Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome Resequencing

Chugang Mei,[†,1] Hongcheng Wang,[†,1] Qijun Liao,[†,2] Lizhong Wang,[†,2] Gong Cheng,[1] Hongbao Wang,[1] Chunping Zhao,[1] Shancen Zhao,[2] Jiuzhou Song,[3] Xuanmin Guang,[2] George E. Liu,[4] Anning Li,[1] Xueli Wu,[2] Chongzhi Wang,[2] Xiaodong Fang,[2] Xin Zhao,[1,5] Stephen B. Smith,[6] Wucai Yang,[1] Wanqiang Tian,[7] Linsheng Gui,[1] Yingying Zhang,[1] Rodney A. Hill,[8] Zhongliang Jiang,[1] Yaping Xin, [1] Cunling Jia,[1] Xiuzhu Sun,[1] Shuhui Wang,[1] Huanming Yang,[9,10] Jian Wang,[9,10] Wenjuan Zhu,[*,2] and Linsen Zan[*,1]

[1]College of Animal Science and Technology, Northwest A&F University, Yangling, China

[2]BGI Genomics, BGI-Shenzhen, Shenzhen, China

[3]Department of Animal and Avian Sciences, University of Maryland, Maryland, USA

[4]Animal Genomics and Improvement Laboratory, USDA-ARS, Maryland, USA

[5]Department of Animal Science, McGill University, Montreal, Canada

[6]Department of Animal Science, Texas A&M University, Texas, USA

[7]Yangling Vocational & Technical College, Yangling, China

[8]School of Biomedical Sciences, Charles Sturt University, New South Wales, Australia

[9]BGI-Shenzhen, Shenzhen, China

[10]James D. Watson Institute of Genome Sciences, Hangzhou, China

[†]These authors contributed equally to this work.

The sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession PRJNA283480.

***Corresponding authors:** E-mails: zanlinsen@163.com; wenjuan.zhu@genomics.cn.

Associate editor: Bing Su

Abstract

The bovine genetic resources in China are diverse, but their value and potential are yet to be discovered. To determine the genetic diversity and population structure of Chinese cattle, we analyzed the whole genomes of 46 cattle from six phenotypically and geographically representative Chinese cattle breeds, together with 18 Red Angus cattle genomes, 11 Japanese black cattle genomes and taurine and indicine genomes available from previous studies. Our results showed that Chinese cattle originated from hybridization between *Bos taurus* and *Bos indicus*. Moreover, we found that the level of genetic variation in Chinese cattle depends upon the degree of indicine content. We also discovered many potential selective sweep regions associated with domestication related to breed-specific characteristics, with selective sweep regions including genes associated with coat color (*ERCC2*, *MC1R*, *ZBTB17*, and *MAP2K1*), dairy traits (*NCAPG*, *MAPK7*, *FST*, *ITFG1*, *SETMAR*, *PAG1*, *CSN3*, and *RPL37A*), and meat production/quality traits (such as *BBS2*, *R3HDM1*, *IGFBP2*, *IGFBP5*, *MYH9*, *MYH4*, and *MC5R*). These findings substantially expand the catalogue of genetic variants in cattle and reveal new insights into the evolutionary history and domestication traits of Chinese cattle.

Key words: **whole genome sequencing, selection, Chinese cattle, indicine components, admixture.**

Article

Introduction

Domesticated extant cattle can be categorized into two major geographic taxa: humpless taurine (*B. taurus*) and humped indicine (*B. indicus*) cattle, which diverged from each other >250,000 years ago (Hiendleder et al. 2008; Gibbs et al. 2009; Canavez et al. 2012; Porto-Neto et al. 2013). According to previous reports, taurine cattle were domesticated in the Fertile Crescent ~8,000–10,000 years ago, and indicine cattle were domesticated in the Indus Valley ~6,000–8,000 years ago (Loftus et al. 1994; Van Vuure 2002; Bickhart et al. 2016). As a representative ruminant, cattle provide hides, meat, and milk for human needs and work as draught animals for pulling carts, ploughing, and other tasks in less mechanized cultures (Sherratt 1983; Zhang et al. 2013). Through artificial selection, >1,000 cattle breeds were established throughout the world (Scherf and Pilling 2015). Of these breeds, 72 breeds originated from and are endemic to China. These Chinese breeds vary in their intrinsic characteristics and are important genetic resources for cattle worldwide. Chinese cattle have long been used as draught animals and are valued for their parasite resistance, utilization of roughage-based diets and tolerance to environmental challenges (Qiu et al. 1993; Wang and Ding 1996). Chinese cattle are roughly divided into three groups according to their ecological characteristics and sex

chromosome polymorphisms: a southern group, largely descended from the indicine lineage; a northern group, belonging to the taurine lineage; and a central group, which originated from *B. taurus* × *B. indicus* hybrids (Qiu et al. 1993; Cai et al. 2006).

High-throughput whole genome sequencing can be used to exploit population structure and characteristics to identify the effects of selection upon the cattle genome in different breeds. This approach has been performed with dairy cattle such as Holstein–Friesian, Fleckvieh, and Jersey populations for traits including embryonic death, lethal chondrodysplasia, milk production, and curly coat (Daetwyler et al. 2014). Studies have also been performed on economic traits with breeds such as Hereford, Black Angus, and Limousin (Gibbs et al. 2009; Stothard et al. 2011). Several studies of traits under positive selection have been performed with many European breeds; however, positive selection signatures in Chinese cattle have yet to be determined. A limited number of phylogenetic studies of Chinese cattle have been performed with Y chromosomal and mitochondrial DNAs (Lei et al. 2000; Cai et al. 2007, 2014). These sequences reflect the histories of individual loci and thus do not have the power to track artificial selection signals, complex histories of introgression, or admixture of genomes. Thus, the population stratification of Chinese cattle and signatures of selection in these breeds remain poorly understood.

In this study, we performed whole-genome sequencing on six phenotypically and geographically diverse domestic Chinese cattle breeds (Qinchuan cattle, QCC, $n = 37$; Nanyang cattle, NYC, $n = 2$; Luxi cattle, LXC, $n = 1$; Yanbian cattle, YBC, $n = 2$; Yunnan cattle, YNC, $n = 2$; and Leiqiong cattle, LQC, $n = 2$), and two non-Chinese breeds (Japanese black cattle, JBC, $n = 11$ and Red Angus cattle, RAN, $n = 18$). Using the obtained whole-genome sequence data together with publicly available whole-genome sequence data for additional seven breeds, we explored the genetic diversity, phylogenetic relationships, and demographic history of Chinese cattle. We also integrated patterns of hybridization and detected genes and corresponding variants that are associated with agriculturally important traits. Our analyses provide new insights into the population stratification and local breeding of Chinese cattle and the interface with worldwide domestic breeds.

Results and Discussion

Whole-Genome Sequencing and Genetic Variation

Whole-genome sequencing of 75 samples generated a total of 27.52 billion paired-end reads with 500-bp insert size. Alignment with the reference genome of *B. taurus* (UMD3.1) showed an average depth of 11.4× and an average coverage of 98.46% (supplementary table S1, Supplementary Material online). To place these cattle in a more detailed phylogeographic context, we also analyzed previously published whole-genome sequence data from individuals of representative taurine and indicine breeds ($n = 76$, supplementary tables S2 and S3 and fig. S1, Supplementary Material online). We detected a total of 57.22 million single-nucleotide polymorphisms (SNPs) and 5.27 million small insertions and deletions (InDels) (supplementary table S3 and fig. S2, Supplementary Material online). More than half (59.90% and 72.45%) of the SNPs and InDels were absent in the SNP Database (dbSNP, release 140); the novel variants, which substantially expanded the set of genetic variants in cattle, were mainly contributed by *B. indicus* and Chinese breeds, especially LQC and QCC (supplementary table S3 and figs. S3 and S4, Supplementary Material online). Rare variants (<1%) captured 37.64% of the data set. Approximately 21.54 million autosomal variants had a minor allele frequency <1%, ~16.38 million had frequencies between 1% and 5%, and ~19.30 million had a frequency >5% (supplementary fig. S3*a*, Supplementary Material online). The most common variants (~80.94% of 19.30 million) with a >5% minor allele frequency were found in the dbSNP database. In contrast, only 30.91% (5.06 million of 16.38 million) of variants were in the range of 1–5% in frequency, and 10.50% (2.26 million of 21.54 million) of variants had frequencies <1%.

