

Impact of Parental *Bos taurus* and *Bos indicus* Origins on Copy Number Variation in Traditional Chinese Cattle Breeds

Liangzhi Zhang^{1,2,†}, Shangang Jia^{1,†}, Martin Plath¹, Yongzhen Huang¹, Congjun Li³, Chuzhao Lei¹, Xin Zhao¹, and Hong Chen^{1,*}

¹Shaanxi Key Laboratory of Molecular Biology for Agriculture, College of Animal Science and Technology, Northwest A & F University, Yangling, Shaanxi, China

²Key laboratory of adaptation and evolution of plateau biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences

³United States Department of Agriculture-Agricultural Research Service, Bovine Functional Genomics Laboratory, Beltsville, Maryland

*Corresponding author: E-mail: chenhong1212@263.net.

[†]These authors contributed equally to this work.

Associate editor: Greg Elgar

Accepted: July 28, 2015

Abstract

Copy number variation (CNV) is an important component of genomic structural variation and plays a role not only in evolutionary diversification but also in domestication. Chinese cattle were derived from *Bos taurus* and *Bos indicus*, and several breeds presumably are of hybrid origin, but the evolution of CNV regions (CNVRs) has not yet been examined in this context. Here, we used CNVRs, mtDNA D-loop sequence variation, and Y-chromosomal single nucleotide polymorphisms to assess the impact of maternal and paternal *B. taurus* and *B. indicus* origins on the distribution of CNVRs in 24 Chinese domesticated bulls. We discovered 470 genome-wide CNVRs, only 72 of which were shared by all three Y-lineages (*B. taurus*: Y₁, Y₂; *B. indicus*: Y₃), whereas 265 were shared by inferred taurine or indicine paternal lineages, and 228 when considering their maternal taurine or indicine origins. Phylogenetic analysis uncovered eight taurine/indicine hybrids, and principal component analysis on CNVs corroborated genomic exchange during hybridization. The distribution patterns of CNVRs tended to be lineage-specific, and correlation analysis revealed significant positive or negative co-occurrences of CNVRs across lineages. Our study suggests that CNVs in Chinese cattle partly result from selective breeding during domestication, but also from hybridization and introgression.

Key words: CNV, Y-chromosomal SNPs, mtDNA D-loop, parental origin.

Introduction

Cattle have a long history of domestication and selective breeding (Bradley et al. 1996). Several cattle genomes have been published in recent years (Elsik et al. 2009; Canavez et al. 2012), and so systematic investigations of genomic changes accompanying the domestication process—such as breed-specific genomic differentiation, hybridization, and introgression—have become feasible, including the identification of genome-wide single nucleotide polymorphisms (SNPs) (Decker et al. 2009; Gibbs et al. 2009; Jansen et al. 2013; McTavish et al. 2013) and copy number variations (CNVs) (Liu et al. 2010; Pollinger et al. 2010). Genic CNV regions (CNVRs) have the potential to affect phenotypes through the disruption of gene dosages, unmasking of recessive alleles, and loss of regulatory elements or regulatory polymorphisms (Weischenfeldt et al. 2013). Furthermore, gene duplications

played a role in the evolution of sexual dimorphism (Novembre and Ramachandran 2011) and facilitated gene neofunctionalization or subfunctionalization during evolutionary diversification (Adler et al. 2014).

Several investigations of cattle CNVs have been conducted, focusing mostly on breed-specific patterns of CNV (Bae et al. 2010; Fadista et al. 2010; Liu et al. 2010; Hou et al. 2011; Stothard et al. 2011; Zhan et al. 2011; Bickhart et al. 2012; Jiang et al. 2012, 2013). However, modern breeding practices commonly used in the cattle industry include artificial insemination and hybridization of breeds (thereby introducing more variation in CNVs), for example, when importing foreign excellent cattle sperm for insemination of Chinese cows (Li, Xie, et al. 2013). This highlights the necessity for additional, comparative investigations of lineage-specific patterns of CNVs. Moreover, artificial selection during the domestication process

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

or in modern breeding programs leads to rapid genomic evolution (Novembre and Ramachandran 2011; Parsch and Ellegren 2013), and it remains imperative to understand the links between domestication history and artificial selection, genomic changes (including CNVs), and phenotypic divergence in production animals (Kijas et al. 2009; Novembre and Ramachandran 2011). Selection for phenotypic traits associated with high CNV of certain genomic regions, for example, could help explain why some CNVs are present in domestic animals despite the predicted loss of redundant gene copies with time (Liu et al. 2010; Zhang et al. 2014).

With 28 recognized breeds, China is particularly rich in genomic resources for research on cattle (Qiu et al. 1988). The high degree of phenotypic and genomic differences between breeds is partly due to a long domestication history and adaptation to different local/climatic conditions throughout China (Chan et al. 2010; Jansen et al. 2013). Another important source of variation is the origin of breeds from either *Bos taurus* (taurine) or *Bos indicus* (indicine cattle): Phylogeographic studies on sequence variation of maternally (mitochondrial) and paternally inherited (Y chromosomal) markers found Chinese cattle to be divided into two main clades (taurine and indicine), the ancestors of which likely originated from the Near East and the Indian subcontinent, respectively (MacHugh et al. 1997). Today, taurine cattle are dominant in northern China and indicine cattle in southern China (Chen and Qiu 1993; Lai et al. 2006; Lei et al. 2006; Jia et al. 2007, 2010). Analysis of Y-chromosome SNPs identified three haplotypes, namely Y₁ (taurine origin), Y₂ (taurine) and Y₃ (indicine) (Götherström et al. 2005), and subsequent investigations using Y-SNPs and Y-STRs confirmed that Y₂ dominated in the north (91.4%) and Y₃ in the south of China (90.8%) (Li, Zhang, et al. 2013). A number of studies also suggest that hybridization and introgression of taurine and indicine cattle occurred, especially in central parts of China, but inferences were mostly made on the basis of mtDNA and Y-chromosomal information only (Lai et al. 2006; Jia et al. 2010; Li, Xie, et al. 2013). Y-chromosomes and mtDNA, however, generally lack recombination and thus, are of limited use to unravel patterns of genome evolution after hybridization and artificial selection (McTavish et al. 2013).

