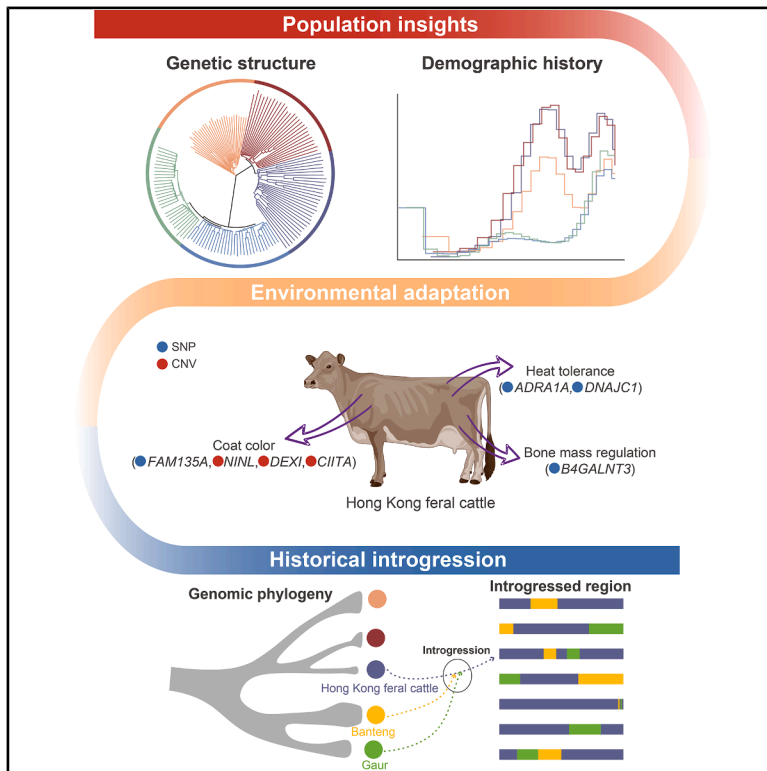


The genome landscape of Hong Kong feral cattle as a unique genetic resource

Graphical abstract



Authors

Xiaoyu Luo, Xiaoran Lu, Yan Ren, ...,
Chuzhao Lei, Ningbo Chen, Wai Yee Low

Correspondence

ningbochen@nwafu.edu.cn (N.C.),
wai.low@adelaide.edu.au (W.Y.L.)

In brief

Phylogenetics; Genomic analysis;
Genomics

Highlights

- HKF cattle represent a unique gene pool of indigenous East Asian populations
- HKF cattle harbor substantial yet untapped genetic potential
- Introgression shaped HKF cattle genomes before splitting from other East Asian populations
- Selected genomic loci associated with environmental adaptation were found in HKF cattle



Article

The genome landscape of Hong Kong feral cattle as a unique genetic resource

Xiaoyu Luo,^{1,7} Xiaoran Lu,^{1,7} Yan Ren,² Yifan Cao,¹ Xuewei Liu,¹ Mario Barbato,^{3,4} Paolo Ajmone-Marsan,⁴ John L. Williams,^{2,4} Rick Tearle,² Richard A.L. Brown,⁵ Michael P. Reichel,^{5,6} Chuzhao Lei,¹ Ningbo Chen,^{1,*} and Wai Yee Low^{2,8,*}

¹Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China

²Davies Livestock Research Centre, School of Animal and Veterinary Sciences, The University of Adelaide, Roseworthy, SA 5371, Australia

³Department of Veterinary Sciences, Università degli Studi di Messina, 98168 Messina, Italy

⁴Department of Animal Science, Food and Technology–DIANA, Università Cattolica del Sacro Cuore, 29122 Piacenza, Italy

⁵College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong 999077, China

⁶Department of Population Medicine and Diagnostic Sciences, Cornell University College of Veterinary Medicine, Ithaca, NY 14853, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: ningbochen@nwfau.edu.cn (N.C.), wai.low@adelaide.edu.au (W.Y.L.)

