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 Zan[1,4]

[1]Northwest A&F University, Yangling 712100, Shaanxi, China; [2]Inner Mongolia Agricultural University, Hohhot 010000, China; [3]Yangling Vocational & Technical College, Yangling 712100, China; [4]National Beef Cattle Improvement Center, Yangling 712100, China; [5]Yunnan Agricultural University, Kunming 650031, China; [6]Qinghai University, Xining 810008, China; [7]Utah 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., *R3HDM1, 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[2–6] 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.[9] 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.[11,12] 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

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 (≥1×)	97.62%	97.98%
Coverage rate (≥4×)	92.41%	92.08%
Coverage rate (≥10×)	41.62%	34.59%
Total reads (bp)	36,297,349,362 (100%)	33,546,860,746 (100%)
Clean reads (bp)	35,533,395,809 (97.85%)	32,492,362,240 (96.83%)
Q30	92.45%	91.43%
Mapped reads	98.37%	99.21%

Table 2. Functional classification of the detected single-nucleotide polymorphisms (SNPs)

SNP	MN	MG	Total
Total numbers	23,644,213	15,309,216	25,502,400
Heterozygous/homozygous	1.47	2.54	2.00
Transition/transversion ratio	2.40	2.38	2.39
Novel rate (%)	9.73	9.76	9.78
Intergenic	13,605,091	8,827,342	14,641,248
Up/downstream	323,597	214,385	399,595
Gene	9,715,525	6,267,489	10,461,557
Intronic	8,521,597	5,477,158	9,211,436
Splicing	492	382	557
Exonic	191,076	129,293	210,716
UTR	1,002,360	660,656	1,038,848
Non-synonymous	66,551	47,117	74,679
Synonymous	114,421	75,073	124,859
Stop-gain	573	399	669
Stop-loss	102	76	107
Others	9,429	6,628	10,402

(Table 1). A total of 907.97 gigabases of high-quality paired-end reads (150 bp) was produced, and an average mapping depth of 13.9 (MN) and 12.8 (MG) fold were presented, respectively. 97.62% (MN) and 97.98% (MG) of the acquired reads were plotted to distinctive positions on the reference genome (*Bos_taurus*_UMD_3.1).[12]

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] 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, over the past thousands of years, the population size has decreased significantly. This could be caused by natural or artificial selection of the genotype. With the use of the RefSeq and Ensembl gene sets, variations were annotated. 8,521,597 and 5,477,158 of the total SNPs in MN and MG were in introns, 492 and 382 in splice sites, and 191,076 and 129,293 in exonic regions, which includes 1,002,360 and 660,656 in untranslated regions (UTRs), 66,551 and 47,117 non-synonymous, 114,421 and 75,073 synonymous, 573 and 399 stop-gain, and 102 and 76 stop-loss, respectively. Moreover, 2,800,408 and 1,785,757 indels were found in MN and MG genomes, respectively, 9.94% of which were newfound (Table S1). The insertion/deletion ratio of the total indels reached 0.73 (Table S1). The indels were mostly 1 bp in length (Figure S2). Interestingly, concerning the detected variations, 10,193,184 and 1,858,187 SNPs and 1,248,546 and 233,895 indels were breed specific in MN and MG, respectively (Figures S3 and S4). The higher ratios of total variations and breed-specific variations in MN than those in MG are consistent with the fact that cattle in South China have higher genetic diversity.

Moreover, it is believed that MG have experienced more robust artificial selection than MN.