With an increasing number of genomic data sets being published every year, genome-wide markers are increasingly utilized to analyze the evolutionary/genomic histories (sometimes including domestication effects) not only of model species such as *Drosophila* fruit flies (Emerson et al. 2008), humans (Novembre and Ramachandran 2011) and chimpanzee (Gatto et al. 2006), but also of nonmodel organisms (Qu et al. 2013) and increasingly domestic animals: Cattle (MacHugh et al. 1997; Gibbs et al. 2009; McTavish et al. 2013), sheep (Kijas et al. 2009), dogs (Pollinger et al. 2010), horses (McCue et al. 2012; Metzger et al. 2013), and pig (Li, Tian, et al. 2013). In cattle, SNPs have been applied to study their genomic diversity and to make inferences about their

domestication history, and a recent study corroborated cross-breeding of taurine and indicine cattle in central Asia (Decker et al. 2014). However, to the best of our knowledge, breed-specific differences in CNV, and especially the evolution of CNVRs after hybridization between taurine and indicine cattle, have not yet been addressed. In our present study, we inferred the origins (taurine or indicine) of 24 Chinese bulls from 12 different breeds (supplementary table S1, Supplementary Material online) based on Y-chromosomal SNPs (Ginja et al. 2009) and mtDNA D-loop sequence variation (Jia et al. 2010) and reanalyzed a genome-wide CNV data set generated by means of microarray-based comparative genomic hybridization (array CGH) (Zhang et al. 2014). Simultaneously considering the maternal and paternal origins of those breeds allowed interpreting breed-specific differences in CNVR in light of their domestication history that involved not only prolonged artificial selection but also hybridization between distant lineages (Lai et al. 2006; Jia et al. 2010).

Materials and Methods

Sample Collection

We collected blood samples of $n = 24$ bulls from 12 typical and common cattle breeds throughout China (supplementary table S1 and fig. S1, Supplementary Material online): Anxi (AX), Bohaihei (BH), Chinese Holstein (HD), Jiaxian (JX), Jinnan (JN), Hainan (HN), Luxi (LC), Mongolian (MG), Nanyang (NY), Qinchuan (QQ), Wannan (WN), and Zaosheng cattle (ZS). Of these, MG, AX, and ZS stemmed from the northern range of China, HN and WN from southern parts, whereas the others came from central China (supplementary fig. S1, Supplementary Material online). BH is the only black breed and HD is the main dairy cow breed in China. For our quantitative real-time polymerase chain reaction (qPCR) approach, we additionally collected five pure Angus bulls (AG)—an introduced breed—from Shaanxi Province as reference samples of confirmed taurine origin. Sample collection was carried out in accordance with the ethical guidelines approved by the Animal Care Commission of the College of Animal Science and Technology, Northwest A & F University. Genomic DNA was extracted (Sambrook and Russell 2001) and purified from whole blood using the DNA purification kit (Plus Minipreps DNA Purification System; Promega, Beijing, China), and quantified using spectrophotometry and agarose gel electrophoresis.

Y-Chromosomal and mtDNA (D-loop) Haplotyping

We determined Y-chromosomal haplotypes of the 24 bulls according to previously published protocols (Li, Xie, et al. 2013; Li, Zhang, et al. 2013). In brief, two primer pairs were used for PCR amplification (supplementary table S2, Supplementary Material online), and after purification PCR products were Sanger-sequenced by Sangon Biotech

(Shanghai, China). We distinguished Y_1 from Y_2 - and Y_3 -haplotypes based on the SNP (C/A, position 423 in AY936543) of *UTY-19*, and Y_3 from Y_1 and Y_2 by using the SNP (T/C, position 655 in AF241271) of *ZFY-10*.

Maternal haplotypes of 19 bulls were determined by sequencing the entire D-loop (supplementary table S2, Supplementary Material online, for primers), and sequences were aligned using ClustalW (Larkin et al. 2007). The clades of taurine and indicine origins inferred by means of a maximum-likelihood tree following the Tamura–Nei model (Tamura and Nei 1993) with 1,000 bootstrap replicates calculated using MEGA6 (Tamura et al. 2013), and the general time reversible (GTR) substitution model (Gatto et al. 2006) in MrBayes v.3.2 (Ronquist et al. 2012). For another $n = 5$ individuals, no information could be retrieved as we ran out of DNA isolate as a result of the CNV analyses.

Array CGH Platform and Data Analysis

The 24 Chinese bulls were included in our CGH array approach described elsewhere (Zhang et al. 2014), and one Angus bull was used as the reference sample. CNV calling for copy number gains and losses was based on \log_2 signal intensity (Olshen et al. 2004): Segments with mean $|\log_2$ ratios ≥ 0.5 and covering at least five consecutive probes were considered to represent CNVRs. Chromosome (i.e., karyogram) plotting, heat map drawing, indicator analysis (De Cáceres et al. 2012), and nonmetric multidimensional scaling (NMDS) were carried out using the R-packages ggbio, heatmap.2, indicpecies, and vegan, respectively. Principal component analysis (PCA) was performed in STAMP v.2.02 (Parks and Beiko 2010). A phylogenetic tree was constructed based on binary data (CNVR present/absent) in MrBayes v.3.2 (Ronquist et al. 2012) using the GTR substitution model (Gatto et al. 2006).

Validation of CNVRs by Using qPCR

We performed qPCR analysis to validate copy number differences detected by our CGH approach on CFX-96 Real-Time PCR Detection System (Bio-Rad). The primers were designed by using Beacon Designer (PREMIER Biosoft, USA) or from the previous paper (Fadista et al. 2010; Zhang et al. 2014) (supplementary table S5, Supplementary Material online). Two control primers (Primer1 and Primer7, in supplementary table S5, Supplementary Material online), which served as an internal standard, were coamplified with the corresponding test primers. PCR reaction was done in a volume of 20 μ l containing 20 ng of genomic DNA, 0.4 μ M of each primer, and SYBR Premix Ex Taq II reagents (TaKaRa Biotechnology, Dalian, China). Three replicate reactions were performed for primer pairs, and a comparative CT method was used to calculate the copy number. The $\Delta\Delta Ct$ values were determined by comparing test samples and Angus reference (two-copy

state) with internal control. Finally, the relative copy number for each sample was calculated as $2^{-\Delta\Delta Ct}$ (Zhang et al. 2014).

Results

Identifying Y-Chromosomal and mtDNA Lineages

Distinguishing taurine from indicine cattle based on morphological characteristics can lead to misclassifications, and Y-chromosomal SNPs in combination with mitochondrial sequence variation are widely accepted for exploring the paternal and maternal origins of cattle (Ginja et al. 2009; Jia et al. 2010; Li, Zhang, et al. 2013). By combining SNPs of the Y-chromosomal *UTY* and *ZFY* genes, we could unambiguously assign $n = 24$ bulls to two taurine Y-chromosomal haplotypes (Y_1 , $n = 2$; Y_2 , $n = 13$) and an indicine haplotype (Y_3 , $n = 9$; supplementary table S2, Supplementary Material online).