<https://doi.org/10.1016/j.isci.2026.115465>

SUMMARY

In the Hong Kong Special Administrative Region of China, there is a feral cattle population that has not been well characterized genetically. In this study, we used high-coverage (~30×) whole-genome sequencing from 30 Hong Kong feral (HKF) cattle and compared them to 116 individuals from four representative populations worldwide. Our analyses revealed that the HKF cattle have high genetic diversity in the face of a declining effective population size, suggesting their substantial and yet untapped genetic potential. We also identified introgression events that occurred prior to the divergence between HKF cattle and other East Asian indicine populations, which shaped the adaptation of HKF cattle in Asian agro-ecologies. Moreover, we identified positive selection in HKF cattle for environmental adaptation, particularly in traits related to heat tolerance, bone strength, and coat color. Our findings provide insights into the genetic origin and unique adaptation of HKF cattle.

INTRODUCTION

Cattle are one of the main livestock species worldwide and have played an important role in agriculture and society since their domestication. They were introduced into new regions through human migrations, where exposure to novel ecological conditions imposed strong selective pressures on their genomes over a brief period.¹ The evolutionary consequences of these pressures were further shaped by human management. Populations under intensive management often exhibited restricted genetic and phenotypic diversity, whereas unmanaged populations typically retained higher diversity and were better adapted to local environments.²

In the Hong Kong Special Administrative Region of China, there are about 1,000 feral cattle³ that inhabit urban environments, which represents a unique context for feral populations worldwide, but little is known about their genetic background and adaptive changes. Previous studies indicated that the early geographical history of Hong Kong region was attributed to the Lingnan region in southern China.⁴ In this region, cattle likely arrived by a coastal route during the eastward spread of domesticated indicine cattle.⁵ Subsequent admixture with local wild or

domesticated banteng and gaur led to increased genetic diversity.^{5–7} Due to the effects of geographical isolation, the Hong Kong cattle population may be distinct from mainland populations. In the 1970s, rapid urbanization and economic development in Hong Kong led to the abandonment of cattle farming, and as a result, the Hong Kong cattle started to roam freely and are now considered a feral population. They are therefore exposed to disease, have free mate choice, and other pressures associated with surviving in natural and sometimes urban environments. Interestingly, despite being a small population, the Hong Kong feral (HKF) cattle displayed high phenotypic diversity in terms of horn shape and orientation, coat color, hump size and shape, tail length, and conformation. Hence, the HKF population offers a unique opportunity to study the effects on the genome and to characterize the needs and diversity of this unique population.

One previous study based on single-nucleotide polymorphism (SNP) array data showed no evidence of gene flow from any of the Asian local breeds to the HKF cattle.⁸ Meanwhile, a study based on Y chromosome and mitochondrial DNA data reported the pattern of introgression of the indicine genome across southern China.⁵ Notably, both studies support the introgression from



