Luo et al. *Zool. Res.* 2025, 46(3): 518−526
https://doi.org/10.24272/j.issn.2095-8137.2024.442
https://cstr.cn/32248.14.zr.2024.442

ZR | Zoological Research

Article Open Access

Genetic basis and origin of coat color in Leiqiong cattle

Fu-Nong Luo[1,2,#], Shu-Jun Chen[1,2,#], Hojjat Asadollahpour Nanaei[1,3], Xin-Yu Wang[1,2], Rasmus Heller[4], De-Xiang Hu[5], Hong Cheng[6], Jie Li[1,2], Shi-Heng Ni[7], Mao Li[8], Xue-Lei Dai[9,*], Yu Jiang[1,2,*]

[1] *College of Animal Science and Technology*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[2] *Key Laboratory of Livestock Biology*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[3] *College of Life Sciences*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*
[4] *Department of Biology*, *University of Copenhagen*, *Copenhagen* 2200, *Denmark*
[5] *State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan*, *Yunnan University*, *Kunming*, *Yunnan* 650091, *China*
[6] *Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen*, *Shenzhen*, *Guangdong* 518000, *China*
[7] *Animal Husbandry Technology Extending Stations of Hainan Province*, *Haikou*, *Hainan* 570203, *China*
[8] *Tropical Crops Genetic Resources Institute*, *Chinese Academy of Tropical Agricultural Sciences*, *Danzhou*, *Hainan* 571737, *China*
[9] *Yazhouwan National Laboratory*, *Yazhou Bay Science and Technology City*, *Sanya*, *Hainan* 572024, *China*

ABSTRACT

Coat color polymorphism in domestic animals provides a robust framework for elucidating mechanisms of species adaptation, domestication, and genomic diversity. Leiqiong cattle, a representative indicine breed from southern China, are predominantly yellow-coated, although a subset exhibits a solid black phenotype. To determine the genetic basis of this variation, a genome-wide association study (GWAS) was performed in 212 Leiqiong bulls. A pronounced association signal was detected on chromosome 6 within the fifth intron of the *CORIN* gene, providing the first evidence of the potential influence of *CORIN* on bovine coat color variation. Integration of these results with publicly available genomic datasets and haplotype analyses indicated that the yellow coat phenotype is derived from Indian indicine ancestry, whereas the black coat phenotype emerged through introgression from wild bovine lineages and artificial hybridization with Wagyu cattle. Comparative analysis of Indian indicine cattle with divergent coat colors revealed distinct *LEF1* haplotypes within a shared *CORIN* background, suggesting an ancient and complex domestication history underlying coat color variation. These findings provide direct evidence that introgression has shaped phenotypic variation in East Asian cattle and offer novel insights into the genetic architecture of pigmentation, with implications for future breeding strategies.

Keywords: Genome-wide association studies; Indicine cattle; *CORIN* gene; Haplotype diversity; Introgression

INTRODUCTION

Coat color polymorphism serves as a powerful model for investigating the mechanisms of species adaptation, the evolutionary processes involved in domestication, and the genetic architecture underlying phenotypic diversity in both wild and domestic animals (Cieslak et al., 2011). To date, over 150 genes involved in pigmentation have been identified in mammals (Caro & Mallarino, 2020; Hofreiter, 2013). These genes regulate various aspects of pigmentation biology, including pheomelanin and eumelanin production, melanocyte development, melanogenesis, and the intracellular transport and transfer of melanin (Cieslak et al., 2011; Hubbard et al., 2010). The remarkable spectrum of coat color phenotypes observed in modern domestic animals offers a tractable system for dissecting the genetic determinants of morphological variation.

Indicine cattle (*Bos indicus*) were domesticated approximately 8 000 years before present (YBP) in the Indus Valley, located in present-day Pakistan, and subsequently dispersed across diverse geographic regions (Verdugo et al., 2019). Genomic analyses have demonstrated that modern indicine cattle can be broadly classified into three major lineages: Indian indicine, African indicine, and East Asian indicine (Chen et al., 2018a). East Asian indicine cattle are thought to have migrated into China via Southeast Asia, during which they experienced repeated introgression events

Received: 12 December 2024; Accepted: 20 January 2025; Online: 21 January 2025

Foundation items: This work was supported by the National Key R&D Program of China (2022YFF1000100). Shaanxi Innovation Team Project (2022TD-10), Shaanxi Livestock and Poultry Breeding Double-chain Fusion Key Project (2022GD-TSLD-46-0401), and Postdoctoral Research Funding of Hainan Province

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: daixuelei2014@163.com; yu.jiang@nwafu.edu.cn

from other *Bos* species (Chen et al., 2023; Dai et al., 2023). Coat color phenotypes in Indian and African indicine populations are typically dominated by solid yellow (e.g., Sahiwal, Gir) or solid white (e.g., Bhagnari, Dajal) morphs (Iqbal et al., 2019; Kim et al., 2020). In East Asian indicine cattle, yellow remains the predominant coat color (Dai et al., 2023), a trait that is particularly valued in China due to longstanding cultural preferences. Despite substantial research on coloration in taurine cattle, studies on domesticated indicine cattle remain relatively limited (Fan et al., 2014; Hofstetter et al., 2019).

Leiqiong cattle, a distinctive indicine breed native to southern China, have been historically maintained for meat production and are characterized by excellent heat tolerance, resilience to low-quality forage, and robust resistance to disease. The current population exceeds 400 000 individuals annually (China National Commission of Animal Genetic Resources, 2011). This breed typically exhibits the characteristic yellow coat associated with native Chinese cattle, along with hallmark *Bos indicus* features such as prominent humps and thin, pliable skin (Gao et al., 2017). Although the black coat phenotype is uncommon among indigenous Chinese cattle, a subset of Leiqiong individuals displays a uniform black coloration. The presence of this rare phenotype, coupled with a complex breeding history, potentially including hybridization with other breeds or *Bos* species, suggests underlying genetic contributions that have not been fully characterized. This study investigated the genetic basis of coat color variation in Leiqiong cattle and explored the evolutionary origins of their yellow and black pigmentation phenotypes.