Likewise, we analyzed the maternal origins of the studied bulls in a Bayesian phylogenetic approach, based on Sanger-sequenced mtDNA (D-loop) amplicons, which divided the samples into two major clades (i.e., taurine, $n = 8$; indicine, $n = 11$; fig. 1a). Combining information from both analyses on the taurine/indicine clustering, we inferred 11 samples to have unambiguous taurine or indicine origin, that is, both parents had the same origins (supplementary table S2, Supplementary Material online). In eight samples from five central Chinese breeds (BH, JN, JX, NY, and QQ), however, we found a mismatch between paternal and maternal information and classified those bulls as hybrids (supplementary table S2, Supplementary Material online). In summary, we confirmed northern breeds to be of taurine origin, southern breeds to be of indicine origin, whereas central Chinese breeds appear to be of hybrid origin (supplementary fig. S1, Supplementary Material online).

CNVs in Different Clades of Cattle

Emerson et al. (2008) demonstrated that the genome-wide distribution of CNVs varied significantly among genomic regions of *Drosophila melanogaster*. In our present study, we identified 356 CNVRs in the $n = 24$ examined bulls that could be mapped to chromosomes 1 through 29, the X-chromosome and the mitochondrion, amounting to 38.8 Mb (i.e., 1.33% of the 2,918.1 Mb Btau_4.0 cattle genome [Elsik et al. 2009]; fig. 2). When including the 114 unassigned sequences, the total number of 470 discovered CNVRs (62.1 Mb) corresponds to 2.13% of the cattle genome. We inferred 314 sequence losses, 112 gains, and 44 cases of combined losses and gains within the same CNVR (table 1 and supplementary table S3, Supplementary Material online). Eighty-two CNVRs were unique, that is, found in only one individual, whereas 388 CNVRs were shared by at least two individuals, 48 of which had a frequency ≥ 0.5 (table 1).

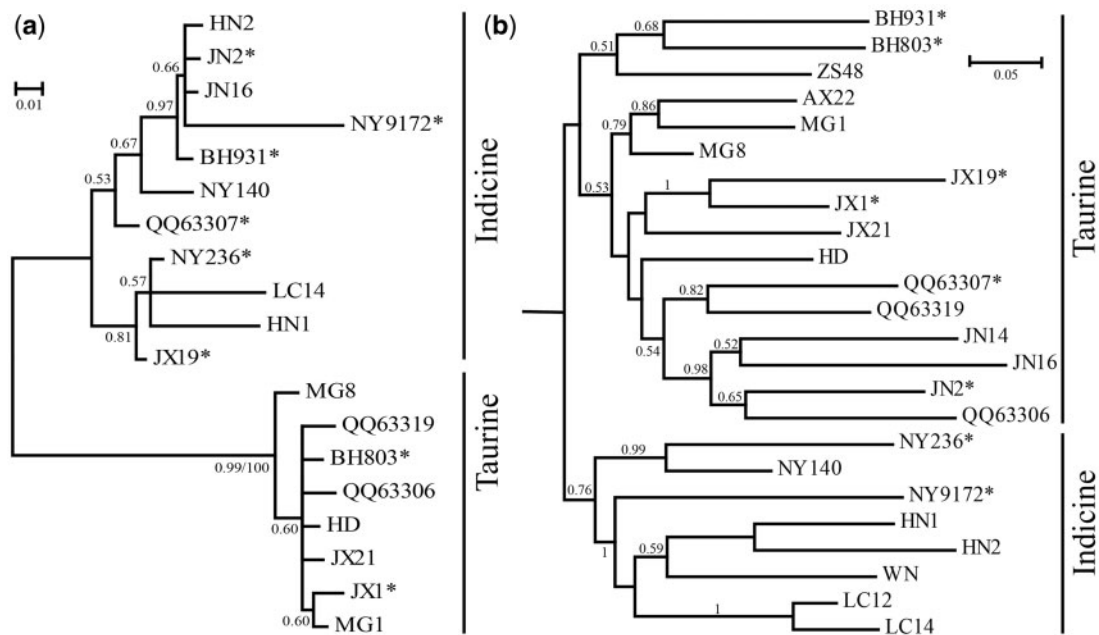


FIG. 1.—Phylogenetic trees of Chinese bulls derived from (a) mitochondrial (D-loop) sequence variation and (b) CNVR information (present/absent). Putative hybrids of taurine and indicine cattle (based on the D-loop and Y-chromosomal SNPs) are marked with “*”. Posterior probabilities/bootstraps values are shown if >0.5 or >50 .

We found a significant difference in the composition of the 469 CNVRs (excluding one CNVR in mtDNA) in our $n=24$ samples (Cochran’s Q -test: $Q_{23, 468} = 135.4$, $P = 2.2 \times 10^{-16}$). When considering the three Y-chromosomal haplotypes, CNVRs were more frequent in Y_2 - and Y_3 -haplotypes than in Y_1 -bulls (table 1 and [supplementary fig. S2, Supplementary Material](#) online). We compared the extent of CNVR-sharing and found that 105 CNVRs were shared by Y_1 (105/125; 84%) and Y_2 (105/390; 27%), 254 by Y_2 (254/390; 65%) and Y_3 (254/324; 78%), 83 by Y_1 (83/125; 66%) and Y_3 (83/324; 26%), and 72 by all three Y-haplotypes ([supplementary fig. S2a, Supplementary Material](#) online). We considered the paternal (and maternal) taurine/indicine origins of the examined bulls and found 265 (and 228) CNVRs to be shared by taurine and indicine cattle, 145 (and 60) to be unique to taurine and 59 (and 166) to indicine cattle ([supplementary fig. S2b and c, Supplementary Material](#) online).

Phylogenetic Analysis Based on CNVRs

To shed light on the potential evolutionary contributions of CNVs to the formation of major cattle breeds in China, we conducted Bayesian phylogenetic analysis of the different breeds based on CNVR information. Individuals of the same breed clustered closely in the phylogenetic tree (fig. 1b), which was corroborated by sample correlation plot analysis ([supplementary fig. S3b, Supplementary Material](#) online).