Taurine breeds had an average of 5.06 million single-nucleotide variants (SNVs) per sample, 96.2% of which were found in the dbSNP database. Indicine breeds had an average of 11.91 million SNVs per sample, 2.35 times higher than taurine breeds (supplementary table S3, Supplementary Material online). Only 59.03% of the indicine SNVs were found in the database. Specifically, LQC from South China had an average of 16.78 million SNVs with only 48.63% found in the database, which is 3.8 times the number for the taurine breeds (supplementary tables S3 and S4, Supplementary Material online). In addition, 95.85% of the singletons were found in Chinese cattle breeds, especially LQC (supplementary figs. S3*a*, S4, and S5*c*, Supplementary Material online), indicating that the Chinese cattle had high genetic diversity.

Most of the cattle groups experienced population bottlenecks during domestication. Taurine cattle showed a similar level of nucleotide diversity (θ_π, $\sim 1.2 \times 10^{-3}$) as that of yak (Qiu et al. 2015) and giant panda ($\sim 1.3 \times 10^{-3}$) (Zhao et al. 2013). Moreover, they showed higher nucleotide diversity than that estimated for human populations ($\sim 1.0 \times 10^{-3}$) and lower than that of indicine cattle ($\sim 2 \times 10^{-3}$) (supplementary table S3, Supplementary Material online). This low level of variation in taurine cattle was also reflected by the extensive linkage disequilibrium (LD) levels among taurine breeds, especially JBC and JER (supplementary fig. S6*b*, Supplementary Material online), indicating a severe bottleneck in taurine cattle (Gibbs et al. 2009; Stothard et al. 2011; Daetwyler et al. 2014).

Compared with European cattle ($\sim 1 \times 10^{-3}$), Chinese cattle ($2 \times 10^{-3} \sim 4 \times 10^{-3}$) showed relatively high nucleotide diversity. The nucleotide diversity of LQC (4.2×10^{-3}) was approximately two times higher than that of indicine breeds (2×10^{-3}). This distinction between Leiqiong (LQC) and indicine cattle was also apparent from the high level of fixation index (F_{ST}) between them (supplementary table S5, Supplementary Material online). The difference of Chinese cattle from European cattle was also reflected by the values of heterozygosity, haplotype diversity, runs of homozygosity

(ROH), inbreeding coefficients, and identity-by-descent (IBD) (supplementary table S6 and figs. S6*a* and *c* and S7, Supplementary Material online). These findings are consistent with previous studies (Lai et al. 2006; Lei et al. 2006; Decker et al. 2014) in which *B. taurus* × *B. indicus* admixture events in Chinese cattle breeds were found to have significantly increased the genetic diversity of Chinese cattle relative to European taurine cattle.

By comparing the population-specific SNPs among *B. indicus* (including only BRM, GIR, and NEL), LQC, and *B. taurus* (including RAN, JBC, HOL, JER, FLV, and LIM), we found 938, 79, and 4 population-specific nonsynonymous variants (PSNVs) with high variant allele frequency (>0.9), respectively (supplementary table S7, Supplementary Material online). Strikingly, some genes had more than three novel nonsynonymous variants (supplementary tables S8–S10, Supplementary Material online). For example, there were 12, 5, and 4 nonsynonymous mutations in the *SPTBN5*, *RP1L1*, and *GHRHR* genes, respectively, of *B. indicus* and 9, 8, and 5 novel nonsynonymous mutations in the *LOC616720*, *LOC101903496*, and *RNF213* genes, respectively, of LQC (supplementary tables S8 and S10, Supplementary Material online). In the *SPTBN5* gene, there were 14 *B. indicus*-specific missense mutations with high allele frequency (>0.93; supplementary table S8, Supplementary Material online). These missense mutations were only observed in indicine cattle, and only two of them have been found in dbSNP. In humans, *SPTBN5* imparts a certain resistance to the parasite *Plasmodium falciparum* and to enterohemorrhagic *Escherichia coli* (Ruetz et al. 2012; Labrecque et al. 2013). Indicine cattle are more resistant to thermal stress, parasites, and disease than taurine cattle (Hansen 2004; Sartori et al. 2010). This knowledge led us to speculate that the *SPTBN5* gene may contribute to parasite resistance in indicine cattle. Therefore, these genes with specific nonsynonymous variants might have played roles in the formation of the characteristic phenotypes of each breed.

Population Structure and Characterization of Chinese Cattle Breeds

Using yak (*Bos grunniens*) as an outgroup, we explored the phylogenetic relationships among 151 cattle samples based on whole genome SNP data. The resulting neighbor-joining tree supported the clustering of the taurine clade (RAN, JBC, HOL, FLV, LIM, JER, and YBC) and the indicine clade (BRM, NEL, and GIR). NYC, LXC, YNC, and LQC were grouped together near the indicine clade (fig. 1*a*). Interestingly, QCC was situated between these two clades and had many branches connecting to the trunk, consistent with previous studies (Lai et al. 2006; Lei et al. 2006; Decker et al. 2014), suggesting that this breed has two main ancestor components: taurine and indicine. The principle component analysis (PCA) provided similar results, with all of the taurine cattle breeds except JBC and YBC forming a tight cluster clearly separate from indicine cattle and most of the Chinese populations occupied intermediate positions between the two major clusters (fig. 1*b* and supplementary figs. S8 and S9, Supplementary Material online). The PC2 tended to separate populations sampled in East Asia from those of India and Europe. Both the phylogenetic and PCA analyses indicated a heterogeneous nature of Chinese cattle. The genetic influence of *B. taurus* was greater on QCC and YBC than on the other breeds of central and southern China, whereas *B. indicus* contributed more to LQC, YNC, NYC, and LXC than to the remaining breeds. These results are consistent with the hypothesis that Chinese cattle breeds are admixtures of taurine (*B. taurus*) and indicine (*B. indicus*) cattle (Yu et al. 1999).

We used clustering models for estimating ancestral populations setting $K = 2$ through $K = 6$ with ADMIXTURE (Alexander et al. 2009) for all 151 cattle samples (fig. 1*c*). With K changing progressively from 2 to 6, we found that Chinese breeds showed evidence of admixture, with the extent varying among the different breeds. The average ancestry proportions for each of the admixed populations, assuming $K = 2$ ancestral populations, are shown in supplementary table S11, Supplementary Material online. We found a strong association between genetic diversity and indicine descent in Chinese breeds (supplementary fig. S10, Supplementary Material online). Our results indicate that the heterogeneous nature of Chinese cattle mainly originated from hybridization between *B. taurus* and *B. indicus*.

Inference of Population Size from Whole-Genome Sequencing Data

Historical fluctuations in the effective population size (N_e) for all cattle were reconstructed using the Pairwise Sequential Markovian Coalescent (PSMC) model, and two bottlenecks and two expansions were identified for all cattle (fig. 2*a* and supplementary figs. S11–S13, Supplementary Material online). The split time between the ancestor of indicine cattle and the ancestor of taurine cattle occurred ~1.6 Ma, around which time the uplift of the Himalayas (Yuanmu movement, ~1.6 Ma) (Zheng et al. 2002) established geographical isolation. It is highly possible that the habitat of *B. primigenius* was split into two regions (South Asia and South China), resulting in population separation between the ancestor of *B. indicus* and the ancestor of *B. taurus*, consistent with the aforementioned results.