CNV discovery and annotation

Copy number variations (CNVs) refer to DNA segments no less than 1 kb with varying copy numbers compared to a reference genome. In the genome, CNVs have wide distribution. In comparison with the SNPs, CNVs are capable of affecting most genomes and leading to effects similar to changing gene structure and dosage, as well as regulating and exposing recessive alleles. CNVs also have a correlation with many diseases and probably promote several missing heritabilities. After quality control, a total of 4,760 and 3,946 CNVs were found in MN and MG genomes (Tables S2 and S3), respectively, covering 32.19 Mb (1.13%) and 26.41 Mb (0.93%) of the reference and containing 442 and 385 genes (Tables S4 and S5). As previous reports suggested, some of these CNV-related genes were correlated with coat color, dairy traits, meat production/quality traits, and reproduction traits, as well as immunity and adaption. *KIT*, *ASIP*, and *AP3B1* are likely to be involved in coat color, not only in cattle but also in other domesticated mammals.[15,20,21] In particular, because of the expression of the *ASIP* gene in adipocytes and its significance in obese yellow mice, this transcript was also believed to be related to the milk composition character of the dairy line and the intramuscular fat content of other varieties.[22] It has been shown that the *SPP1* 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
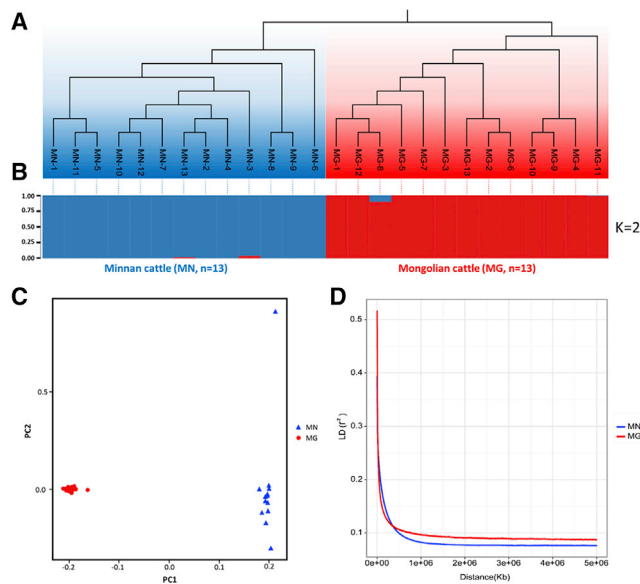
Figure 1. Population analysis of the two sequenced cattle breeds
(A) Neighbor-joining phylogenetic. (B) Genetic structure. The length of each colored segment represents the proportion of the individual's genome from $K = 2$ ancestral populations. (C) Principal-component analysis (PCA). (D) Linkage disequilibrium (LD) analysis.

quality.[25] *SLC27A6* and *ITFG1* were considered as candidate genes of milk fat mammary glands and daily production.[26,27] *PRKG1* is involved in feed efficiency and gap junction, acting as a candidate gene for intramuscular fat in pigs.[16] It is well known that *TG* affects meat quality traits and carcass in beef cattle.[28,29] *CRH* can significantly suppress glucose uptake and stimulate fat breakdown. This gene is critical for lipid metabolism, highly associated with ribeye area and marbling as well as the OMEGA 6:3 ratio.[30] 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 artificial selection.[31] It has been recently found that *NEB* is associated with both lipid and organoleptic traits.[32] *ATP8A1* is also correlated with weight gain, residual feed intake, feed conversion ratio, and feed intake.[16] In addition, it was reported that some other genes are involved in body size, fertility, production, and milk fatty acid profile (e.g., *FANCC*,[33] *SRY*,[13] *SLC5A4*,[34] and *CTNNBL1*[35]). Moreover, several immune system activation genes respond to environmental stress, including *CD59*,[36] *CDH9*,[37] and *PROCR*.[38] *RLBP1*,[39] *BOLA*,[40] and *STOM*[15] 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.[21] 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 (indicine 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 S7). 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.[9] 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.[9] According to further investigation of this study, a considerable number of the PSGs were correlated with shaping 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*, *HSPA4*, *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
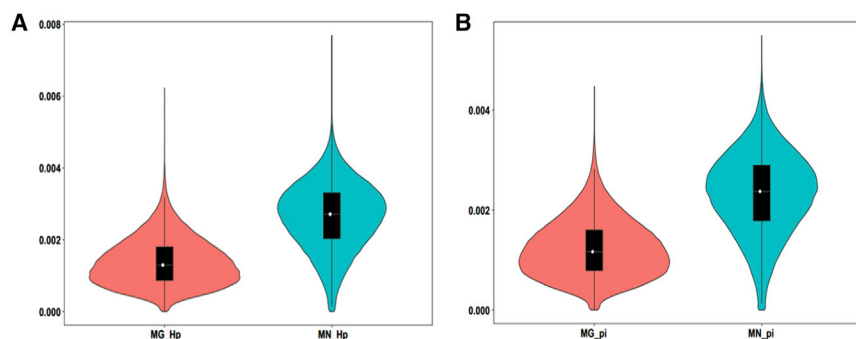
growth and the candidate for wool production because of its link to the hair follicle cycle.[42] In this study, the frequencies of *DVL2* mutations are obviously different between the MG and MN, especially the three mutations in CDS including Chr19:27574636A>G, Chr19:27574810G>A, and Chr19:27578136G>A (Figure 3B). These variations might affect the normal transcription and expression of the *DVL2* 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.[41] *KLHL3* was over-represented in biological processes correlated with kidney development, which can affect water reabsorption in extreme temperature conditions in several species.[41] *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.[15] *GPR50* plays a role in thermogenesis and might be directly correlated with wisent adaptation to colder climatic conditions.[43] 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.[44] 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 *MON1A*, *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.[33,38,39,43]