A geographic pattern emerged in which the northern (taurine) samples (MG and AX cattle) clustered together and were separated from others; a similar pattern was apparent for a southern cluster (WN, HN, and LC breeds) of indicine decent (fig. 1b and [supplementary fig. S3a, Supplementary Material](#) online). Analysis of variance (ANOVA) confirmed that the grouping of taurine and indicine cattle was statistically significant; that is, ANOVA yielded significant results for the comparisons of $Y_1 + Y_2$ versus Y_3 ($F_{1, 468} = 6.49$, $P = 0.011$), and Y_1 - versus Y_2 - versus Y_3 -haplotypes ($F_{2, 468} = 6.65$, $P = 0.0013$). The position of putative hybrid samples in the CNVR phylogenetic tree (fig. 1b) was partly in accordance with their inferred maternal (e.g., BH803 in the taurine clade and NY9172 in the indicine clade) or paternal origins (e.g., JN2 and QQ63307 in the taurine clade). In the eight hybrids, six samples clustered in the taurine clade of the CNVR-based tree, but in the indicine clade in the mtDNA (D-loop)-based tree. This suggests that introgression of indicine into taurine cattle in central Chinese breeds was mainly female-mediated, involving indicine cows and taurine bulls.

PCA to Detect Hybridization and Introgression

In another approach to explore the effects of hybridization among breeds, we performed PCA using CNVR information. Plotting the samples according to their paternal Y-chromosomal or maternal mt-haplotype uncovered intermediate positions of putative hybrid individuals (marked with * in fig. 3)

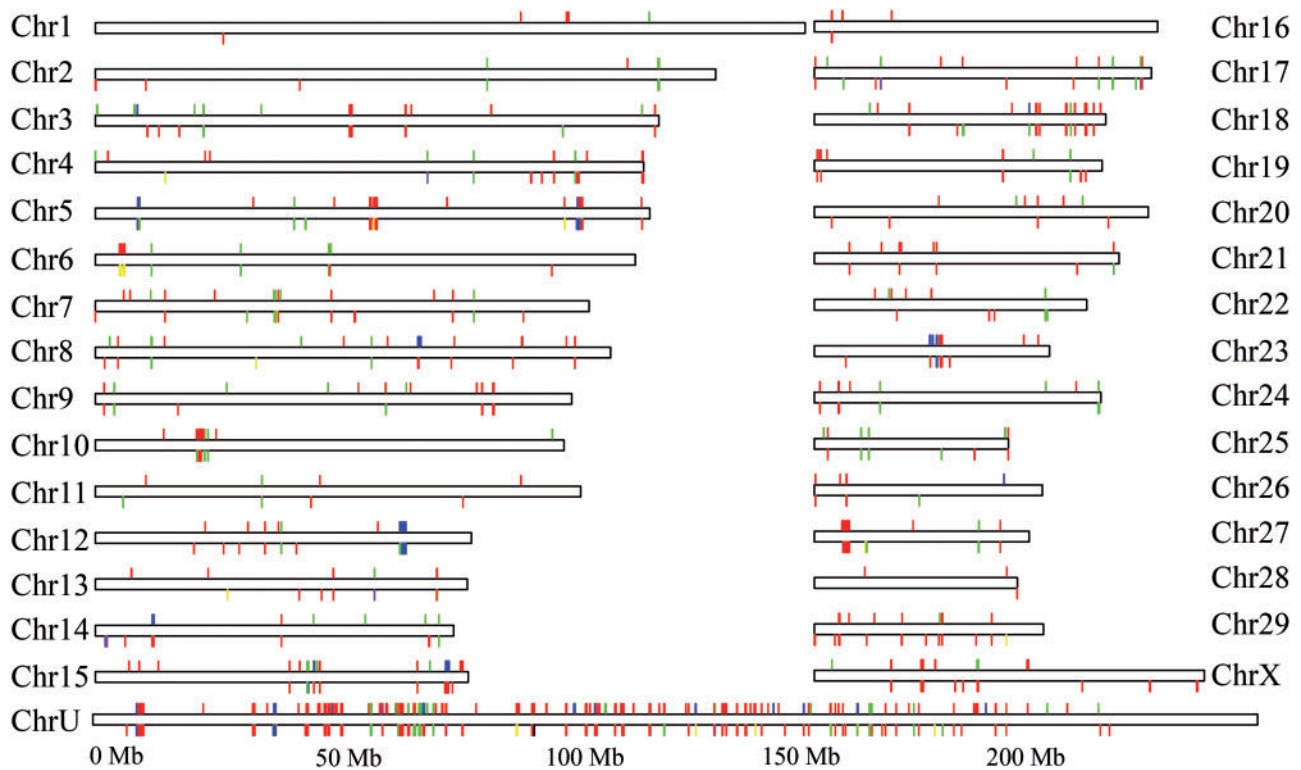


Fig. 2.—Distribution of different CNVRs across chromosomes (excluding the mitochondrion) in the cattle genome. Y_2 -individuals are shown above the chromosomes in green (gains), red (losses) and dark blue (both), whereas below chromosomes, information for Y_3 -individuals (same colors as for Y_2) and Y_1 -bulls (purple: gains, yellow: losses, black: both) is given.

Table 1

Summary Statistics for CNV in Chinese Bulls ($n=24$)

Paternal Haplotype	Sample Size (n)	Count	Unique	Gains	Losses	Both	Total Length (kb)
Y_1 (taurine)	2	69 (37.0)	55 (37.5)	30 (15.0)	38 (19.0)	1 (0.5)	10,497 (152)
Y_2 (taurine)	13	337 (25.9)	87 (6.7)	90 (6.9)	218 (16.8)	29 (2.2)	44,785 (133)
Y_3 (indicine)	9	251 (27.9)	47 (5.2)	65 (7.2)	178 (19.8)	8 (0.9)	32,895 (131)
Total	24	470 (19.6)	82 (3.4)	112 (4.7)	314 (13.1)	44 (1.8)	62,073 (132)

NOTE.—Numbers in parentheses are normalized by sample size or CNVR counts.

between the taurine and indicine clusters. The central Chinese QQ breed, for example, has recently received foreign (taurine) sperm during artificial insemination (Li, Xie, et al. 2013), which explains why QQ63307 (with an indicine maternal background) had a large number of Y_1 -specific CNVRs and, as an exception among the hybrid bulls, was placed in the Y_1/Y_2 cluster (figs. 1b and 3 and [supplementary fig. S4a](#), [Supplementary Material](#) online). It is interesting to note that JN16 with indicine parents (maternal type: indicine; paternal type: Y_3) clustered in the taurine group. Overall, our results were in good agreement with our a priori predictions regarding the distinctness of northern and southern breeds as well as ongoing hybridization and introgression in central parts of

China and, more generally, validate the use of CNVR information to study the domestication history and evolution of cattle breeds.