The cattle population declined ~0.8 Ma, at the same time as the three largest Pleistocene glaciations: the Xixiabangma Glaciation (XG, 1.1–0.8 Ma), the Naynayxungla Glaciation (NG, 0.78–0.5 Ma), and the Penultimate Glaciation (0.30–0.13 Ma) (fig. 2*a*). However, after a very short bottleneck in the ancestor of indicine cattle (BRM, GIR, and NEL) ~500,000 years ago, N_e recovered very quickly and reached a peak ~140,000 years ago. In contrast, the ancestor of taurine cattle (RAN, JBC, HOL, FLV, LIM, and JER) experienced a long, stable bottleneck until 70,000 years ago, consistent with their lower genetic diversity relative to that of indicine cattle. Divergence among the ancestors of indicine and taurine cattle may have begun ~0.5 Ma, coinciding with the uplift of the Tibetan–Pamir Plateau, which caused drying and desertification that were dramatically enhanced ~0.5 Ma (Fang et al. 2002). The demographic trajectories of Chinese cattle breeds (LQC, LXC, NYC, QCC, YBC, and YNC) were distinct from those of typical indicine cattle or taurine cattle due to the

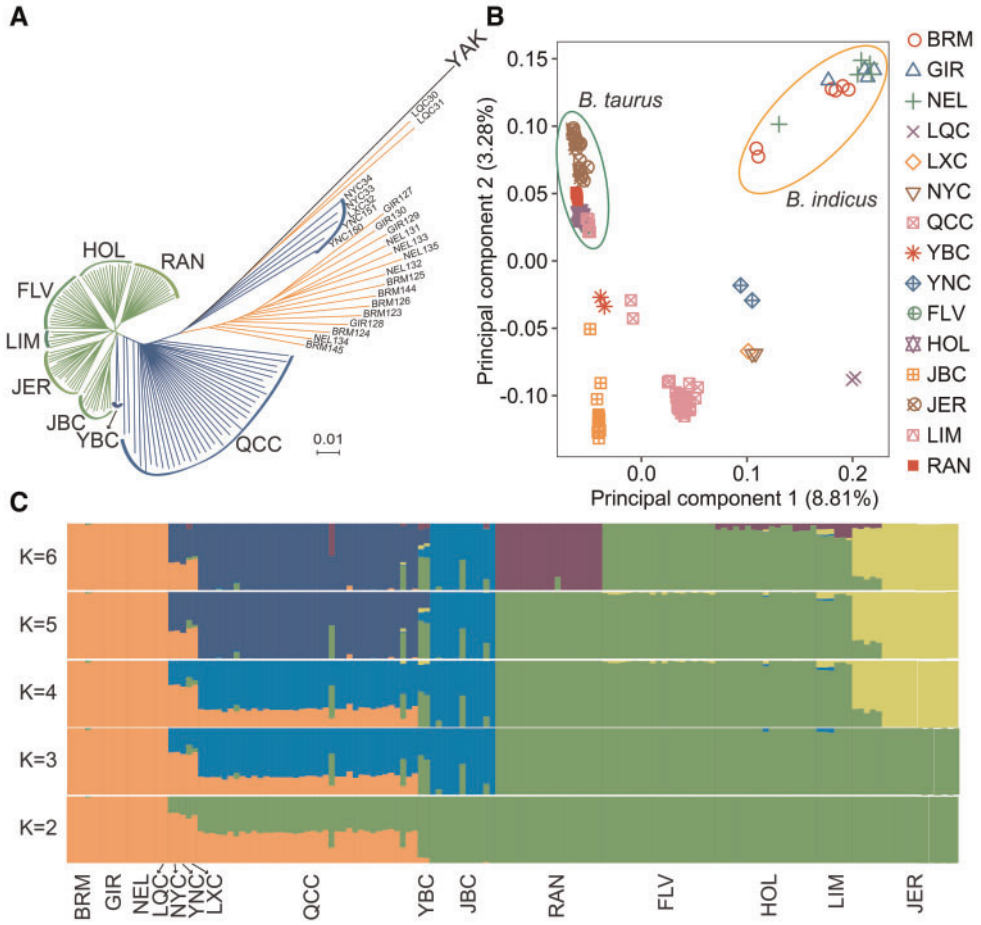

Fig. 1. Population genetic analysis. (*a*) Neighbor-joining tree of indicine, taurine, and hybrid cattle based on our data and publicly available whole-genome sequencing data. Orange branches indicate indicine cattle, green branches represent taurine cattle and blue branches represent hybrid cattle. The scale bar represents the identity-by-state (IBS) score between individuals. (*b*) Principal component analysis of cattle. (*c*) Genetic structure of cattle breeds using the ADMIXTURE program.

influence of *B. taurus* × *B. indicus* admixture events. The historical pattern of four Chinese cattle breeds (LQC, NYC, LXC, and YNC) with more descent contributed by *B. indicus* (>69%, supplementary table S11, Supplementary Material online) roughly correlates with the indicine lineage, but is distinct from QCC and YBC.

It is noteworthy that a historical pattern of two bottlenecks and two expansions has been observed in many mammals, such as yak (Qiu et al. 2015), giant panda (Zhao et al. 2013), wild boar (Choi et al. 2013), snub-nosed monkey (Zhou et al. 2014), gayal (Mei et al. 2016), and bear (Miller et al. 2012). These concordant patterns suggest that terrestrial mammals might share similar demographic histories and that the evolution of terrestrial mammals at the Early-Middle Pleistocene boundary was strongly affected by global glaciations and severely cold climates.

The Multiple Sequentially Markovian Coalescent (MSMC) analysis was used to study the genetic separation between two populations as a function of time by modeling the relationships of multiple haplotypes. For each population split scenario, the relative cross coalescence rate estimates were obtained by dividing the cross-population coalescence rate by the average within-population coalescence rate (Schiffels and Durbin 2014). Based on the analysis of four haplotypes for each pair of populations, the MSMC results show that the beginning of the split between the NEL ancestors and the GIR and BRM ancestors occurred ~11,000 years ago. This split occurred shortly after the Younger Dryas epoch (an abrupt rapid cooling period that occurred 12,800–11,500 years ago; Chen et al. 2006) (fig. 2*b*). The split between the NEL and BRM ancestors occurred ~6,600 years ago. After separation, the N_e of indicine cattle expanded, reaching a peak ~1,500 years ago, whereas the N_e of taurine cattle remained stable due to inbreeding and artificial domestication (supplementary fig. S14, Supplementary Material online). These data suggest that NEL, GIR, and BRM shared the same ancestor and that NEL separated from indicine ancestors earlier than GIR and BRM did. As we observed, the separation was slow and might have been the result of continuous gene exchange among these breeds.

Different from the split among indicine cattle groups, a sharp separation among the ancestors of taurine breeds occurred 5,000–2,000 years ago, with a split time at 3,500 years ago (fig. 2*c*), coinciding with the Unetice culture (4,200–3,500 years ago). Taurine breeds underwent strong domestication in a very short time. We infer that the considerable