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.[11,32,46,47] 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.[9,42,46] Previous reports suggested that mutations in *MC1R* generate red (or chestnut) coat colors in many different species (e.g., dogs, mice, horses, and cattle)[15] 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 olfaction and food conversion efficiency,[22,43] respectively. Moreover, *PIK3CB* was reported to be involved in the Akt/PI3K and MAPK signaling pathways, which is important for high feed efficiency.[16] Selection of these genes in MN may facilitate the adaption to available food resources and vegetation diversity (forest/steppe habitat). These genes are recognized the promising candidates for further investigation.

Candidate genes correlated to production and quality traits

As agriculture mechanization has been developed, these traits are gradually depreciated, and the demand for beef is increasing with the rapid improvement of people's living standards. Hence, genetic improvement through selection on genes that are correlated with economic characters (e.g., meat production and quality) has been largely considered during artificial selection.

In MN, because of their role in fertilization, growth, and embryonic development, some of the PSGs (e.g., *SLC38A3*, *EXOC3*, *STAT5B*, and *EMD*) were also recommended as candidate genes for meat traits by whole-genome resequencing of the Japanese native cattle Kuchinoshima-Ushi.[48] *FGF1* was implicated in reproduction traits.[9] It was found that polymorphisms in *ADRB2* have impacts on the selected carcass traits of Qinchuan cattle.[49] *PDHB*, located in the regions of homozygosity of Hanwoo (Korean cattle), has been involved in the tricarboxylic acid cycle.[32,38] High-resolution analysis between fine-wool Merino and coarse-wool Churra sheep breeds found the *SLC2A4* gene was harbored in the selection sweeps and linked to meat production/quality traits.[42] In addition, *PCSK4*, *SPOCK1*, and *CABIN1* have vital roles in fertility and are recognized to be vital candidates for reproduction traits in animals as well.[9,34] Likewise, in MG, some genes were found to have relationships with animal production traits. *LEMD3* showed extreme allele frequencies among different sheep breeds in previous research.