NMDS to Identify Lineage-Specific CNVRs

To examine the contribution of lineage-specific CNVRs to the observed differentiation between breeds in more detail, we applied a powerful statistical approach derived from community ecology, namely NMDS (Dixon 2003), which visualizes similarity between individuals in complex data sets. Congruent with the results from phylogenetic analyses and PCA, samples were generally divided into two groups of

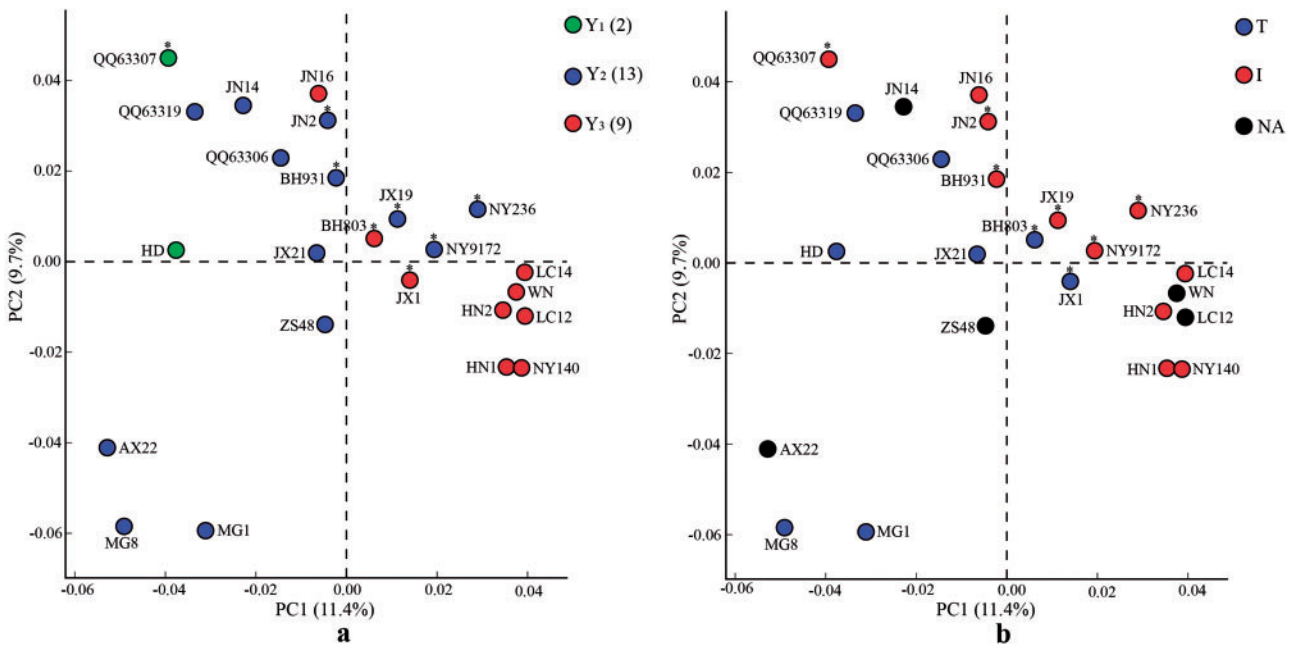


Fig. 3.—PCA of CNVs. (a) Plotting split by Y-chromosomal haplotypes, and (b) split into maternal groups of taurine (T) and indicine (I) origins. NA, information not available.

taurine and indicine cattle (supplementary fig. S4a and b, Supplementary Material online). We used indicator analysis (De Cáceres et al. 2012) to detect CNVRs that were statistically more abundant in groups than expected by chance (table 2 and supplementary table S4, Supplementary Material online) and thus identified 63 indicator CNVRs. Twenty of these corresponded to specific paternal Y-haplotypes, 24 to the paternal taurine/indicine grouping, 9 to maternal taurine/indicine haplotypes, and 40 to CNVR-clades of taurine and indicine origin as inferred in our Bayesian phylogenetic analysis (fig. 1b; $P < 0.05$ in all cases). Finally, rerunning the NMDS using only indicator CNVRs resulted in a much stronger clustering (supplementary fig. S4c–f, Supplementary Material online).

Validation of Selected CNVRs by Quantitative Real-Time PCR

Five CNVRs were selected to validate our CGH results. CNVR117 and CNVR209 were chosen as examples of lineage-specific CNVRs, whereby CNVR117 and CNVR226 had already been examined in previous studies, and data were reanalyzed in this study (Fadista et al. 2010; Zhang et al. 2014), and CNVR217 and CNVR283 were chosen as examples of CNVRs with low and high frequencies of occurrence, respectively (table 2). We designed five primer pairs for qPCR amplification, and two primer pairs for our internal reference (supplementary table S5, Supplementary Material online). In all five assays, qPCR confirmed the patterns predicted from

our prior analyses on CNV (see above), demonstrating the reliability of our results (supplementary table S5, Supplementary Material online). We selected CNVR117 for additional validation of its specificity to bulls showing the Y_1 -haplotype (see results from our indicator analysis) using a larger sample size ($n=29$) including $n=5$ Angus bulls of known taurine (Y_1) origin (supplementary table S2, Supplementary Material online), and found all Y_1 -bulls to show more copy numbers than Y_2 or Y_3 -bulls (supplementary table S6 and fig. S5, Supplementary Material online). This result underscores the validity of using CNVR117 to distinguish Y_1 from Y_2 or Y_3 -bulls.