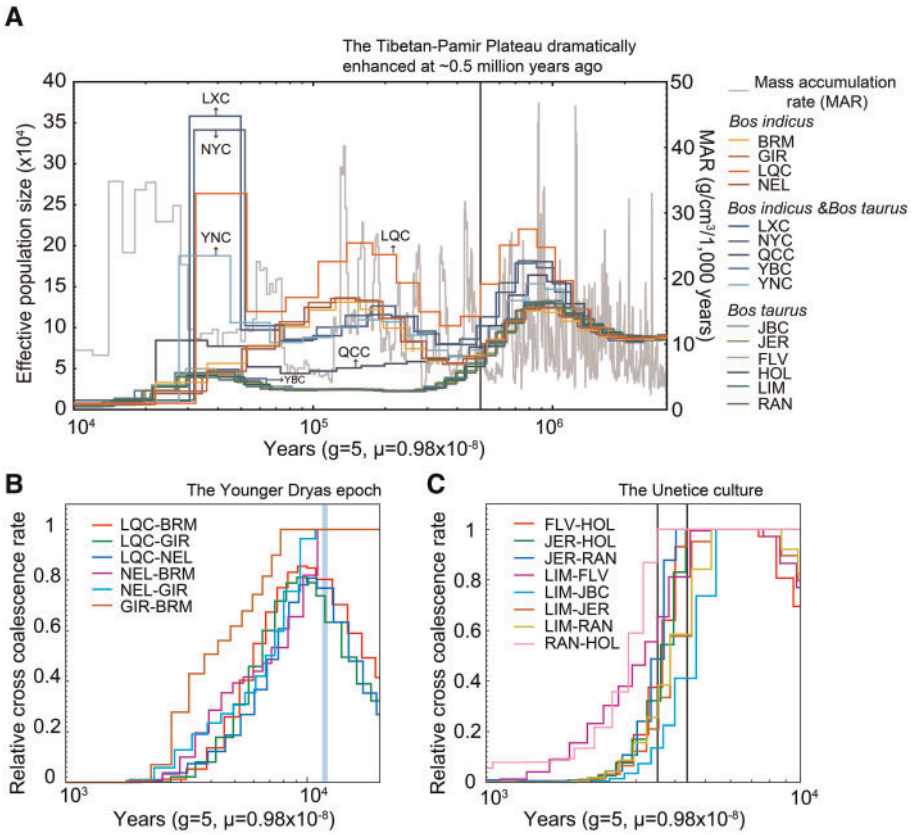

FIG. 2. Demographic history of cattle. (*a*) Ancestral population size is inferred using PSMC. A generation time of 5 years and a mutation rate of 0.98×10^{-8} mutations per nucleotide per generation are used. The relative cross coalescence rates over time between indicine (*b*) and taurine (*c*) breeds are estimated using MSMC, with four haplotypes each pair.

economic prosperity based on the diversified agriculture of the Unetice culture contributed to the early formation of European cattle breeds (Svizzero and Tisdell 2016).

Signatures of Positive Selection in Cattle Genome

We identified regions that exhibit high levels of differentiation among cattle breeds using the d_i statistic in a reduced data set containing breeds with at least 10 samples (fig. 3). We treated GIR, NEL, and BRM as one group (IND) based on their close genetic relationships as evidenced by both the PCA and phylogenetic results (fig. 1*a* and *b*). Those windows with the highest average d_i values within each breed, which fell into the upper 99th percentile of the empirical distribution, were considered putative signatures of selection (supplementary fig. S15, Supplementary Material online).

In total, we identified 2,842 potential selective sweep regions in one or more of the seven breeds (full genomic regions are detailed in supplementary tables S12–S18, Supplementary Material online), which had an average size of 67 kb (ranging from 11 kb to 1,150 kb). These regions harbored 1,429 protein-coding genes, 682 (47.72%) of which were previously identified as under positive selection in cattle (Randhawa et al. 2016). More specifically, we detected 357, 381, 232, 234, 307, 300, and 307 potentially positively selected genes on breed-specific selection events in the IND, QCC, JBC, RAN, FLV, HOL, and JER genomes, respectively (fig. 3*a* and supplementary tables S19–S25, Supplementary Material online).

To obtain a broad overview of the molecular functions of these genes and to test the hypothesis that particular functional classes are enriched in the most differentiated regions of cattle genome, we performed a gene ontology (GO) analysis using ClueGO (Bindea et al. 2009) for each group separately. A potential concern regarding this analysis is the low power to detect enrichment due to the low expected counts for many categories. Nonetheless, several categories showed enrichment for signals of positive selection in one or more groups (supplementary table S26, Supplementary Material online), including the related categories of cellular response to UV as well as immune response and pathogen defence. These findings suggest that immune-related genes are pervasive targets of positive selection because of their critical role in immune and defence functions.

Many genes associated with shaping particular characteristics of the populations are presented within these regions (table 1). These include morphological (coat color, horn/polledness) and production traits (dairy, muscle formation, skeletal development, energy partitioning, fertility, draft traits).

Several genes involved in coat color phenotypes were identified as targets of positive selection in one or more groups (table 1), including *ERCC2* in QCC and IND, *MC1R* in IND, *ZBTB17* in QCC and *MAP2K1* in JBC. One of these genes, *MC1R*, is well known for its role in regulating the switch

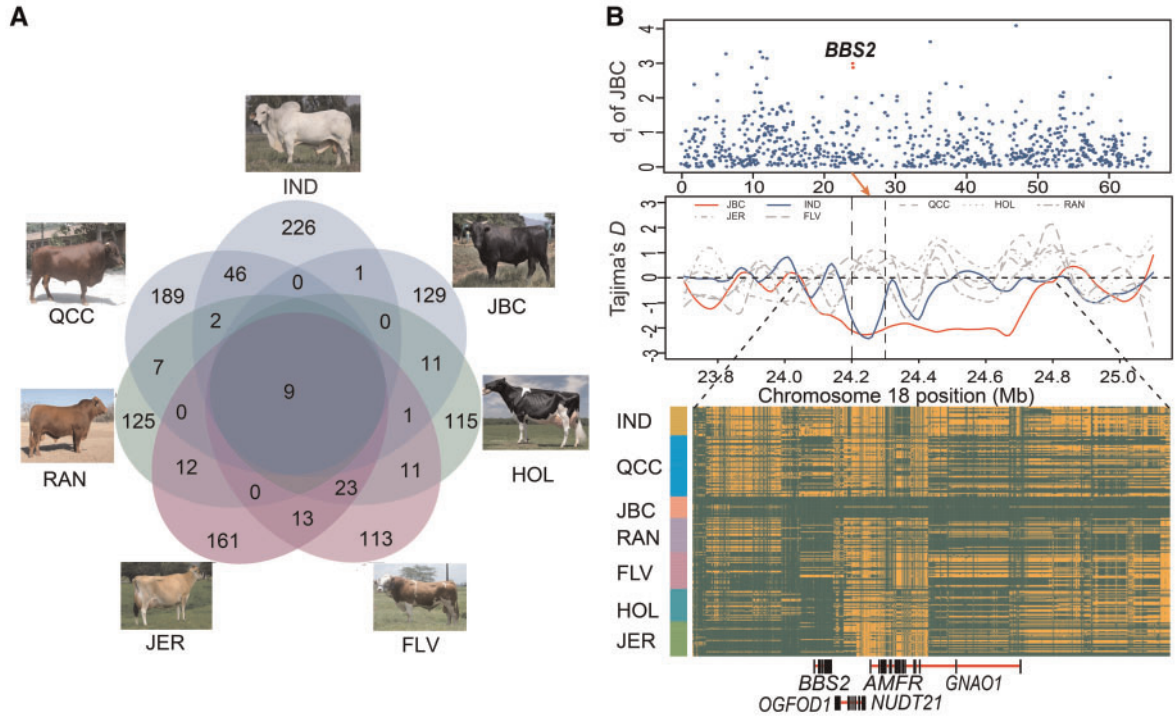

Fig. 3. Candidate positively selected genes. (*a*) Shared candidate positively selected genes among groups. Only partial numbers are shown. (*b*) An example of selective sweep at the *BBS2* gene in JBC; only positive values of d_i are shown (top). The Tajima's *D* values in each group are shown (middle). SNPs with minor allele frequencies > 0.05 are used to construct haplotype patterns (bottom). The major alleles in JBC are green, and the minor alleles in JBC are yellow.

Table 1. Summary of Partial Traits Associated with Positively Selected Genes.