MATERIALS AND METHODS

Sample collection

Fresh blood samples were collected from a total of 212 Leiqiong bulls across different regions in Hainan province, including Haikou City (*n*=64), Wenchang City (*n*=40), Ding'an County (*n*=73), Baoting County (*n*=17), and Chengmai County (*n*=18). Each sampling site consisted of small-scale breeding farms housing several hundred Leiqiong cattle. To maximize representativeness, nearly all available adult, uncastrated bulls aged between one and three years were included. In addition to these newly collected samples, whole-genome resequencing data from 641 individuals representing diverse *Bos* species were retrieved from public databases for comparative population genomic analyses. Detailed information on all individuals, including breed, country of origin, geographical location, and sequencing depth, is provided in Supplementary Table S1.

Phenotypic assessment

Photographs of all 212 Leiqiong bulls were captured during sample collection. To quantify coat color as a continuous trait, seven anatomical regions, including the back, hump, abdomen, and head, among others, were scored using a binary scale, assigning 1 for black and 0 for yellow. Phenotypic evaluations were independently conducted by three observers to avoid subjective bias. Final scores were determined by consensus following discussion among evaluators. This scoring approach yielded individual-level trait profiles suitable for principal component analysis (PCA) (Supplementary Table S2). The first principal component (PC1) correlated strongly with the degree of black coloration, with lower scores corresponding to predominantly yellow coats and higher scores reflecting predominantly black coats. The distribution of these phenotypes is shown in Supplementary Figure S1. Additionally, coat color was categorized as a binary trait—yellow or black—for downstream analyses.

Whole-genome sequencing and variant detection

Genomic DNA was isolated from whole blood samples using a standard cetyltrimethylammonium bromide (CTAB) extraction protocol. Paired-end short-insert (150 bp) libraries were constructed and sequenced on the Illumina HiSeq platform (Beijing Genomics Institution, China). Raw short-read data were filtered using Fastp v.0.20.0 (Chen et al., 2018b) with default parameters. Quality-filtered reads were aligned to the bovine reference genome (ARS-UCD2.0) using default settings. Variant calling was conducted using SImgvcf_gpu v.2.0.7 (Zhang et al., 2021). Raw single nucleotide polymorphisms (SNPs) were filtered with BCFtools v.1.1 (Chen et al., 2018b) using the parameters: "-v snps -e 'QD<2.0 | QUAL<30.0 | SOR>3.0 | FS>60.0 | MQ<40.0 | MQRankSum<–12.5 | ReadPosRankSum<–8.0' -m2 -M2". To evaluate the influence of sequencing depth on SNP detection, recall, precision, and concordance were assessed across a range of depths (1× to 13×). At the average sequencing depth of 6.6×, SNP precision exceeded 92%, concordance was greater than 95%, and recall reached approximately 75%. These results indicated that although variant discovery was limited at lower depths, identified SNPs remained highly accurate (Supplementary Figure S2). Validation of SImgvcf_gpu SNP calls was conducted by comparing results from a subset of samples analyzed using the BWA-MEM and GATK pipelines under identical conditions. SNP concordance between the two methods reached 99.9%, confirming the accuracy of SImgvcf_gpu (Supplementary Figure S2). Only biallelic SNPs with a missing genotype rate below 0.2 were retained. Genotype imputation and phasing were simultaneously performed using Beagle v5.4 (Browning et al., 2021) with default parameters. To assess imputation and phasing accuracy, 13 samples with sequencing depths ranging from 9.98× to 13.48× were selected. Six samples were randomly downsampled to 5–6× as the test set, while the remaining seven samples were used as the reference. These samples, together with additional publicly available data, were re-phased using Beagle. Even at 5× depth, phasing accuracy exceeded 95.5% (Supplementary Figure S2). All SNPs were annotated using ANNOVAR (Wang et al., 2010).

Genome-wide association study

Genotype data for 212 Leiqiong cattle were extracted using BCFtools (Danecek et al., 2021). Variants with a minor allele frequency (MAF) below 0.05 were excluded, yielding 25 066 549 SNPs for downstream analysis. Nucleotide diversity was calculated for four groups (*n*=20 per group) using a sliding window approach with a window size of 50 kb (Supplementary Figure S2). In subsequent GWAS analyses, the first eight PCs from genome-wide PCA were used as covariates to correct for population structure effects.

Association testing for the quantitative phenotype (PC1 scores of color phenotype), was performed using the mixed linear model with the leave-one-chromosome-out (MLMA-LOCO) procedure in GCTA v1.94 (Yang et al., 2011). For binary phenotype analysis, logistic regression analysis was performed using the "--logistic" option in PLINK (Purcell et al.,

2007).

To identify association signals potentially driven by multiple causal variants either independent or partially correlated, conditional GWAS analysis was performed by including the top-scoring SNP as a fixed effect, with PC1 as the phenotype, using the "--condition" option in PLINK v.1.9 (Purcell et al., 2007). A genome-wide significance threshold of 1.995×10^{-9} was adopted, derived from Bonferroni correction for multiple testing (0.05/25 066 549).

Assessment of non-additive genetic effects

Non-additive GWAS was performed following previously described methods (Mapel et al., 2024). Briefly, SNP genotypes were recoded under recessive inheritance assumptions for either the reference (R) or alternate (A) allele. Specifically, genotypes were coded as RR=1, RA=0, AA=0 (M1) in the former case and as AA=1, RA=0, RR=0 (M2) in the latter case. Only SNPs with a MAF greater than 0.5% and a frequency of the encoded "1" genotype between 0.5% and 0.95% were retained for association testing.

To further examine deviations from additive genetic effects, a dominance deviation test was conducted using the genotypic model in PLINK (Purcell et al., 2007) (option --linear genotypic). This model tested each SNP independently for evidence of a deviation from additivity in the *CORIN* haplotype region (Supplementary Table S3).

Population structure and phylogenetic analysis

To elucidate the genetic structure and phylogenetic relationships among cattle breeds/populations, PCA, neighbor-joining (NJ) tree construction, and admixture analyses were conducted using autosomal SNPs.

A total of 747 individuals, comprising representatives from four wild *Bos* species and 79 domestic cattle breeds, were included. Prior to analysis, linkage disequilibrium (LD)-based pruning was applied to the genotype data using PLINK v.1.9 with the option "--indep-pairwise 50 10 0.1". PCA was then performed on the pruned dataset using EIGENSOFT v8.0.0 (Patterson et al., 2006). Pairwise genetic distances were calculated with PLINK v.1.9 (--distance-matrix), with the resulting distance matrix used to construct a NJ tree. The final phylogenetic tree was visualized with Interactive Tree of Life (iTOL) (Letunic & Bork, 2016).