Co-Occurrence of Different CNVRs

We investigated positive (Spearman’s correlation coefficient $r_s \geq 0.6$, $P < 0.01$) and negative links ($r_s \leq -0.6$, $P < 0.01$) between different CNVRs, as statistically significant links may be indicative of correlated selection on different genomic regions in the different breeds, as well as random genetic drift and subsequent breed-specific accumulation of certain CNVR combinations. Significant positive and negative links decreased strongly at CNVRs with prevalence greater than 3/24 (i.e., CNVRs present in more than 3 of the 24 bulls; supplementary table S7 and fig. S6, Supplementary Material online), which likely reflects our sample of $n=1-3$ individuals per breed and implies that most “links” are driven by breed-specific CNVR patterns (i.e., the contribution of single or few breeds led to the statistically significant results). Likewise,

Table 2

Lineage-Specific CNVRs in Paternal (Y-Chromosomal) and Maternal (mtDNA) Haplotypes

CNVRs	Frequency in Groups								Significance (Indicator Analysis)
	Y ₁	Y ₂	PT	PI	MT	MI	CT	CI	
CNVR68	0.00	0.31	0.27	0.78	0.00	0.64	0.25	0.88	Y ₃ [*] , PI [*] , MI ^{**} , CI ^{**}
CNVR99	0.00	0.00	0.00	0.33	0.00	0.18	0.00	0.38	Y ₃ [*] , PI [*] , CI [*]
CNVR162	0.00	0.00	0.00	0.33	0.00	0.18	0.00	0.38	Y ₃ [*] , PI [*] , CI [*]
CNVR355	0.00	0.00	0.00	0.33	0.00	0.18	0.00	0.38	Y ₃ [*] , PI [*] , CI [*]
CNVR469	0.00	0.00	0.00	0.44	0.25	0.18	0.19	0.13	Y ₃ [*] , PI ^{**}
CNVR213	0.00	0.00	0.00	0.33	0.13	0.18	0.06	0.25	Y ₃ [*] , PI [*]
CNVR308	0.00	0.00	0.00	0.44	0.00	0.27	0.06	0.38	Y ₃ [*] , PI [*]
CNVR468	0.00	0.00	0.00	0.33	0.13	0.18	0.06	0.25	Y ₃ [*] , PI [*]
CNVR409	0.00	0.31	0.27	0.89	0.25	0.55	0.31	0.88	Y ₃ ^{**} , PI ^{**} , CI [*]
CNVR10	0.00	0.15	0.13	0.78	0.38	0.36	0.25	0.63	Y ₃ ^{**} , PI ^{**}
CNVR326	0.00	0.54	0.47	0.78	0.38	0.64	0.38	1.00	Y ₂ +Y ₃ [*] , CI ^{**}
CNVR117	0.00	0.92	0.80	1.00	0.75	0.82	0.81	1.00	Y ₂ +Y ₃ ^{**}
CNVR126	0.00	0.85	0.73	0.78	0.75	0.73	0.63	1.00	Y ₂ +Y ₃ [*]
CNVR56	1.00	0.77	0.80	0.22	0.88	0.27	0.88	0.00	Y ₁ +Y ₂ ^{**} , PT [*] , CT ^{***}
CNVR209	1.00	0.62	0.67	0.00	0.38	0.27	0.56	0.13	Y ₁ +Y ₂ ^{***} , PT ^{**}
CNVR395	1.00	0.38	0.47	0.11	0.50	0.18	0.50	0.00	Y ₁ [*] , CT [*]
CNVR416	1.00	0.38	0.47	0.00	0.50	0.27	0.44	0.00	Y ₁ ^{**} , PT [*] , CT [*]
CNVR41	1.00	0.31	0.40	0.00	0.38	0.09	0.38	0.00	Y ₁ ^{**}
CNVR172	1.00	0.23	0.33	0.11	0.38	0.09	0.38	0.00	Y ₁ [*]
CNVR440	1.00	0.38	0.47	0.22	0.63	0.36	0.50	0.13	Y ₁ [*]
CNVR96	0.50	0.38	0.40	0.22	0.63	0.18	0.50	0.00	CT [*]
CNVR165	0.50	0.46	0.47	0.11	0.38	0.27	0.50	0.00	CT [*]
CNVR12	0.50	0.77	0.73	0.11	0.75	0.27	0.75	0.00	PT ^{**} , CT ^{**}
CNVR123	0.50	0.38	0.40	0.00	0.13	0.27	0.38	0.00	PT [*]
CNVR200	0.00	0.54	0.47	0.00	0.38	0.18	0.31	0.25	PT [*]
CNVR227	1.00	0.92	0.93	0.44	0.75	0.64	0.88	0.50	PT [*]
CNVR224	0.00	0.23	0.20	0.67	0.00	0.55	0.13	0.88	PI [*] , MI [*] , CI ^{***}
CNVR76	0.00	0.08	0.07	0.56	0.13	0.27	0.06	0.63	PI [*] , CI ^{**}
CNVR389	0.50	0.00	0.07	0.56	0.00	0.36	0.06	0.63	PI ^{**} , CI ^{**}
CNVR322	0.00	0.08	0.07	0.67	0.13	0.36	0.19	0.50	PI ^{**}
CNVR26	0.50	0.31	0.33	0.78	0.38	0.45	0.38	0.75	PI [*]
CNVR146	0.00	0.00	0.00	0.33	0.13	0.09	0.06	0.25	PI [*]
CNVR379	0.50	0.62	0.60	1.00	0.75	0.73	0.69	0.88	PI [*]
CNVR75	0.50	0.46	0.47	0.89	0.38	0.82	0.44	1.00	MI [*] , CI ^{**}
CNVR170	0.50	0.54	0.53	0.89	0.38	0.91	0.50	1.00	MI ^{**} , CI [*]
CNVR27	0.00	0.31	0.27	0.56	0.00	0.45	0.31	0.50	MI [*]
CNVR60	0.00	0.38	0.33	0.33	0.00	0.45	0.19	0.63	MI [*]
CNVR199	0.00	0.23	0.20	0.44	0.00	0.36	0.19	0.50	MI [*]
CNVR338	0.50	0.31	0.33	0.44	0.00	0.45	0.31	0.50	MI [*]
CNVR421	0.00	0.31	0.27	0.22	0.00	0.45	0.13	0.50	MI [*]
CNVR228	0.00	0.31	0.27	0.67	0.13	0.55	0.13	1.00	CI ^{***}
CNVR254	0.00	0.15	0.13	0.33	0.00	0.36	0.00	0.63	CI ^{**}
CNVR366	0.00	0.15	0.13	0.56	0.00	0.36	0.06	0.75	CI ^{***}
CNVR399	0.00	0.46	0.40	0.67	0.25	0.55	0.25	1.00	CI ^{***}
CNVR405	0.00	0.38	0.33	0.78	0.25	0.55	0.25	1.00	CI ^{***}
CNVR62	0.00	0.08	0.07	0.33	0.00	0.27	0.00	0.50	CI ^{**}
CNVR95	0.00	0.46	0.40	0.78	0.38	0.55	0.31	1.00	CI ^{**}
CNVR307	0.00	0.08	0.07	0.44	0.00	0.27	0.00	0.63	CI ^{**}
CNVR331	0.00	0.08	0.07	0.44	0.00	0.27	0.00	0.63	CI ^{**}
CNVR20	0.00	0.08	0.07	0.22	0.00	0.18	0.00	0.38	CI [*]
CNVR23	0.50	0.69	0.67	0.78	0.75	0.64	0.56	1.00	CI [*]
CNVR38	0.00	0.15	0.13	0.11	0.00	0.27	0.00	0.38	CI [*]