Gene	Breed	Trait	Reference
MC1R	IND	CC	(Lee et al. 2002; Gan et al. 2007)
MAP2K1[a]	JBC	CC	(Gutierrez-Gil et al. 2015)
ZBTB17	QCC	CC	(Gutierrez-Gil et al. 2015)
ERCC2[a]	QCC, IND	CC	(Gutierrez-Gil et al. 2015)
FST, ITFG1, SETMAR, PAG1	HOL	DY	(Bech-Sabat et al. 2008; Rincon et al. 2009; Bloise et al. 2010; Xu et al. 2015)
RPL37A[a], *CSN3*	JER	DY	(Wedholm et al. 2006; Yahvah et al. 2015)
MAPK7[a]	FLV	DY	(Lin et al. 2013)
NCAPG	FLV, JER	DY	(Eberlein et al. 2009; Setoguchi et al. 2011)
BBS2[a], *S1PR3*[a], *LRP2BP*[a], *IGFBP2, IGFBP5, MYH9, ASGR1*	JBC	MT, GT	(Forti et al. 2007; Sattler and Levkau 2009; Zhang et al. 2012; Lee et al. 2013; Sorbolini et al. 2015; Yoon and Ko 2016)
SRPK3[a], *POLDIP2*[a], *SLC2A5, TMEM97, MYH4,*	RAN	MT, GT	(Smith et al. 2001; Clark et al. 2011; Xu et al. 2011; Zhao et al. 2012; Lee et al. 2013; Zhang et al. 2016)
CASP9[a], *DIO1, SREBF2, PLOD3*	QCC	MT, GT	(Ouali et al. 2006; Lee et al. 2013)
SLC2A4[a], *OSTN, CPT2, CSF2RB*	FLV	MT, GT	(Grindflek et al. 2002; Lee et al. 2013; Xu et al. 2015)
MC5R[a]	IND	MT, GT	(Kováčik et al. 2012; Switonski et al. 2013)
AOX1[a]	QCC, RAN, FLV, JER, and IND	MT, GT	(Brandes et al. 1995)
R3HDM1	QCC, FLV, HOL, JER, and IND	MT, FC	(Gibbs et al. 2009)

CC, coat color; MT, meat traits; GT, growth traits; DY, dairy traits; FC, food conversion efficiency.
[a]Newly identified genes associated with phenotypic features of cattle.

between eumelanin and pheomelanin biosynthesis pathways in mammals, including cattle. The selection signals of IND in *MC1R* represent the significant role of light coloration (light gray to white in BRM and NEL, yellowish-red to white in GIR) associated with the adaptation of IND to its tropical environment.

Some of the strongest signals of selection appeared in various types of genes related to production traits (table 1). For example, several genes involved in milk production showed clear evidence of positive selection in dairy cattle (*NCAPG* in both FLV and JER; *MAPK7* in FLV; *FST*, *ITFG1*, *SETMAR*, and *PAG1* in HOL; *CSN3* and *RPL37A* in JER). Various genes involved in meat traits have also been targets of recent positive selection (table 1). Some genes related to skeletal muscle development and muscle fibre type appeared to be targets of positive selection, including *CASP9*, *DIO1*,

SREBF2, and *PLOD3* in QCC, *ASGR1*, *IGFBP2*, *IGFBP5*, and *MYH9* in JBC, *OSTN*, *CPT2*, *CSF2RB*, and *SLC2A4* in FLV, and *SLC2A5*, *TMEM97*, *MYH4*, *SRPK3*, and *POLDIP2* in RAN. We also found that a set of important genes associated with lipid metabolism were putatively positively selected (*AOX1* in QCC, RAN, FLV, JER, and IND; *MC5R* in IND; *BBS2*, *S1PR3*, and *LRP2BP* in JBC). Interestingly, we identified a missense mutation in *BBS2* (exon15 rs135889003, c.A1880G, p.Q627R) that was almost fixed (allele frequency > 0.95) in JBC, a breed known for producing the intensely marbled Wagyu beef (with >30% intramuscular fat of beef) (Gotoh et al. 2014). *BBS2* is a member of the Bardet–Biedl syndrome gene family, the primary clinical feature of which is obesity, and has been found to play a significant role in adipogenesis (Forti et al. 2007). The positive selection signals near the *BBS2* region are further confirmed by significantly lower values of Tajima's *D* and the long haplotype patterns in JBC (fig. 3*b*), which may be useful as a genetic target for breeding selection for beef marbling improvement. In addition, *R3HDM1*, a gene associated with efficient food conversion and intramuscular fat content, showed signals of positive selection in five groups (QCC, FLV, HOL, JER, and IND). These genes might be associated with the tenderness and quality of meat in cattle.

Conclusion

Whole-genome sequencing of representative Chinese cattle breeds and two additional breeds (JBC and RAN) generated a comprehensive catalogue of genetic variations. This is the first population genomic study on Chinese cattle to use next-generation whole-genome sequencing data and is an important source of genetic information for cattle worldwide. Bovine haplotypes have been inferred in Mongolian yaks, with recent admixture at least 1,500 years ago (Medugorac et al. 2017). It is highly possible that there was recent introgression from yak (*B. grunniens*) to Chinese cattle, as suggested by previous studies (Lei et al. 2000; Cai et al. 2007, 2014). The genetic influence of yak is too limited to have been detected in the representative cattle breeds examined in our study. We also discovered many potential selective sweeps associated with domestication related to breed-specific characteristics, with selective sweep regions including genes associated with coat color, dairy traits, and meat production/quality traits. Collectively, these findings substantially expand the catalogue of genetic variants in cattle and reveal new insights into the evolutionary history and domestication traits of Chinese cattle.

Materials and Methods

Sample Collection and Sequencing

To represent the overall genetic diversity of Chinese cattle, we selected 46 samples from 6 representative Chinese cattle breeds with divergent phenotypic characters across the main geographic distribution: Qinchuan cattle (QCC, $n = 37$), Nanyang cattle (NYC, $n = 2$), Luxi cattle (LXC, $n = 1$), Yanbian cattle (YBC, $n = 2$), Yunnan cattle (YNC, $n = 2$), and Leiqiong cattle (LQC, $n = 2$). For comparison, samples from two specialized beef cattle breeds, Red Angus (RAN, $n = 18$), and JBC ($n = 11$), were also collected (supplementary table S2, Supplementary Material online). Total genomic DNA was extracted from the blood samples of the animals using a standard phenol–chloroform protocol. For each individual, at least 5-μg genomic DNA was used to construct paired-end libraries with an insert size of 500 bp according to the Illumina's library preparation protocol. Moreover, we collected 76 genome sequences from previous studies for the breeds Brahman (BRM, indicine, $n = 6$), Nelore (NEL, indicine, $n = 5$), Gir (GIR, indicine, $n = 4$), Limousin (LIM, taurine, $n = 6$), Jersey (JER, taurine, $n = 18$), Fleckvieh (FLV, taurine, $n = 19$), and Holstein (HOL, taurine, $n = 18$) (details in supplementary tables S1 and S2, Supplementary Material online).

Alignments and Variant Identification

Paired-end reads (100 bp) obtained from sequencing in the present study and previous studies were mapped to the *B. taurus* genome (UMD3.1) (Zimin et al. 2009) using BWA (Li and Durbin 2009) with the default parameters. Sequence Alignment Map (SAM) format files were imported into SAMtools (Li et al. 2009) for sorting and merging and into Picard (http://broadinstitute.github.io/picard/, version 1.92) to remove duplicated reads. To identify the ancestral state of cattle, we mapped the raw reads of yak (Qiu et al. 2012), sequenced to 65×, to the reference genome.

Initial variant site identification was performed using SAMtools mpileup and GATK UnifiedGenotyper (Genome Analysis Toolkit, version 2.4-9) (McKenna et al. 2010) with the default settings. The overlap subset of 53,979,675 single-nucleotide polymorphisms (SNPs) and 5,924,578 small insertions and deletions (InDels; 91% of InDels were 1–30 bp in length, and the largest InDel was 403 bp in length) was defined as a high-confidence catalogue used for base quality recalibration using GATK with the default set of covariants. The resulting recalibrated bam files were then used as input for a second variant calling with GATK. The resulting variant calls were analyzed, and approximately the highest scoring 10% of the predicted variant sites were used as a training set for variant quality recalibration and filtering by using GATK. These steps resulted in 60,031,459 SNPs and 5,603,383 InDels. To obtain high-quality results for further analyses, we only retained biallelic SNPs and InDels with >90% calling rates, resulting in 57,220,105 SNPs and 5,270,518 InDels. Beagle (Browning and Browning 2007), which has been shown to yield highly accurate solutions, was used to improve the genotype calls using genotype likelihoods from GATK and to infer the haplotypes in the sample. Short InDels were not included in the diversity or divergence estimates and were not included in the other analyses. Variants (SNPs and InDels) were annotated using ANNOVAR (Wang et al. 2010).