[42] Intronic variation in this gene has direct correlations with a pleiotropic quantitative trait locus (QTL) reported in cattle for calving ease direct, birth weight, ribeye muscle area, and marbling.[42] According to Gautier et al.,[43] *NCOA1* could be considered as a candidate gene for skeletal and muscular development, reproduction, and embryonic and organ development in cattle. *APOF* has key functions in lipid metabolism.[43] Furthermore, it was reported that some genes, including *HEYL*, *SLC29A1*, *TFIP11*, *ASNS*, *RPL15*, *RELA*, *RDH5*,
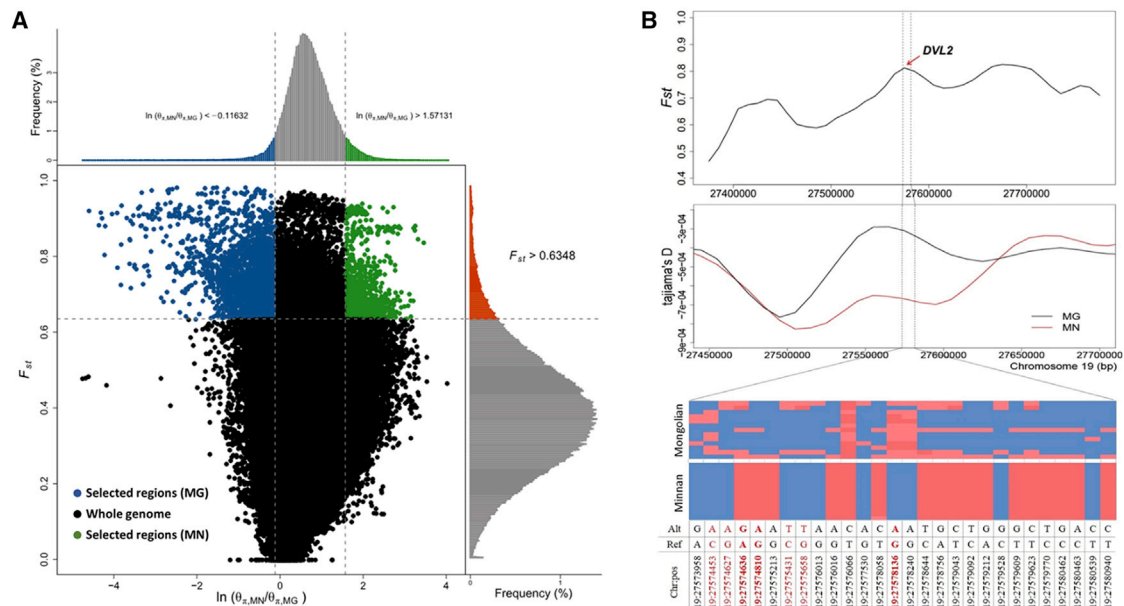
Figure 3. Overview of selective sweeps in the MG and MN genomes

(A) Distribution of θπ ratios ln($θ_{π,MN}/θ_{π,MG}$) and F_{ST} values, calculated in 50 kb windows sliding in 10 kb steps. Data points respectively located to the left and right of the vertical dashed lines (representing 5% left and right tails of the empirical θπ ratio distribution) and above the horizontal dashed line (the 5% right tail of the empirical F_{ST} distribution) were recognized as the regions selected for MG (blue points) and MN (green points), respectively. (B) Example of genes with strong selective sweep signals. F_{ST} values and Tajima's D values are shown on the top and middle. Haplotypes inferred from genotyping the SNPs across the *DVL2* gene are shown on the bottom (blue, red, and deep red colors represent non-mutation, heterozygous mutations, and heterozygous mutations in comparison with the reference, respectively).

and *ITPR3*, have underlying correlations with economically significant traits (e.g., domestication, growth rate, and meat/milk production) in other cattle breeds.[48] 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 Covaries system (Life Technologies). The DNA library was built from these fragments through ligation of paired-end adopters 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[12] 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."[50] The output was aligned and converted to bam format from SAM format with the use of SAMtools.[51] 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).[52] Beagle[53] 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[54] 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.[55] 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