Population structure was further investigated using ADMIXTURE v.1.23 (Alexander et al., 2009). Ancestry proportions were estimated by testing *K* values ranging from 2 to 5, enabling the quantification of genome-wide admixture for each individual.

Selective sweep detection and pairwise sequence divergence analysis

To identify potential selection signals associated with color variation in Leiqiong cattle, individuals were divided into two groups based on coat phenotype: solid black and yellow (non-solid black). Selection signatures were evaluated using several metrics, including population differentiation index (F_{ST}), nucleotide diversity (π), and Tajima's *D*. These statistics were calculated using a window size of 40 kb and a step size of 10 kb with VCFtools v.0.1.17 (Danecek et al., 2011), unless otherwise specified.

Nucleotide differences between populations were calculated using an in-house script. Briefly, SNP differences were quantified by comparing each haplotype to all other haplotypes across the dataset, and results were categorized by population group. To estimate divergence times between haplotypes, a mutation rate of 1.26×10^{-8} per site per generation and a generation interval of 6 years were applied (Liu et al., 2006). Although the mutation rate may vary across lineages, including geographically isolated and specifically managed local populations such as Leiqiong cattle, the adopted mutation rate reflects values widely utilized in multiple studies on cattle.

Maximum-likelihood (ML) phylogenetic tree

To determine the potential origins of haplotypes across the *Bos* genus, a ML phylogenetic tree was constructed following our previous study (Dai et al., 2023). Briefly, SNPs located within the target genomic region were extracted from *Bos* samples, and heterozygous SNP sites were encoded using standard International Union of Pure and Applied Chemistry chemical nomenclature (IUPAC) codes. Phylogenetic reconstruction was performed in MEGA v.11.0.10 (Tamura et al., 2021) using the Tamura-Nei model. Tree topology was evaluated with 100 bootstrap replicates to assess node support.

Haplotype visualization and network construction

To characterize haplotype-specific genotype patterns, phased SNPs from the target region were extracted and visualized as a heatmap using an in-house script. Haplotype networks were inferred using the PEGAS package in R (Paradis, 2010) based on pairwise differences. To minimize redundancy, no more than four individuals sharing identical haplotypes within the same population were retained. SNPs with a MAF below 0.05 were excluded from the analysis to ensure robustness.

Analysis of exonic variants

Subsequent to functional annotation, exonic SNPs significantly associated with coat color phenotype were selected for further investigation. Amino acid sequences corresponding to these variants were retrieved from the Ruminant Genome Database (http://animal.omics.pro/code/index.php/RGD) (Fu et al., 2022) to assess cross-species conservation across representative ruminants. Local hydrophobicity of adjacent amino acids was analyzed using ProtScale (https://web.expasy.org/ProtScale/) (Wilkins et al., 1998). Additionally, the amino acid sequences of two types of CORIN (CORIN540H and CORIN540Y) were submitted to SWISS-MODEL for protein structure prediction (https://swissmodel.expasy.org/interactive) (Waterhouse et al., 2018).

RESULTS

***CORIN* underlies coat color variation in Leiqiong cattle**

To elucidate the genetic basis of pigmentation differences in Leiqiong cattle, a GWAS was performed on 212 bulls sampled from different regions, with an average sequencing depth of 6.6× (Supplementary Table S1). Coat color phenotypes were quantified by scoring pigmentation across seven anatomical regions, revealing significant variation in the extent of black fur coverage, although most individuals exhibited predominantly yellow coats. PCA of these phenotypic scores indicated that PC1 explained 59.36% of the total variance (Supplementary Figure S1). Additionally, association tests on all 212 Leiqiong bulls, classifying coats as either black (*n*=38) or yellow (*n*=174), yielded consistent association signals (Figure 1A; Supplementary Figure S3). Both approaches identified a ~30 kb region on chromosome 6 (chr6:66380000–66410000) as