Phylogenetic and Population Structure Analyses

A phylogenetic tree was constructed from the SNP data by using the neighbor-joining method in the program PHYLIP v3.695 (http://evolution.genetics.washington.edu/phylip.html), and distance matrices were calculated using PLINK (Purcell et al. 2007). The ancestral states of the SNPs were

determined using a close relative of cattle, *B. grunniens*, as the outgroup. Population structure was further inferred using ADMIXTURE (Alexander et al. 2009) with kinship (K) set from 2 to 7. Principle component analysis was carried out using the smartPCA program of the EIGENSOFT (Patterson et al. 2006) package.

Genome-Wide Patterns of Genetic Diversity and Divergence

The average pairwise nucleotide diversity (θ_π) and Tajima's *D* statistic of each breed were calculated using a sliding window approach (50-kb sliding windows in 10-kb steps) with the default parameters of VCFtools (Danecek et al. 2011). Population differentiation was measured by pairwise F_{ST} using the unbiased estimator of Weir and Cockerham (1984) with the default parameters.

Linkage Disequilibrium

To estimate the genome-wide LD of each breed, we calculated the mean r^2 values for pairwise markers with Haploview (Barrett et al. 2005) software. Only SNPs with a minor allele frequency >0.05 in three groups (Chinese cattle, indicine, and taurine) were used. The parameters of Haploview were set to "-maxdistance 200 -dprime -memory 5000 -minGeno 0.6 -minMAF 0.05 -hwcutoff 0.001." To minimize the influence of sample size, only breeds with at least five individuals were used, and breeds with more than five samples were down-sampled to five.

Haplotype Diversity

For the haplotype diversity analysis, the same breeds and SNP set were used as in the linkage disequilibrium analysis. To calculate haplotype diversity, the genome was divided into 5- to 500-kb bins (detailed in supplementary table S6, Supplementary Material online). Windows with fewer than two SNPs per 5 kb were removed, and those with more than four SNPs, four SNPs were randomly selected. Considering the substantial variation in the recombination rate across the cattle genome, we adopted a sliding-window strategy and allowed the window to slide by half its length each time. The frequencies of haplotypes were counted, and haplotype diversity (H) was calculated as described previously (Daetwyler et al. 2014).

PSMC Analysis

We inferred the demographic history of *B. taurus* and *B. indicus* using the Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011). In the default PSMC approach, a whole genome diploid consensus sequence was generated using the alignment file from one sample. Recalling that most of our genomes have not been sequenced to a high average depth of coverage (mostly ~10×) and that PSMC has high false-negative rates at low depths of coverage (i.e., <20×) leading to a systematic underestimation of true event times (Orlando et al. 2013; Nadachowska-Brzyska et al. 2016), we applied a modified PSMC approach: the SNPs of one sample were extracted from variants called on cohorts of all samples and converted to consensus sequences. This procedure was followed for samples (marked in supplementary table S1, Supplementary Material online) with relatively high sequencing depth in each breed to ensure the quality of consensus sequences. We then transformed the consensus sequence into a fasta-like format using "fq2psmcfa." The PSMC parameters were set as follows: "-p 4 + 25*2 + 4 + 6." The mutation rate per generation per site was estimated as: $\mu = D \times g/2\,T$, where *D* is the observed frequency of pairwise differences between two species, *T* is the estimated divergence time, and *g* is the estimated generation time for the two species. The cattle generation time (g) was set to an estimate of 5 years and the estimated divergence time was set to 4.9 Ma based on a previous study on cattle and yak (Qiu et al. 2012). These values yielded an estimated mutation rate of 9.796×10^{-9} mutations per generation per site. We obtained mass accumulation rate (MAR) of Chinese loess of the past 3.6 My (Sun and An 2005), an index indicating cold and dry or warm and wet climatic periods in China (fig. 2*a* and supplementary figs. S13 and S14, Supplementary Material online).

To evaluate the differences between our revised PSMC approach and the default method, we reconstructed trajectories from two samples with different depth of coverage (SRR1262805 with 24× and SRR1262808 with 9×) of the same breed (FLV) which should yield similar inferences. The PSMC profiles retrieved from the default and revised approach of the high depth sample were found to be almost identical (supplementary fig. S11, Supplementary Material online), both with regarding to the timing and the magnitude of demographic events, except for the most recent expansion phase, in which a lower intensity was found using the revised approach. We found that PSMC inference based on the low depth sample showed a biased demographic model and could be satisfactorily corrected with our revised PSMC approach. Additionally, we note that the detected bias observed for genomes with low depth (<20×) could also be corrected assuming a uniform False Negative Rate (uNFR) by using the option "–M" of the plotting script "psmc_plot.pl" to specify the uFNR correction rate (Orlando et al. 2013; Hung et al. 2014). The uFNR correction showed a similar plot of a low depth sample compared with high depth PSMC inference (supplementary fig. S11, Supplementary Material online). No striking differences were observed among the PSMC profiles reconstructed from different taurine breeds with different sequencing depth of coverage (range from 9× to 24×, supplementary table S1, Supplementary Material online and fig. 2*a*). Consequently, we found our revised approach to be a suitable method that introduced acceptable new biases to estimate the PSMC inference of low average sequencing depth samples.

To explore the potential impact of the reference genome on the PSMC results of indicine breeds, we mapped sequence reads of indicine samples against the assembly of *B. indicus* (Nelore breed, GenBank assembly accession: GCF_000247795.1) and repeated the PSMC analysis (default setting with uFNR correction). Although the PSMC profiles reconstructed from different references were not identical, the qualitative results hold for indicine breeds

with the *B. indicus* reference genome (fig. 2*a* and supplementary fig. S12, Supplementary Material online).

MSMC Analysis

The Multiple Sequential Coalescent Markovian (MSMC) model (Schiffels and Durbin 2014) was used to infer changes in effective population size (N_e) and divergence time between breeds (samples marked in supplementary table S1, Supplementary Material online). MSMC is an extension of the PSMC model, which uses a hidden Markov model to scan genomes and analyze patterns of heterozygosity, with long DNA segments with low heterozygosity reflecting recent coalescent evens. The rate of coalescent events is then used to estimate N_e at a given time. To scale the output of MSMC to real-time population sizes, we used the generation time and mutation rate mentioned earlier (description of PSMC analysis). We obtained atmospheric surface air temperature (SAT) and global sea level (GSL) data of the past 3 My (Bintanja and van de Wal 2008) (supplementary figs. S13 and S14, Supplementary Material online).

Selective Sweep Analysis

Considering the sample size and close genetic background of indicine cattle (NEL, BRM, and GIR), we pooled these three indicine breeds into one group (IND) in our selection analysis. Seven cattle groups (QCC, RAN, JBC, IND, HOL, FLV, and JER) with sample sizes >10 were retained for the following analysis. To identify candidate loci for breed-specific phenotypes that are known to be under positive selection, we used the d_i statistic (Akey et al. 2010) to measure the locus-specific divergences in allele frequency for each group based on unbiased estimates of pairwise F_{ST}. Briefly, for each SNP, we calculated the statistic $d_i = \sum_{j \neq i} \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{sd[F_{ST}^{ij}]}$, where $E[F_{ST}^{ij}]$ and $sd[F_{ST}^{ij}]$ denote the expected value and SD of F_{ST} between group *i* and *j* calculated from all SNPs. For each group, d_i was averaged over the SNPs in nonoverlapping 50-kb windows. Windows with SNP number <10 were removed. The top 1% of windows with highest mean d_i score were defined as candidate selective sweep regions. Adjacent sweeps within a distance of 50 kb were merged into one sweep. Selective sweep regions were annotated with cattle QTLdb release 29 from the Animal Quantitative Trait Loci Database (Hu et al. 2016). Candidate genes under positive selection were defined as those in which more than half of the gene interval was found in selective sweep regions. Tajima's *D* statistic was computed by using VCFtools for each candidate gene. Gene Ontology (GO) enrichment analysis for genes in selective sweep regions was performed with a hypergeometric test using ClueGO (Bindea et al. 2009). The false discovery rate (FDR) was used to correct the *P* values with the Benjamini–Hochberg approach.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Author Contributions

L.-S.Z., W.-J.Z., G.C., and H.-B.W. led the experiments and designed the analytical strategy. L.-S.Z., C.-G.M., H.-C.W., W.-Q.T., L.-S.G., Y.-Y.Z. Z.-L.J., Y.-P.X., and X.-Z.S. performed animal work and prepared biological samples. C.-G.M., H.-C.W., G.C., H.-B.W., C.-P.Z., A.-N.L., W.-C.Y., C.-L.J., and S.-H.W. constructed the DNA library and performed sequencing. W.-J.Z., Q.-J.L., L.-Z.W., X.-L.W., X.-M.G., and C.-Z.W. detected, annotated, and summarized up variants. W.-J.Z., Q.-J.L., C.-G.M., and H.-C.W. performed selection analysis. W.-J.Z., L.-Z.W., C.-G.M., and H.-C.W. analyzed origination of China indicine cattle and population history. C.-G.M., H.-C.W., W.-J.Z., Q.-J.L., and L.-Z.W. wrote the manuscript. L.-S.Z., S.-C.Z., J.-Z.S., G.L., X.-D.F., X.Z., S.S. H.-M.Y., J.W., and R.H. revised the manuscript. All the authors reviewed and approved the final manuscript.