Table 3. Candidate genes putatively selected affecting phenotypes in cattle

Gene	Description	Chr.	F_{ST}	Traits	Reference
Minnan cattle					
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	[42]
HSPA4	heat shock 70kDa protein 4	7	0.6965	adaptability	[15]
GPR50	G protein-coupled receptor 50	X	0.7483	adaptability	[43]
FGF9	fibroblast growth factor 9	12	0.6802	adaptability	[44]
MON1A	MON1 homolog A, secretory trafficking associated	22	0.8486	parasite resistance	[39]
MST1R	macrophage stimulating 1 receptor	22	0.7560	parasite resistance	[39]
DNAH2	dynein axonemal heavy chain 2	19	0.7540	parasite resistance	[38,45]
LCT	lactase	2	0.6614	parasite resistance	[33]
THPO	thrombopoietin	1	0.7626	parasite resistance	[11]
SAR1B	secretion associated Ras related GTPase 1B	7	0.7129	immune response	[26]
IFNAR2	interferon alpha and beta receptor subunit 2	1	0.6516	immune response	[11]
HSPA9	heat shock protein family A member 9	7	0.6390	immune response	[46]
CD5	CD5 molecule	29	0.6496	immune response	[32]
VAMP7	vesicle-associated membrane protein 7	X	0.8344	immune response	[47]
IL15	interleukin 15	17	0.6597	immune response	[11]
FOXC2	forkhead box C2	18	0.7322	disease resistance	[32]
TPM2	tropomyosin 2	8	0.6497	disease resistance	[32]
CNGB1	cyclic nucleotide gated channel beta 1	18	0.6554	feed efficiency	[43]
R3HDM1	R3H domain containing 1	2	0.6497	feed efficiency	[22]
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta	1	0.7196	feed efficiency	[16]
FGF1	fibroblast growth factor 1	7	0.6860	reproduction	[46]
ADRB2	adrenoceptor beta 2	7	0.7676	meat traits	[14]
PDHB	pyruvate dehydrogenase beta	22	0.7017	meat traits	[32,38]
SLC2A4	solute carrier family 2 member 4	19	0.7437	meat traits	[42]
PCSK4	proprotein convertase subtilisin/kexin type 4	7	0.7427	fertility	[46]
SPOCK1	SPARC/osteonectin, cwcv and kazal like domains proteoglycan 1	7	0.7334	fertility	[46]
CABIN1	calcineurin binding protein 1	17	0.6795	fertility	[34]
SLC38A3	solute carrier family 38 member 3	22	0.8155	meat traits	[48]
EXOC3	exocyst complex component 3	20	0.8345	meat traits	[48]
STAT5B	signal transducer and activator of transcription 5B	19	0.6768	meat traits	[48]
EMD	emerin	X	0.9695	meat traits	[48]
Mongolian cattle					
MC1R	melanocortin 1 receptor	18	0.6532	coat color	[15]
RXFP2	relaxin/insulin like family peptide receptor 2	12	0.7337	horn development	[42,46]
CDHR4	cadherin related family member 4	22	0.6483	parasite resistance	[39]
IP6K1	inositol hexakisphosphate kinase 1	22	0.7258	parasite resistance	[39]
RNF123	ring finger protein 123	22	0.7129	parasite resistance	[39]
LEMD3	LEM domain containing 3	5	0.6871	growth, meat traits	[42]
NCOA1	nuclear receptor coactivator 1	11	0.6699	muscular development, reproduction	[42,43]
APOF	apolipoprotein F	5	0.7752	meat traits	[43]

Table 3. Continued

Gene	Description	Chr.	F_{ST}	Traits	Reference
HEYL	hes related family bHLH transcription factor with YRPW motif-like	3	0.6418	meat traits	[48]
SLC29A1	solute carrier family 29 member 1	23	0.7452	meat traits	[48]
TFIP11	tuftelin interacting protein 11	17	0.6634	meat traits	[48]
ASNS	asparagine synthetase	4	0.6548	meat traits	[48]
RPL15	ribosomal protein L15	27	0.7344	meat traits	[48]
RELA	RELA proto-oncogene, NF-kB subunit	29	0.6763	dairy traits	[48]
RDH5	retinol dehydrogenase 5	5	0.7020	meat traits	[48]
ITPR3	inositol 1,4,5-trisphosphate receptor type 3	23	0.6633	meat traits	[48]

"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 refseq, 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 Haploview[59] 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 ($\theta\pi$) 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 $\theta\pi$ 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 ($\theta\pi$) and Hp value of each breed through VCFtools software with default parameters and 50 kb sliding windows in 10 kb steps.[61] The $\theta\pi$ ratio between group MN and MG was calculated as $\ln(\theta\pi,MN/\theta\pi,MG)$, 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.