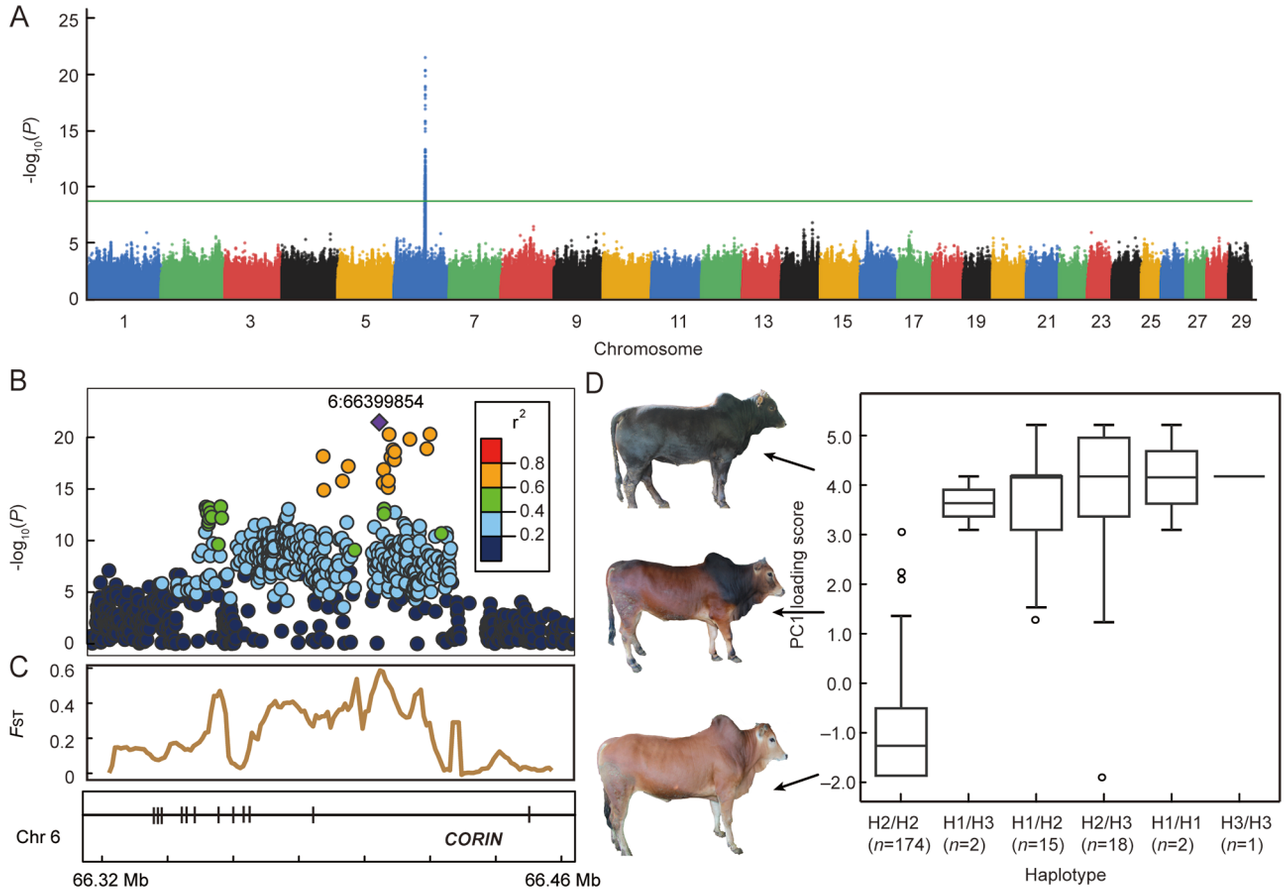

Figure 1 Genome-wide association study of coat color variation in Leiqiong cattle

A: A single genomic region on chromosome 6 exhibited a strong association with coat color, with the most significant SNP located at chr6:66399854 (P=3.56×10^{-22}). B: Magnified view of peak association reveals that the signal is located within the *CORIN* gene (chr6: 66272531–66581941), with r^2 indicating degree of linkage between the lead SNP and surrounding variants. C: Genetic differentiation (F_{ST}) across the significant region. D: Phenotypic variation stratified by *CORIN* haplotypes. Haplotypes H1 and H3 were associated with increased black pigmentation (PC1, Supplementary Figure S1). H2–H2 denotes individuals homozygous for haplotype H2 (as defined in Figure 2A), with the other five genotypes denoted accordingly. Photos by Fu-Nong Luo.

the primary locus of association. The most strongly associated SNP, located at chr6:66399854 A>G (P=3.56×10^{-22}), exhibited an alternative allele frequency of 87.5% in the population, situated within the fifth intron of *CORIN* (Figure 1B). The GWAS signal disappeared when this SNP was included as a fixed effect, suggesting that the signal is likely driven by a single underlying causal variant or haplotype (Supplementary Figure S3).

To identify putative genes associated with the dark black coat color in Leiqiong cattle, population genetic analyses were conducted to detect signatures of selection across the genome. Individuals were stratified by coat color and subjected to genome-wide calculations of the population differentiation index (F_{ST}). A pronounced F_{ST} peak was detected at the *CORIN* locus compared to other genomic regions (Figure 1C; Supplementary Figure S4). Additional evidence of selection at *CORIN* was provided by elevated π-ratio values and shifts in Tajima's D (Supplementary Figure S4). Haplotype heatmap analysis identified three distinct haplotype clusters in the *CORIN* region, designated as H1 (4.95%), H2 (89.86%), and H3 (5.19%) (Figure 2A), which were clearly associated with coat color phenotype. H2 was predominantly associated with yellow coats (lower PC1 values), whereas H1 and H3 were associated black coats (higher PC1 values) (Fisher test P<0.001, odds ratio=6401) (Figure 1D). Notably, most black-coated individuals were heterozygous (e.g., H1/H2, H2/H3), suggesting that H1 and H3 alleles may exhibit dominance over H2. This hypothesis was corroborated by non-additive GWAS and dominance deviation analyses, both of which yielded significant results (P<0.001) (Supplementary Figure S5 and Table S3). The observed genotype frequencies did not deviate from Hardy-Weinberg equilibrium (P=0.098).

To investigate whether any exon variants near the GWAS signal were associated with coat color, SNPs within coding regions were analyzed. However, no significant exonic variants were identified (Supplementary Figure S6), suggesting that the causal mechanism may involve regulatory elements or splicing factors, with potential causative variants requiring further investigation.

Black coat color in Leiqiong Cattle originates from introgression and artificial hybridization

To trace the evolutionary origins of the haplotype groups H1 and H3 associated with the black coat phenotype in Leiqiong cattle, whole-genome sequencing data from 641 previously published *Bos* genomes were analyzed. This dataset included 603 domestic cattle covering 78 breeds and 38 individuals from eight wild *Bos* species (Supplementary Table S1). PCA and phylogenetic reconstructions recovered six major cattle