(continued)

Downloaded from <http://gbe.oxfordjournals.org/> at Northwest Plateau Institute of Biology, CAS on December 19, 2016

Table 2 Continued

CNVRs	Frequency in Groups								Significance (Indicator Analysis)
	Y ₁	Y ₂	PT	PI	MT	MI	CT	CI	
CNVR46	0.00	0.15	0.13	0.11	0.00	0.27	0.00	0.38	CI*
CNVR47	0.00	0.08	0.07	0.22	0.00	0.18	0.00	0.38	CI*
CNVR69	0.00	0.15	0.13	0.33	0.00	0.27	0.06	0.50	CI*
CNVR221	0.00	0.08	0.07	0.22	0.00	0.27	0.00	0.38	CI*
CNVR323	0.00	0.08	0.07	0.22	0.00	0.18	0.00	0.38	CI*
CNVR344	1.00	0.38	0.47	0.89	0.38	0.73	0.44	1.00	CI*
CNVR371	0.00	0.23	0.20	0.22	0.13	0.36	0.06	0.50	CI*
CNVR397	0.00	0.38	0.33	0.56	0.38	0.45	0.25	0.75	CI*
CNVR407	0.50	0.62	0.60	0.89	0.88	0.64	0.56	1.00	CI*
CNVR438	0.00	0.38	0.33	0.56	0.25	0.45	0.25	0.75	CI*
CNVR461	0.00	0.15	0.13	0.33	0.13	0.27	0.06	0.50	CI*

NOTE.—Y-haplotypes indicate paternal taurine (PT; Y₁, Y₂) or indicine origins (PI; Y₃). Mt-haplotypes indicate maternal taurine (MT) or indicine (MI) origins. Also given are inferred CNVR-clades for taurine (CT) or indicine bulls (CI).

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

some breed/group-specific CNVRs were not found in other breeds/groups, resulting in negative links.

However, we identified 20 positive and 6 negative links for CNVRs with a prevalence ≥ 0.5 (supplementary table S8 and fig. S6, Supplementary Material online), which cannot be explained by the contributions of specific CNVR combinations in single breeds. These links partly reflect different origins from taurine or indicine ancestors; for example, all CNVRs with a frequency ≥ 0.5 that showed negative links were connected to CNVR56, which was specific to taurine cattle (see also table 2).

Discussion

Strictly paternally (Y-chromosomal) (Cai et al. 2006; Li, Xie, et al. 2013) and maternally inherited (mitochondrial) markers (Lai et al. 2006; Lei et al. 2006; Jia et al. 2010), both of which largely lack recombination, are well established to infer the paternal and maternal origins of cattle, respectively. Besides the two genes analyzed here, previous studies used sequence variation of the Y-chromosomal genes *USP9Y*, *ZFY*, *DDX3Y*, and *UTY* to determine the paternal origins of cattle breeds (Götherström et al. 2005; Li, Xie, et al. 2013; Li, Zhang, et al. 2013); however, our present study demonstrates that SNPs of the two genes *ZFY* and *UTY* were sufficient to unambiguously assign the 24 bulls studied herein to three paternal haplotypes (Y₁, Y₂, and Y₃)—in several cases (like Angus bulls) matching our a priori predictions. A previous study reported breed-specific differences in mtDNA D-loop sequence variation using a much larger data set of Chinese cattle and detected four taurine (T_{1A}, T₂, T₃, and T₅) and two indicine haplotypes (I₁ and I₂) (Jia et al. 2010). Our present study made use of the discriminatory power of D-loop sequence variation to determine the maternal origins of breeds, and our combined analysis not only identified two clades of taurine and indicine origins but also confirmed hybrid status of eight bulls stemming from Central China (Lei et al. 2006).

Crossbreeding of taurine and indicine in Chinese cattle breeds occurred historically (Chen and Qiu 1993) and is likely to increase in the future as regional interconnectivity increases. Hybridization is expected to introduce additional variation into cattle genomes, for example, due to recombination and assortment, and our different analyses on CNVRs (like PCA and NMDS) support this view, as a considerable degree of overlap between taurine and indicine groups suggests admixture of *B. taurus* and *B. indicus*.

Y-chromosomal and mtDNA markers are of limited use for reconstructing the evolutionary and/or domestication history of a given species as they do not allow tracing complex patterns of genome evolution (Ginja et al. 2009). Therefore, traditional methods, based on Y-chromosomal and mtDNA sequence variation, failed to unravel evolutionary processes shaping cattle genomes, although those markers are invaluable to detect crossbreeding of cattle breeds (Cai et al. 2006; Lai et al. 2006; Ginja et al. 2010). Our present study exemplifies the utility of genome-wide CNVRs, determined through whole-genome CGH arrays, to study the domestication history and genome evolution of cattle breeds, and our results are in good agreement with the results of a recent study using genome-wide SNPs (Decker et al. 2014). Given financial and time constraints of population-wide cattle genome sequencing, the method presented in this study offers a convenient and useful novel approach that provides a high number of polymorphic positions over the entire genome and thus represents an elegant alternative to other approaches based on SNPs, or microsatellite length polymorphisms.

Our strategy of combining the analysis of CNVRs with analyses of the paternal and maternal origins of cattle enabled us to trace signatures of mosaic genomic evolution (O'Brien et al. 2014). For example, an interesting finding in our present study was that one bull (JN16), which was inferred to be of indicine paternal and maternal origin, was placed into the taurine

clade when analyzing CNVRs (fig. 1b). This finding suggests that complex crossbreeding and backcrossing after hybridization of the two lineages—possibly in combination with selection for preferable phenotypic traits associated with CNVRs—may affect patterns of genomic introgression. Based on our observations of lineage-specific distribution patterns and (positive and negative) co-occurrences of different CNVRs, we argue that our findings at least in part reflect selection during the domestication history of cattle. Signatures of domestication and artificial selection have left their imprints in the cattle genome, as determined through SNPs (Kijas et al. 2009; Habier et al. 2010).