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.R.M., Sousa-Canavez, J.M., 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., Utsunomiya, Y.T., Garcia, J.F., Gondro, 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., Alanazi, 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., Garcia, J.F., Sonstegard, T.S., Van Tassell, C.P., et al. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Res. 23, 253–262.

9. Randhawa, I.A.S., Khatkar, M.S., Thomson, P.C., and Raadsma, H.W. (2016). A Meta-Assembly of Selection Signatures in Cattle. PLoS ONE 11, e0153013.

10. Mengjie, W., and Yi, D. (1996). The importance of work animals in rural China. World Anim. Rev. 86, 65–67.

11. Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L., Guigó, R., et al.; Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324, 522–528.

12. Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., et al. (2009). A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10, R42.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

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

22. Gutiérrez-Gil, B., Arranz, J.J., and Wiener, P. (2015). An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Front. Genet. 6, 167.

23. Leonard, S., Khatib, H., Schutzkus, V., Chang, Y.M., and Maltecca, C. (2005). Effects of the osteopontin gene variants on milk production traits in dairy cattle. J. Dairy Sci. 88, 4083–4086.

24. Sheehy, P.A., Riley, L.G., Raadsma, H.W., Williamson, P., and Wynn, P.C. (2009). A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. Anim. Genet. 40, 492–498.

25. Puigserver, P., and Spiegelman, B.M. (2003). Peroxisome proliferator-activated receptor-γ coactivator 1 α (PGC-1 α): transcriptional coactivator and metabolic regulator. Endocr. Rev. 24, 78–90.

26. Xu, L., Bickhart, D.M., Cole, J.B., Schroeder, S.G., Song, J., Tassell, C.P., Sonstegard, T.S., and Liu, G.E. (2015). Genomic signatures reveal new evidences for selection of important traits in domestic cattle. Mol. Biol. Evol. 32, 711–725.

27. Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G., and Arranz, J.J. (2017). Variant discovery in the sheep milk transcriptome using RNA sequencing. BMC Genomics 18, 170.

28. Gan, Q.-F., Zhang, L.-P., Li, J.-Y., Hou, G.-Y., Li, H.D., Gao, X., Ren, H.Y., Chen, J.B., and Xu, S.Z. (2008). Association analysis of thyroglobulin gene variants with carcass and meat quality traits in beef cattle. J. Appl. Genet. 49, 251–255.

29. Bennett, G.L., Shackelford, S.D., Wheeler, T.L., King, D.A., Casas, E., and Smith, T.P.L. (2013). Selection for genetic markers in beef cattle reveals complex associations of thyroglobulin and casein1-s1 with carcass and meat traits. J. Anim. Sci. 91, 565–571.

30. Zhang, L., Michal, J.J., O'Fallon, J.V., Pan, Z., Gaskins, C.T., Reeves, J.J., Busboom, J.R., Zhou, X., Ding, B., Dodson, M.V., and Jiang, Z. (2012). Quantitative genomics of 30 complex phenotypes in Wagyu x Angus F1 progeny. Int. J. Biol. Sci. 8, 838–858.

31. Zhang, X., Wang, K., Wang, L., Yang, Y., Ni, Z., Xie, X., Shao, X., Han, J., Wan, D., and Qiu, Q. (2016). Genome-wide patterns of copy number variation in the Chinese yak genome. BMC Genomics 17, 379.

32. Lee, K.-T., Chung, W.-H., Lee, S.-Y., Choi, J.-W., Kim, J., Lim, D., Lee, S., Jang, G.-W., Kim, B., Choy, Y.H., et al. (2013). Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. BMC Genomics 14, 519.

33. Xu, L., Hou, Y., Bickhart, D.M., Zhou, Y., Hay, H.A., Song, J., Sonstegard, T.S., Van Tassell, C.P., and Liu, G.E. (2016). Population-genetic properties of differentiated copy number variations in cattle. Sci. Rep. 6, 23161.