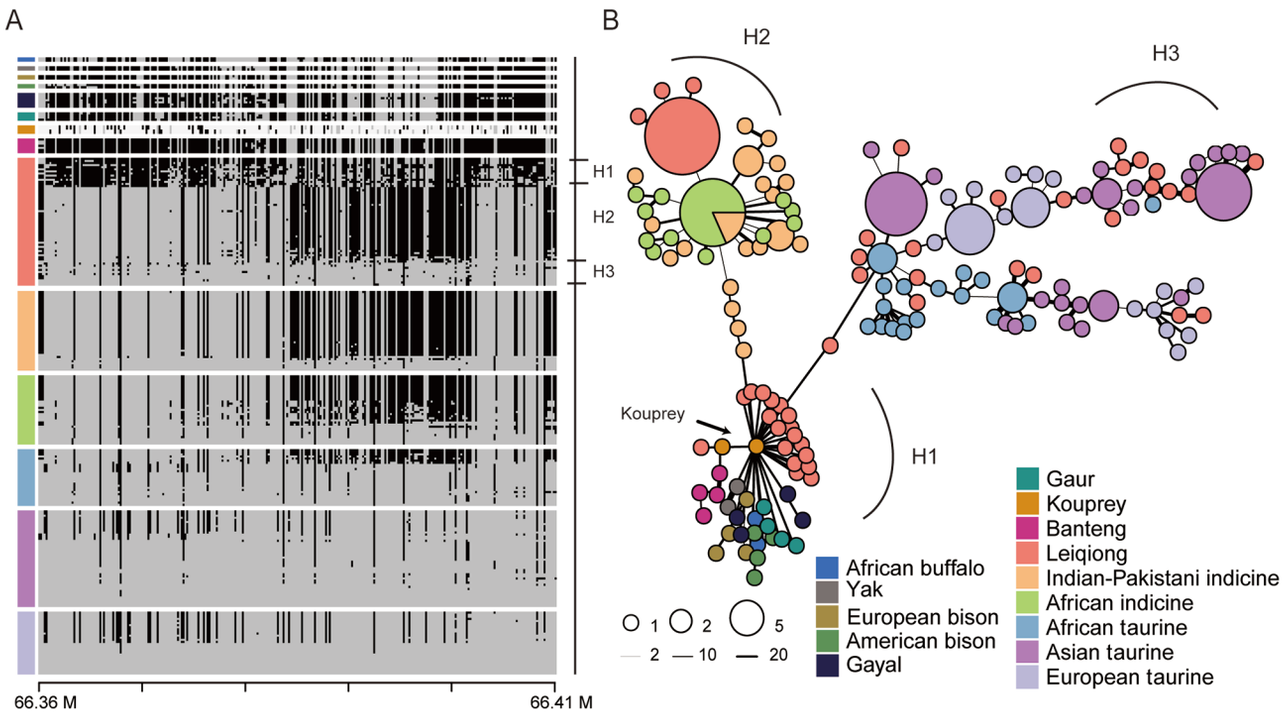

Figure 2 Haplotype pattern at the GWAS-associated loci within *CORIN*

A: Haplotype heatmap of the *CORIN*-associated region (ARS-UCD2.0 chr6:66360000–66410000). Each row represents a phased haplotype, and each column represents a polymorphic SNP variant. Reference and alternative alleles are shown in light gray and black, respectively. B: Haplotype network map of the *CORIN*-associated region (ARS-UCD2.0 chr6:66360000–66410000). Haplotype group H1 was identified as the introgressed haplotype from kouprey.

lineages: European taurine, Asian taurine, African taurine, Indian indicine, African indicine, and Chinese indicine (Supplementary Figure S7). Comparative analysis of *CORIN* haplotypes revealed distinct phylogenetic patterns: H1 grouped with kouprey (*Bos sauveli*), H3 with European taurine cattle, H2 with Indian indicine lineages (Supplementary Figure S8). These relationships were further supported by haplotype network analysis (Figure 2B), which confirmed the clustering of H1 haplotypes from Leiqiong cattle with kouprey, suggesting historical introgression from this now-extinct wild species.

To investigate the origin of H3, the distribution of *CORIN* haplotypes was examined across taurine breeds. While most taurine cattle harbored two major haplotypes, Wagyu cattle exhibited four distinct *CORIN* haplotypes (Figure 3A, B). Among all domesticated breeds, the H3 haplotypes found in Leiqiong cattle were most similar to those in Wagyu, but dissimilar to those in other taurine cattle (Supplementary Figure S9). This close affinity was corroborated by significantly lower pairwise sequence divergence between H3 haplotypes from Leiqiong and Wagyu cattle (P<0.001) (Figure 3C). Additional support for shared ancestry came from genome-wide F_{ST} comparisons, where the strongest signal of differentiation between black-coated Leiqiong and Wagyu individuals also appeared in the *CORIN* region (Figure 3D), suggesting shared mutation patterns in the *CORIN* region. To further evaluate the proportion of Wagyu ancestry in Leiqiong cattle, unsupervised clustering analysis was performed using ADMIXTURE (v.1.3.0). At K=5, Wagyu cattle formed a distinct ancestral component, and the proportion of Wagyu-like ancestry in Leiqiong cattle ranged from 0% to 50.81%. Individuals carrying the H3 haplotype exhibited three-fold higher Wagyu ancestry than those lacking the H3 haplotype (P<0.001) (Supplementary Figure S10). Interestingly, historical records confirm the introduction of Wagyu cattle to Hainan Province in the 1990s to enhance the meat quality and productivity of Leiqiong cattle. These genetic and historical data strongly imply that haplotype group H3 was introduced into Leiqiong cattle through deliberate hybridization with Wagyu cattle, one of the few taurine breeds with black fur. Taken together, the appearance of black coat color in Leiqiong cattle appears to be the result of wild *Bos* introgression and artificial hybridization with Wagyu cattle.

Yellow coat color in Leiqiong Cattle originates from Indian indicine cattle

Haplotype analysis revealed that the H2 haplotype, associated with yellow coat color in Leiqiong cattle, was widely shared with both Indian and African indicine populations (Figure 2). Given the established ancestry of modern-day Pakistani cattle from indicine cattle, we hypothesized that the H2 haplotype originated from Indian indicine cattle. To evaluate this hypothesis, nine Indian indicine breeds were investigated, including five white-coated and four yellow-coated breeds. Comparative analysis of *CORIN* haplotypes revealed that H2 was almost universal across all Indian indicine samples, regardless of coat color (Figure 4A).

Previous studies have identified *LEF1*, a gene implicated in pigmentation regulation, as the most divergent locus among differently colored Indian indicine cattle (Chen et al., 2023; Guenther et al., 2014). Therefore, *LEF1* haplotypes were investigated to further clarify the genetic basis of coat color variation. While both white and yellow Indian indicine cattle shared the same *CORIN* haplotype, they exhibited striking differences in *LEF1* haplotypes (Figure 4B). Two highly divergent LEF1 haplotype groups were identified—designated

