome-wide variation within and between wild and domestic yak. Mol. Ecol. Resour. 14, 794–801.

15. Mei, C., Wang, H., Zhu, W., Wang, H., Cheng, G., Qu, K., Guang, X., Li, A., Zhao, C., Yang, W., et al. (2016). Whole-genome sequencing of the endangered bovine species Gayal (Bos frontalis) provides new insights into its genetic features. Sci. Rep. 6, 19787.

16. Huang, 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.

17. Liu, G.E., and Bickhart, D.M. (2012). Copy number variation in the cattle genome. Funct. Integr. Genomics 12, 609–624.