34. Frischknecht, M., Bapst, B., Seefried, F.R., Signer-Hasler, H., Garrick, D., Stricker, C., Fries, R., Russ, I., Sölkner, J., Bieber, A., et al.; Intergenomics Consortium (2017). Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics 18, 910.

35. Chandra, A., van Maldegem, F., Andrews, S., Neuberger, M.S., and Rada, C. (2013). Deficiency in spliceosome-associated factor CTNNBL1 does not affect ongoing cell cycling but delays exit from quiescence and results in embryonic lethality in mice. Cell Cycle 12, 732–742.

36. Chan, E.K.F., Nagaraj, S.H., and Reverter, A. (2010). The evolution of tropical adaptation: comparing taurine and zebu cattle. Anim. Genet. 41, 467–477.

37. Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., Gredler, B., Bapst, B., and Flury, C. (2017). Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49, 83.

38. Sorbolini, S., Marras, G., Gaspa, G., Dimauro, C., Cellesi, M., Valentini, A., and Macciotta, N.P.P. (2015). Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. Genet. Sel. Evol. 47, 52.

39. Smetko, A., Soudre, A., Silbermayr, K., Müller, S., Brem, G., Hanotte, O., Boettcher, P.J., Stella, A., Mészáros, G., Wurzinger, M., et al. (2015). Trypanosomosis: potential driver of selection in African cattle. Front. Genet. 6, 137.

40. Oddgeirsson, O., Simpson, S.P., Morgan, A.L.G., Ross, D.S., and Spooner, R.L. (1988). Relationship between the bovine major histocompatibility complex (BoLA), erythrocyte markers and susceptibility to mastitis in Icelandic cattle. Anim. Genet. 19, 11–16.

41. Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., Zhang, F., Zhang, L., Cui, L., He, W., et al. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat. Genet. *47*, 217–225.

42. Gutiérrez-Gil, B., Esteban-Blanco, C., Wiener, P., Chitneedi, P.K., Suarez-Vega, A., and Arranz, J.-J. (2017). High-resolution analysis of selection sweeps identified between fine-wool Merino and coarse-wool Churra sheep breeds. Genet. Sel. Evol. *49*, 81.

43. Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. Mol. Biol. Evol. *33*, 2801–2814.

44. Zhao, Y.-X., Yang, J., Lv, F.-H., Hu, X.-J., Xie, X.-L., Zhang, M., Li, W.-R., Liu, M.-J., Wang, Y.-T., Li, J.-Q., et al. (2017). Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in East Asia. Mol. Biol. Evol. *34*, 2380–2395.

45. Mustafa, H., Khan, W.A., Kuthu, Z.H., Eui-Soo, K., Ajmal, A., Javed, K., Pasha, T.N., Ali, A., Javed, M.T., and Sonstegard, T.S. (2018). Genome-Wide Survey of Selection Signatures in Pakistani Cattle Breeds. Pak. Vet. J. *38*, 214–218.

46. Randhawa, I.A., Khatkar, M.S., Thomson, P.C., and Raadsma, H.W. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BMC Genet. *15*, 34.

47. Kizilkaya, K., Tait, R.G., Garrick, D.J., Fernando, R.L., and Reecy, J.M. (2013). Genome-wide association study of infectious bovine keratoconjunctivitis in Angus cattle. BMC Genet. *14*, 23.

48. Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., Oda, S., Ebihara, S., Yajima, S., Yoshikawa, H., and Kono, T. (2011). Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. BMC Genomics *12*, 103.

49. Mei, C.-G., Gui, L.-S., Wang, H.-C., Tian, W.-Q., Li, Y.-K., and Zan, L.-S. (2018). Polymorphisms in adrenergic receptor genes in Qinchuan cattle show associations with selected carcass traits. Meat Sci. *135*, 166–173.

50. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

52. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

53. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.

54. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. *21*, 974–984.

55. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

56. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

57. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

58. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

59. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263–265.

60. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.

61. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

62. Hu, Z.-L., Park, C.A., and Reecy, J.M. (2016). Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res. *44* (D1), D827–D833.

63. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. *4*, 44–57.