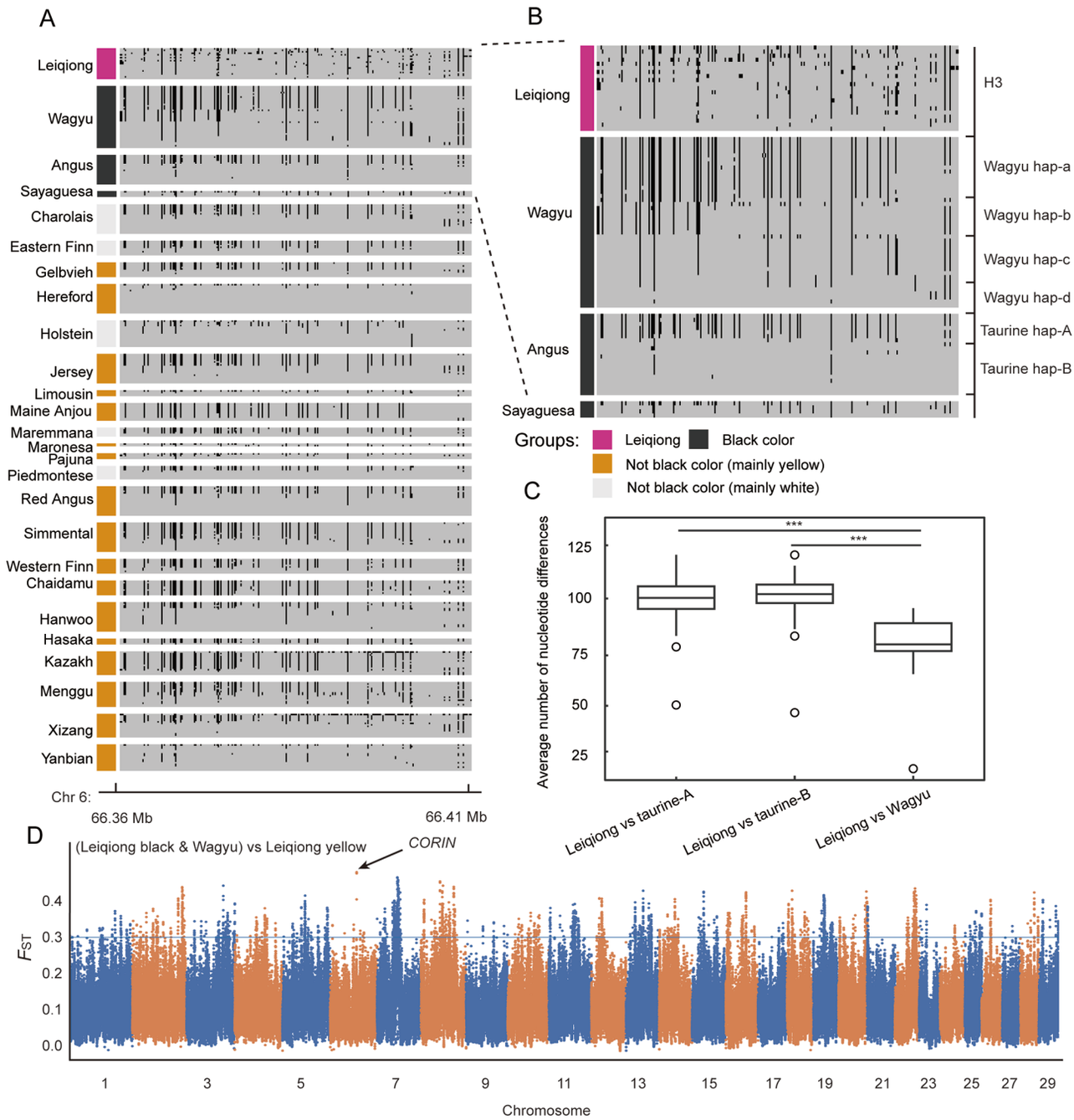

Figure 3 Comparative analysis of *CORIN* haplotype group H3 in Leiqiong cattle and other domestic taurine cattle

A: Haplotype composition across domestic taurine cattle. B: Detailed assessment of haplotype H3 origin, comparing haplotypes between black-coated domestic and Leiqiong cattle. C: Pairwise nucleotide differences showing smaller divergence between Leiqiong and Wagyu cattle than between Leiqiong and other taurine cattle. D: Population differentiation (F_{ST}) between yellow-coated Leiqiong cattle and black-coated cattle (including black Leiqiong and Wagyu individuals). Threshold line represents the top 1% of F_{ST} values. Mann-Whitney *U* test, ***: P<0.001.

LEF1-w (white-associated) and LEF1-y (yellow-associated). The extent of sequence divergence between these haplotypes was substantial, with an estimated nucleotide divergence of 0.61%, corresponding to a divergence time of approximately 2.9 million years (Supplementary Figure S11). This deep divergence suggests that the observed haplotype differences predate cattle domestication and likely reflect ancestral variation within *Bos* lineages. The *LEF1*-y haplotype was widely distributed among European taurine cattle and further subdivided into two distinct subgroups (Figure 4B), indicating that both haplotypes likely coexisted in the early domestication regions of indicine cattle, with *LEF1*-y in *Bos indicus* potentially originating from a common ancestor of *Bos primigenius primigenius* and *Bos primigenius namadicus*, or possibly through introgression from subgroup *LEF1*-y1 of *Bos taurus*. Notably, *LEF1*-w haplotypes were absent in Leiqiong cattle, supporting the inference that the yellow coat color observed in this population was inherited from yellow-coated Indian indicine cattle.

DISCUSSION

Coat color is one of the most visible and historically valued phenotypic traits in domestic animals, often used to distinguish