The Y₁-haplotype is nowadays dominant in north-central Europe, and the Y₂-haplotype prevails in central Europe (Edwards et al. 2011), whereas in Chinese breeds, Y₂ and Y₃ dominate in the north and south, respectively (Li, Zhang, et al. 2013). A previous study found the Y₁-haplotype to originate from European aurochs, specifically, from introgression after hybridization with local domestic cattle (Götherström et al. 2005). Our results indicate that some CNVRs tend to be specific to (or to occur at a higher frequency in) haplotype lineages or taurine/indicine groups, which could be interpreted as a signature of divergent selective regimes during their domestication history, even though genetic drift and random fixation of CNVRs certainly also need to be considered. For example, using a larger sample size we confirmed that CNVR117 was specific of Y₁-bulls. European aurochs, from which the Y₁-haplotype is derived, may have conferred higher copy numbers of CNVR117 compared with the Y₂ and Y₃ groups, and future studies will need to elaborate on the question of how this variation affects phenotypic traits and whether and how selection for those traits may be responsible for higher copy numbers of CNVR117 in Y₁-cattle.

Linking genomic structural variation to phenotypic differences and physiological performance remains a major challenge, and so answering the question of whether and how artificial selection may have affected CNVR evolution during the domestication of cattle will greatly benefit from future studies trying to link CNV to individual's development, physiology, behavior, and morphology.

Supplementary Material

Supplementary figures S1–S6 and tables S1–S8 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 30972080, No. 31272408), the Program of the National Beef Cattle Industrial Technology System (CARS-38), and the National 863 Program of China (Grant No. 2013AA102505).

Literature Cited

- Adler M, Anjum M, Berg OG, Andersson DI, Sandegren L. 2014. High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Mol Biol Evol.* 31:1526–1535.
- Bae J, et al. 2010. Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* 11:232.
- Bickhart DM, et al. 2012. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22:778–790.
- Bradley DG, MacHugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A.* 93:5131–5135.
- Cai X, Chen H, Wang S, Xue K, Lei C. 2006. Polymorphisms of two Y chromosome microsatellites in Chinese cattle. *Genet Sel Evol.* 38:525–534.
- Canavez FC, et al. 2012. Genome sequence and assembly of *Bos indicus*. *J Hered.* 103:342–348.
- Chan EK, Nagaraj SH, Reverter A. 2010. The evolution of tropical adaptation: comparing taurine and zebu cattle. *Anim Genet.* 41:467–477.
- Chen H, Qiu H. 1993. Studies on sex chromosome polymorphism of four local cattle (*Bos taurus*) breeds in China. *Hereditas* 15:14–17.
- De Cáceres M, Legendre P, Wiser SK, Brotons L. 2012. Using species combinations in indicator value analyses. *Methods Ecol Evol.* 3:973–982.
- Decker JE, et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc Natl Acad Sci U S A.* 106:18644–18649.
- Decker JE, et al. 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:1–14.
- Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 14:927–930.
- Edwards CJ, et al. 2011. Dual origins of dairy cattle farming—evidence from a comprehensive survey of European Y-chromosomal variation. *PLoS One* 6:e15922.
- Elsik CG, Tellam RL, Worley KC. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324:522–528.
- Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* 320:1629–1631.
- Fadista J, Thomsen B, Holm L-E, Bendixen C. 2010. Copy number variation in the bovine genome. *BMC Genomics* 11:284.
- Gatto L, Catanzaro D, Milinkovitch MC. 2006. Assessing the applicability of the GTR nucleotide substitution model through simulations. *Evol Bioinform Online.* 2:145–155.
- Gibbs RA, et al. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528–532.
- Ginja C, da Gama LT, Penedo MCT. 2009. Y chromosome haplotype analysis in Portuguese cattle breeds using SNPs and STRs. *J Hered.* 100:148–157.
- Ginja C, et al. 2010. Origins and genetic diversity of New World Creole cattle: inferences from mitochondrial and Y chromosome polymorphisms. *Anim Genet.* 41:128–141.
- Götherström A, et al. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc R Soc Lond B Biol Sci.* 272:2345–2351.
- Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol.* 42:5.
- Hou Y, et al. 2011. Genomic characteristics of cattle copy number variations. *BMC Genomics* 12:127.
- Jansen S, et al. 2013. Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. *BMC Genomics* 14:446.

- Jia SG, et al. 2007. Genetic variation of mitochondrial D-loop region and evolution analysis in some Chinese cattle breeds. *J Genet Genomics* 34:510–518.
- Jia SG, et al. 2010. A new insight into cattle's maternal origin in six Asian countries. *J Genet Genomics*. 37:173–180.
- Jiang L, et al. 2012. Genome-wide identification of copy number variations in Chinese Holstein. *PLoS One* 7:e48732.
- Jiang L, et al. 2013. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics* 14:131.
- Kijas JW, et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4:e4668.
- 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.* 38:146–154.
- Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Lei CZ, et al. 2006. Origin and phylogeographical structure of Chinese cattle. *Anim Genet.* 37:579–582.
- Li M, Tian S, et al. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* 45:1431–1438.
- Li R, Xie WM, et al. 2013. Y chromosome diversity and paternal origin of Chinese cattle. *Mol Biol Rep.* 40:6633–6636.
- Li R, Zhang XM, et al. 2013. Paternal origins of Chinese cattle. *Anim Genet.* 44:446–449.
- Liu GE, et al. 2010. Analysis of copy number variations among diverse cattle breeds. *Genome Res.* 20:693–703.
- MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146:1071–1086.
- McCue ME, et al. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet.* 8:e1002451.
- McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DM. 2013. New World cattle show ancestry from multiple independent domestication events. *Proc Natl Acad Sci U S A.* 110:E1398–E1406.
- Metzger J, et al. 2013. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics* 14:487.
- Novembre J, Ramachandran S. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet.* 12:245–274.
- O'Brien AMP, et al. 2014. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol.* 46:19.
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5:557–572.
- Parks DH, Beiko RG. 2010. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26:715–721.
- Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83–87.
- Pollinger JP, et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898–902.
- Qiu H, Qing Z, Chen YC, Wang A. 1988. Bovine breeds in China. Shanghai (China): Shanghai Scientific and Technical Publishers.
- Qu Y, et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun.* 4:2071.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.
- Sambrook J, Russell D. 2001. Molecular cloning: a laboratory manual. Beijing (China): Science Press.
- Stothard P, et al. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 12:559.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10:512–526.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet.* 14:125–138.
- Zhan B, et al. 2011. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics* 12:557.
- Zhang L, et al. 2014. Detection of copy number variations and their effects in Chinese bulls. *BMC Genomics* 15:480.