Acknowledgments

We thank many people not listed as authors who provided feedback, samples, and encouragement, especially Changguo Yan, Yimin Xu, Shanzhai Liu, Guanli Wang, Xiang Gao, Jianghong Wan, and Kaixing Qu. This work was supported by the National 863 Program of China (2013AA102505), the National Science-technology Support Plan Projects (2015BAD03B04), the Program of National Beef Cattle and Yak Industrial Technology System (CARS-37), the Technical Innovation Engineering Project of Shaanxi Province (2014KTZB02-02-01), the National Beef Cattle Improvement Center, the National & Local Joint Engineering Research Center for Modern Cattle Biotechnology and Application, the Beef Cattle Engineering Technology Research Center of Shaanxi Province, and the State Key Laboratory of Agricultural Genomics.

References

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A*. 107(3):1160–1165.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19(9):1655–1664.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.

Bech-Sabat G, Lopez-Gatius F, Yaniz JL, Garcia-Ispierto I, Santolaria P, Serrano B, Sulon J, de Sousa NM, Beckers JF. 2008. Factors affecting plasma progesterone in the early fetal period in high producing dairy cows. *Theriogenology* 69(4):426–432.

Bickhart DM, Xu L, Hutchison JL, Cole JB, Null DJ, Schroeder SG, Song J, Garcia JF, Sonstegard TS, Van Tassell CP, et al. 2016. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. *DNA Res*. 23(3):253–262.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8):1091–1093.

Bintanja R, van de Wal RSW. 2008. North American ice-sheet dynamics and the onset of 100, 000-year glacial cycles. *Nature* 454(7206):869–872.

Bloise E, Cassali GD, Ferreira MC, Ciarmela P, Petraglia F, Reis FM. 2010. Activin-related proteins in bovine mammary gland: localization and differential expression during gestational development and differentiation. *J Dairy Sci*. 93(10):4592–4601.

Brandes R, Arad R, Bar-Tana J. 1995. Inducers of adipose conversion activate transcription promoted by a peroxisome proliferator's response element in 3T3-L1 cells. *Biochem Pharmacol*. 50(11):1949–1951.

Browning SR, Browning BL. 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(5):1084–1097.

Cai D, Sun Y, Tang Z, Hu S, Li W, Zhao X, Xiang H, Zhou H. 2014. The origins of Chinese domestic cattle as revealed by ancient DNA analysis. *J Archaeol Sci*. 41:423–434.

Cai X, Chen H, Lei C, Wang S, Xue K, Zhang B. 2007. mtDNA diversity and genetic lineages of eighteen cattle breeds from *Bos taurus* and *Bos indicus* in China. *Genetica* 131(2):175–183.

Cai X, Chen H, Wang S, Xue K, Lei C. 2006. Polymorphisms of two Y chromosome microsatellites in Chinese cattle. *Genet Sel Evol*. 38(5):525–534.

Canavez FC, Luche DD, Stothard P, Leite KR, Sousa-Canavez JM, Plastow G, Meidanis J, Souza MA, Feijao P, Moore SS, et al. 2012. Genome sequence and assembly of *Bos indicus*. *J Hered*. 103(3):342–348.

Chen S, Wang Y, Kong X, Liu D, Cheng H, Edwards RL. 2006. A possible Younger Dryas-type event during Asian monsoonal Termination 3. *Sci China D Earth Sci*. 49(9):982–990.

Choi JW, Liao X, Park S, Jeon HJ, Chung WH, Stothard P, Park YS, Lee JK, Lee KT, Kim SH, et al. 2013. Massively parallel sequencing of Chikso (Korean brindle cattle) to discover genome-wide SNPs and InDels. *Mol Cells* 36(3):203–211.

Clark DL, Boler DD, Kutzler LW, Jones KA, McKeith FK, Killefer J, Carr TR, Dilger AC. 2011. Muscle gene expression associated with increased marbling in beef cattle. *Anim Biotechnol*. 22(2):51–63.

Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, Liao X, Djari A, Rodriguez SC, Grohs C, et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 46(8):858–865.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15): 2156–2158.

Decker JE, McKay SD, Rolf MM, Kim J, Molina AA, Sonstegard TS, Hanotte O, Gotherstrom A, Seabury CM, Praharani L, et al. 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet*. 10(3):e1004254.

Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, Klopp N, Furbass R, Weikard R, Kuhn C. 2009. Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (*NCAPG*) gene. *Genetics* 183(3):951–964.

Fang X, Lü L, Yang S, Li J, An Z, Jiang PA, Chen X. 2002. Loess in Kunlun Mountains and its implications on desert development and Tibetan Plateau uplift in west China. *Sci China D Earth Sci*. 45(4):289–299.

Forti E, Aksanov O, Birk RZ. 2007. Temporal expression pattern of Bardet-Biedl syndrome genes in adipogenesis. *Int J Biochem Cell B*. 39(5):1055–1062.

Gan HY, Li JB, Wang HM, Gao YD, Liu WH, Li JP, Zhong JF. 2007. Relationship between the melanocortin receptor 1 (*MC1R*) gene and the coat color phenotype in cattle. *Yi Chuan* 29:195–200.

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, et al. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324(5926):528–532.

Gotoh T, Takahashi H, Nishimura T, Kuchida K, Mannen H. 2014. Meat produced by Japanese Black cattle and Wagyu. *Anim Front*. 4(4):46–54.

Grindflek E, Holzbauer R, Plastow G, Rothschild MF. 2002. Mapping and investigation of the porcine major insulin sensitive glucose transport (*SLC2A4/GLUT4*) gene as a candidate gene for meat quality and carcass traits. *J Anim Breed Genet*. 119(1):47–55.

Gutierrez-Gil B, Arranz JJ, 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.

Hansen PJ. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim Reprod Sci*. 82–83:349–360.

Hiendleder S, Lewalski H, Janke A. 2008. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet Genome Res*. 120(1–2):150–156.

Hu ZL, Park CA, Reecy JM. 2016. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res*. 44(D1):D827–D833.

Hung CM, Shaner PJ, Zink RM, Liu WC, Chu TC, Huang WS, Li SH. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc Natl Acad Sci U S A*. 111(29):10636–10641.

Kováčik A, Bulla J, Trakovická A, Zltný J, Rafayová A. 2012. The effect of the porcine melanocortin-5 receptor (*MC5R*) gene associated with feed intake, carcass and physico-chemical characteristics. *J Microbiol Biotechnol Food Sci*. 1:498–506.

Labrecque R, Vigneault C, Blondin P, Sirard MA. 2013. Gene expression analysis of bovine oocytes with high developmental competence obtained from FSH-stimulated animals. *Mol Reprod Dev*. 80(6):428–440.

Lai SJ, Liu YP, Liu YX, Li XW, Yao YG. 2006. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. *Mol Phylogenet Evol*. 38(1):146–154.