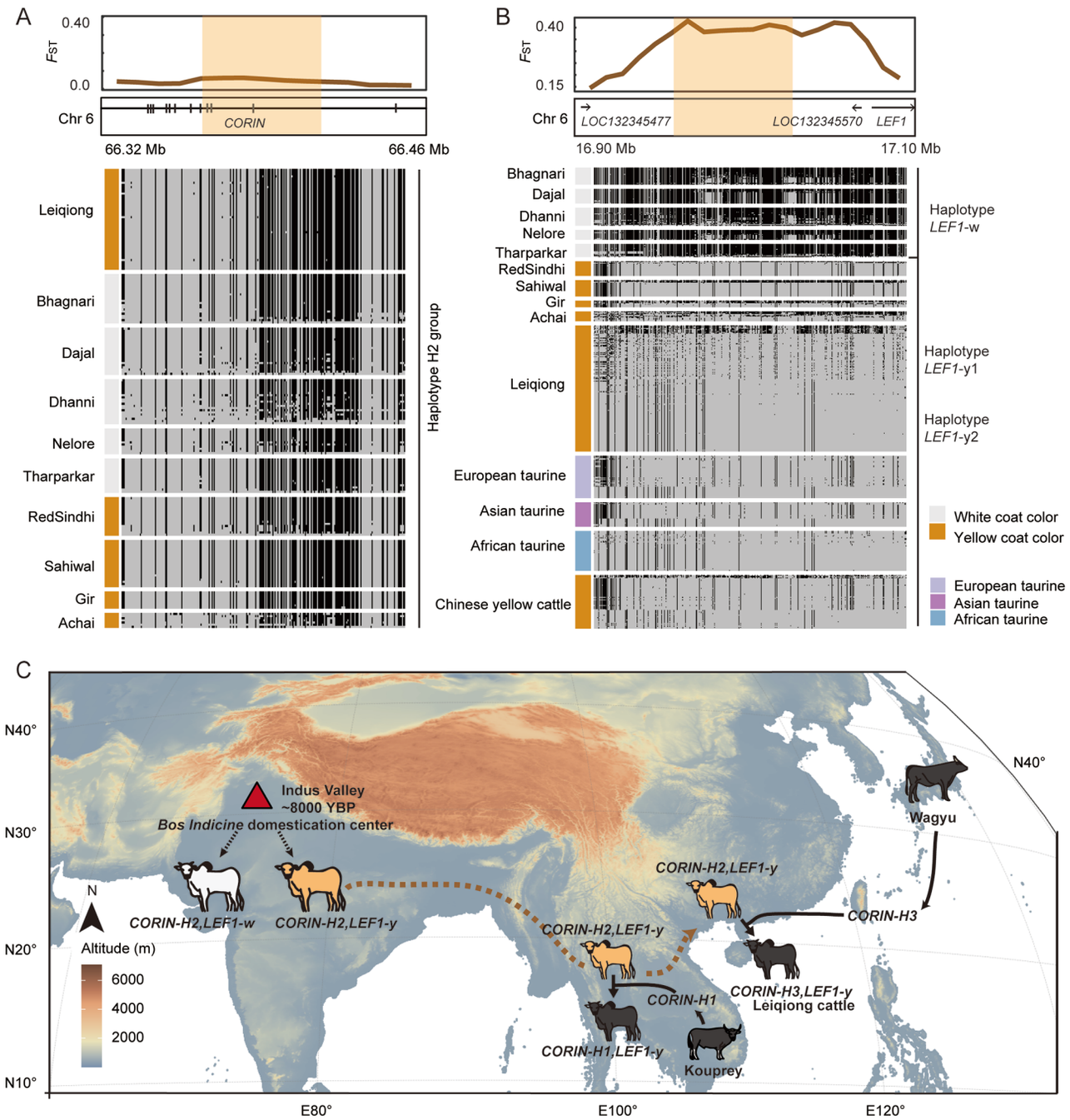

Figure 4 Proposed origins of the yellow coat phenotype in Leiqiong cattle

A: Haplotype heatmap of the *CORIN* region in Leiqiong and Indian indicine cattle with varying coat colors. The yellow-associated haplotype (H2) was present across indicine cattle regardless of pigmentation. B: Haplotype heatmap of the *LEF1* region in different cattle breeds worldwide. *LEF1*-w denotes the haplotype group found in white-coated cattle, *LEF1*-y1 and *LEF1*-y2 denote the haplotype group associated with yellow-coated cattle. F_{ST} values in panels A and B represent comparisons between yellow and white indicine cattle. C: Schematic representation of inferred evolutionary origins of yellow and black coat colors in Leiqiong cattle.

breeds and guide selective breeding. In this study, GWAS, selective sweep analysis, and haplotype-based approaches were applied to map the genetic diversity associated with coat color in Leiqiong cattle. Furthermore, the evolutionary origins of the black and yellow coat phenotypes were investigated, revealing distinct introgression and domestication histories in these cattle.

Chinese indicine cattle, also referred to as East Asian indicine, underwent multiple waves of introgression from other *Bos* species during their dispersal into China. These events likely involved gene flow from species such as gaur (*Bos* gaurus), gayal (*Bos* frontalis), banteng (*Bos* javanicus), and kouprey (*Bos* sauveli) (Chen et al., 2023; Dai et al., 2023; Zhang et al., 2020). The most recent major introgression event occurred approximately 3 000 years ago, likely originating from a source similar to banteng (Dai et al., 2023). Many of these introgression signals are believed to have facilitated the rapid adaptation of indicine cattle to humid regions (Chen et al., 2018a, 2023; Dai et al., 2023; Xia et al., 2023). For example, alleles at pigmentation-related loci such as agouti

signaling protein (*ASIP*) are hypothesized to have been introgressed in East Asian indicine cattle from banteng-like or kouprey-like ancestral populations, potentially contributing to local adaptation through modifications in coat color (Chen et al., 2018a; Dai et al., 2023). However, such hypotheses have remained largely speculative, lacking direct evidence linking introgression to specific phenotypic outcomes in modern East Asian indicine cattle.

In this context, the present study provides the first GWAS evidence demonstrating a phenotypic consequence of introgression in East Asian cattle. Notably, a haplotype segment of the *CORIN* gene—originally derived from kouprey, a now-extinct wild *Bos* species characterized by a black coat (Sinding et al., 2021) —was shown to be introgressed into Leiqiong cattle. This segment was significantly associated with the black coat phenotype. *CORIN* encodes a serine peptidase expressed in hair follicles and functions as a downstream suppressor of ASIP (Enshell-Seijffers et al., 2010). Loss-of-function mutations in *CORIN* have been associated with enlarged pheomelanin bands and lighter pelage in tigers (Xu et al., 2017) and mice (Enshell-Seijffers et al., 2008).

Leiqiong cattle represent a geographically and genetically distinct breed of East Asian indicine. Yellow is the predominant coat color in this breed, aligning with regional cultural preferences across East Asia. Given that the black pigmentation arises from introgression, it is likely that yellow represents the ancestral coat color of East Asian indicine cattle. While indicine cattle in South Asia commonly display both yellow and white pigmentation (Iqbal et al., 2019; Kim et al., 2020), the timing and route of their dispersal into China suggest that the founder populations were likely yellow-coated. Our comparative haplotype analysis of Indian indicine breeds supports this view. Importantly, the partial expression of black pigmentation in Leiqiong cattle appears to be sex-linked, occurring exclusively in bulls. This pattern mirrors sexual dimorphism in pigmentation observed in several wild *Bos* species. To avoid sex-related confounding effects, only male individuals were included in this study.

In conclusion, our findings suggest that coat color variation in Leiqiong cattle reflects a combination of ancient domestication processes and more recent gene flow from wild and domestic sources. Prior to the domestication of indicine cattle, ancestral aurochs in the Indus Valley likely exhibited a spectrum of coat colors. Cultural selection may have favored white individuals, leading to their persistence in some modern indicine populations. In contrast, yellow-coated aurochs —sharing haplotypes with ancestral taurine cattle—were domesticated and later introduced into East Asia, where they contributed to the genetic foundation of Chinese indicine cattle (Figure 4C). This study provides direct evidence linking introgression to phenotypic evolution and offers novel insights into the complex demographic and selective history of coat color variation in cattle.

DATA AVAILABILITY

Raw data were deposited in the National Center for Biotechnology Information database (BioProjectID PRJNA1194778), Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa), (accession number PRJCA034876), and Science Data Bank (https://www.scidb.cn/en) (DOI:10.57760/sciencedb.19821). The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

F.N.L. and S.J.C. conducted comprehensive data analysis, interpreted the results, and prepared the first draft of the manuscript. D.X.H., H.C., and X.L.D. assisted in data analysis, validated data accuracy, and contributed to data visualization. F.N.L., S.H.N., M.L., X.Y.W. and J.L. collected the samples. H.A.N. and R.H. reviewed and edited the manuscript, improving clarity and coherence. F.N.L., X.L.D., and Y.J. designed the research framework and supervised the entire project. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the High-Performance Computing Platform of Northwest A&F University and Computing Center in Xi'an for providing computing resources for bioinformatic analyses.

REFERENCES

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

Browning BL, Tian XW, Zhou Y, et al. 2021. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, **108**(10): 1880–1890.

Caro T, Mallarino R. 2020. Coloration in mammals. *Trends in Ecology & Evolution*, **35**(4): 357–366.

Chen NB, Cai YD, Chen QM, et al. 2018a. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, **9**(1): 2337.

Chen NB, Xia XT, Hanif Q, et al. 2023. Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. *Nature Communications*, **14**(1): 7803.

Chen SF, Zhou YQ, Chen YR, et al. 2018b. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.

China National Commission of Animal Genetic Resources. 2011. Animal Genetic Resources in China: Bovines. Beijing: China Agriculture Press, 111–114. (in Chinese)

Cieslak M, Reissmann M, Hofreiter M, et al. 2011. Colours of domestication. *Biological Reviews*, **86**(4): 885–899.

Dai XL, Bian PP, Hu DX, et al. 2023. A Chinese indicine pangenome reveals a wealth of novel structural variants introgressed from other *Bos* species. *Genome Research*, **33**(8): 1284–1298.

Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.

Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008.

Enshell-Seijffers D, Lindon C, Morgan BA. 2008. The serine protease Corin is a novel modifier of the agouti pathway. *Development*, **135**(2): 217–225.

Enshell-Seijffers D, Lindon C, Wu E, et al. 2010. β-Catenin activity in the dermal papilla of the hair follicle regulates pigment-type switching. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(50): 21564–21569.

Fan YP, Wang P, Fu WX, et al. 2014. Genome-wide association study for pigmentation traits in Chinese Holstein population. *Animal Genetics*, **45**(5): 740–744.

Fu WW, Wang R, Nanaei HA, et al. 2022. RGD v2.0: a major update of the ruminant functional and evolutionary genomics database. *Nucleic Acids Research*, **50**(D1): D1091–D1099.

Gao YH, Gautier M, Ding XD, et al. 2017. Species composition and

environmental adaptation of indigenous Chinese cattle. *Scientific Reports*, **7**(1): 16196.

Guenther CA, Tasic B, Luo LQ, et al. 2014. A molecular basis for classic blond hair color in Europeans. *Nature Genetics*, **46**(7): 748–752.

Hofreiter M. 2013. Genetic basis and evolutionary causes of colour variation in vertebrates. *Seminars in Cell & Developmental Biology*, **24**(6-7): 515.

Hofstetter S, Seefried F, Häfliger IM, et al. 2019. A non-coding regulatory variant in the 5′-region of the MITF gene is associated with white-spotted coat in Brown Swiss cattle. *Animal Genetics*, **50**(1): 27–32.

Hubbard JK, Uy JAC, Hauber ME, et al. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in Genetics*, **26**(5): 231–239.

Iqbal N, Liu X, Yang T, et al. 2019. Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. *PLoS One*, **14**(4): e0215065.

Kim K, Kwon T, Dessie T, et al. 2020. The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics*, **52**(10): 1099–1110.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, **44**(W1): W242–W245.

Liu GE, Matukumalli LK, Sonstegard TS, et al. 2006. Genomic divergences among cattle, dog and human estimated from large-scale alignments of genomic sequences. *BMC Genomics*, **7**: 140.

Mapel XM, Kadri NK, Leonard AS, et al. 2024. Molecular quantitative trait loci in reproductive tissues impact male fertility in cattle. *Nature Communications*, **15**(1): 674.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**(3): 419–420.

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

Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3): 559–575.

Sinding MHS, Ciucani MM, Ramos-Madrigal J, et al. 2021. Kouprey (*Bos sauveli*) genomes unveil polytomic origin of wild Asian *Bos*. *Iscience*, **24**(11): 103226.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, **38**(7): 3022–3027.

Verdugo MP, Mullin VE, Scheu A, et al. 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science*, **365**(6449): 173–176.

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

Waterhouse A, Bertoni M, Bienert S, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, **46**(W1): W296–W303.

Wilkins MR, Gasteiger E, Bairoch A, et al. 1998. Protein identification and analysis tools in the ExPASy server. *In*: Link AJ. 2-D Proteome Analysis Protocols. Totowa: Humana, 531–552.

Xia XT, Zhang FW, Li S, et al. 2023. Structural variation and introgression from wild populations in east Asian cattle genomes confer adaptation to local environment. *Genome Biology*, **24**(1): 211.

Xu X, Dong GX, Schmidt-Küntzel A, et al. 2017. The genetics of tiger pelage color variations. *Cell Research*, **27**(7): 954–957.

Yang JA, Lee SH, Goddard ME, et al. 2011. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, **88**(1): 76–82.

Zhang K, Lenstra JA, Zhang S, et al. 2020. Evolution and domestication of the Bovini species. *Animal Genetics*, **51**(5): 637–657.

Zhang Q, Liu H, Bu F. 2021. High performance of a GPU-accelerated variant calling tool in genome data analysis. BioRxiv, doi: https://doi.org/10.1101/2021.12.12.472266.