Lee KT, Chung WH, Lee SY, Choi JW, Kim J, Lim D, Lee S, Jang GW, Kim B, Choy YH, et al. 2013. Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics* 14:519.

Lee SS, Yang BS, Yang YH, Kang SY, Ko SB, Jung JK, Oh WY, Oh SJ, Kim KI. 2002. Analysis of melanocortin receptor 1 (*MC1R*) genotype in Korean brindle cattle and Korean cattle with dark muzzle. *J Anim Sci Technol*. 44(1):23–30.

Lei C, Chen H, Hu S. 2000. Studies on Y chromosome polymorphism and the origin and classification of Chinese yellow cattle. *Acta Agric Boreali-Occidentalis Sin*. 9:43–47.

Lei CZ, Chen H, Zhang HC, Cai X, Liu RY, Luo LY, Wang CF, Zhang W, Ge QL, Zhang RF, et al. 2006. Origin and phylogeographical structure of Chinese cattle. *Anim Genet*. 37(6):579–582.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357): 493–496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Lin X, Luo J, Zhang L, Zhu J. 2013. MicroRNAs synergistically regulate milk fat synthesis in mammary gland epithelial cells of dairy goats. *Gene Expr*. 16(1):1–13.

Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P. 1994. Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A*. 91(7):2757–2761.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9): 1297–1303.

Medugorac I, Graf A, Grohs C, Rothammer S, Zagdsuren Y, Gladyr E, Zinovieva N, Barbieri J, Seichter D, Russ I, et al. 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat Genet*. 49(3):470–475.

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.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE, et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci U S A*. 109(36):E2382–E2390.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Mol Ecol*. 25(5):1058–1072.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78.

Ouali A, Herrera-Mendez CH, Coulis G, Becila S, Boudjellal A, Aubry L, Sentandreu MA. 2006. Revisiting the conversion of muscle into meat and the underlying mechanisms. *Meat Sci*. 74(1):44–58.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet*. 2(12):e190.

Porto-Neto LR, Sonstegard TS, Liu GE, Bickhart DM, Da SM, Machado MA, Utsunomiya YT, Garcia JF, Gondro C, Van Tassell CP. 2013. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14(1):876.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81(3):559–575.

Qiu H, Ju ZY, Chang ZJ. 1993. A survey of cattle production in China. More attention to animal genetic resources, *Food and Agriculture Organization of the United Nations*.

Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, Zhang X, Ni Z, Hou F, Long R, et al. 2015. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat Commun*. 6:10283.

Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, et al. 2012. The yak genome and adaptation to life at high altitude. *Nat Genet*. 44(8):946–949.

Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW, Barendse W. 2016. A meta-assembly of selection signatures in cattle. *PLoS One* 11(4):e153013.

Rincon G, Islas-Trejo A, Casellas J, Ronin Y, Soller M, Lipkin E, Medrano JF. 2009. Fine mapping and association analysis of a quantitative trait locus for milk production traits on *Bos taurus* autosome 4. *J Dairy Sci*. 92(2):758–764.

Ruetz TJ, Lin AE, Guttman JA. 2012. Enterohaemorrhagic *Escherichia coli* requires the spectrin cytoskeleton for efficient attachment and pedestal formation on host cells. *Microb Pathog*. 52(3):149–156.

Sartori R, Bastos MR, Baruselli PS, Gimenes LU, Ereno RL, Barros CM. 2010. Physiological differences and implications to reproductive management of *Bos taurus* and *Bos indicus* cattle in a tropical environment. *Reprod Domest Rumin Vii* 7(1):357–375.

Sattler K, Levkau B. 2009. Sphingosine-1-phosphate as a mediator of high-density lipoprotein effects in cardiovascular protection. *Cardiovasc Res*. 82(2):201–211.

Scherf BD, Pilling D. 2015. The second report on the state of the world's animal genetic resources for food and agriculture. *Food and Agriculture Organization of the United Nations*.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 46(8):919–925.

Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kuhn C, Kinoshita A, Sugimoto Y, Takasuga A. 2011. The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (*NCAPG*) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Anim Genet*. 42(6):650–655.

Sherratt A. 1983. The secondary exploitation of animals in the Old World. *World Archaeol*. 15(1):90–104.

Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT, Casas E, Wray JE, White J, Cho J, Fahrenkrug SC, et al. 2001. Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res*. 11(4):626–630.

Sorbolini S, Marras G, Gaspa G, Dimauro C, Cellesi M, Valentini A, Macciotta NP. 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.

Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, Moore SS. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 12:559.

Sun YB, An ZS. 2005. Late Pliocene-Pleistocene changes in mass accumulation rates of eolian deposits on the central Chinese Loess Plateau. *J Geophys Res-Atmos*. 110(D23):1193–1194.

Svizzero S, Tisdell C. 2016. Input shortages and the lack of sustainability of bronze production by the Únětice. In: Working Papers on Economics, Ecology and the Environment, No 202. Queensland (Australia): University of Queensland.

Switonski M, Mankowska M, Salamon S. 2013. Family of melanocortin receptor (MCR) genes in mammals-mutations, polymorphisms and phenotypic effects. *J Appl Genet*. 54(4):461–472.

Van Vuure T. 2002. History, morphology and ecology of the aurochs (*Bos primigenius*). Available from: http://citeseerx.ist.psu.edu/viewdoc/summary? doi=10.1.1.534.6285.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 38(16):e164.

Wang M, Ding Y. 1996. The importance of work animals in rural China. *World Anim Rev*. 86:65–67.

Wedholm A, Larsen LB, Lindmark-Mansson H, Karlsson AH, Andren A. 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J Dairy Sci*. 89(9):3296–3305.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.

Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CP, Sonstegard TS, Liu GE. 2015. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol*. 32(3):711–725.

Xu Y, Yu W, Xiong Y, Xie H, Ren Z, Xu D, Lei M, Zuo B, Feng X. 2011. Molecular characterization and expression patterns of serine/arginine-rich specific kinase 3 (*SPRK3*) in porcine skeletal muscle. *Mol Biol Rep*. 38(5):2903–2909.

Yahvah KM, Brooker SL, Williams JE, Settles M, Mcguire MA, Mcguire MK. 2015. Elevated dairy fat intake in lactating women alters milk lipid and fatty acids without detectible changes in expression of genes related to lipid uptake or synthesis. *Nutr Res*. 35(3):221.

Yoon D, Ko E. 2016. Association study between SNPs of the genes within bovine QTLs and meat quality of Hanwoo. *J Anim Sci*. 94(Suppl 4):145.

Yu Y, Nie L, He ZQ, Wen JK, Jian CS, Zhang YP. 1999. Mitochondrial DNA variation in cattle of south China: origin and introgression. *Anim Genet*. 30(4):245–250.

Zhang H, Paijmans JL, Chang F, Wu X, Chen G, Lei C, Yang X, Wei Z, Bradley DG, Orlando L, et al. 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nat Commun*. 4:2755.

Zhang H, Wang S, Wang Z, Da Y, Wang N, Hu X, Zhang Y, Wang Y, Leng L, Tang Z, et al. 2012. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. *BMC Genomics* 13:704.

Zhang Z, Wang Z, Yang Y, Zhao J, Chen Q, Liao R, Chen Z, Zhang X, Xue M, Yang H, et al. 2016. Identification of pleiotropic genes and gene sets underlying growth and immunity traits: a case study on Meishan pigs. *Animal* 10(4):550–557.

Zhao C, Tian F, Yu Y, Luo J, Hu Q, Bequette BJ, Baldwin VR, Liu G, Zan L, Scott UM, et al. 2012. Muscle transcriptomic analyses in Angus cattle with divergent tenderness. *Mol Biol Rep*. 39(4):4185–4193.

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, et al. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet*. 45(1):67–71.

Zheng B, Xu Q, Shen Y. 2002. The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quatern Int*. 97–98:93–101.

Zhou X, Wang B, Pan Q, Zhang J, Kumar S, Sun X, Liu Z, Pan H, Lin Y, Liu G, et al. 2014. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat Genet*. 46(12):1303–1310.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, et al. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol*. 10(4):R42.