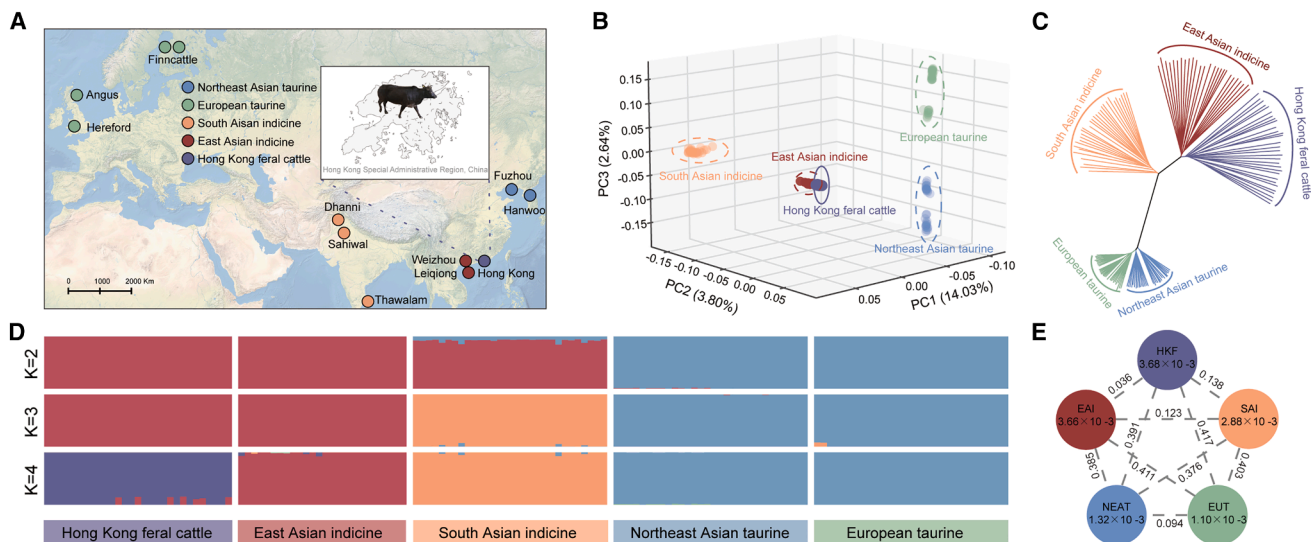


Figure 1. Population structure and relationships of HKF cattle to other cattle populations

(A) Geographic map of sample locations and other cattle populations.

(B) PCA of selected cattle populations.

(C) NJ tree of the relationships between the cattle populations.

(D) The ancestral components inferred using ADMIXTURE for K from 2 to 4.

(E) Genetic diversity and population differentiation across five populations. The values inside the circles represent nucleotide diversity, while the values on the connecting lines between populations indicate the population F_{ST} .

wild bovine species into the genome of Asian and HKF cattle. The findings of these studies indicate that the evolutionary history of HKF cattle has been significantly influenced by introgression. Furthermore, due to biased sampling, different methodological perspectives, and limited genetic material, previous studies have not examined the genetic history and adaptive evolution of HKF cattle. In this study, by analyzing whole-genome sequencing (WGS) data of 30 HKF cattle, we revealed the genetic origin, introgression from wild species, adaptive evolution, and the genetic basis underlying the phenotypic characteristics of HKF cattle. Our findings provide insights into the genetic history of HKF cattle and enhance our understanding of adaptive evolution in feral cattle.

RESULTS

Genome sequencing and SNP variant calling

Individual genomes of 30 HKF cattle were sequenced to an average $29.89\times$ coverage (Table S1). Genotypes for these were jointly called with publicly available genomes of four cattle populations, including European taurine (Hereford, Angus, and Finncattle), Northeast Asian taurine (Hanwoo and Fuzhou), East Asian indicine (EAI) (Weizhou and Leiqiong), and South Asian indicine (SAI) (Sahiwal, Thawalam, and Dhanni) (Figure 1A). In total, all 146 samples were aligned to the *Bos taurus* reference assembly ARS-UCD1.2 with an average mapping rate of 99.50%. Variants were called and filtered using a customized pipeline (see the STAR Methods section), resulting in the retention of ~ 57.06 million bi-allelic SNPs. Functional annotation revealed that the majority of SNPs were located in intergenic (58.59%) and intronic regions (38.23%), while the rest were in the regions

upstream and downstream (1.25%) of open reading frames and untranslated regions (1.01%). Exons contained 0.92% of the SNPs, with 496,136 synonymous SNPs and 11,957 nonsynonymous SNPs (Table S2; Figure S1).

In the HKF cattle population, we compared the SNPs detected between WGS and SNP50k array data.⁸ After pruning for missing data, we retained 33,720 SNPs in the array, of which 16,334 were monomorphic whereas 17,386 are low-frequency variants (minor allele frequency [MAF] ≤ 0.05) (Table S3). In contrast, WGS captured genetic variation comprehensively across the allele frequency spectrum. For comparison with the WGS data, 30,813 variants could be lifted over to ARS-UCD1.2 coordinates. A total of 15,614 SNPs were common to both datasets. We then calculated heterozygosity (H_o) and inbreeding coefficient (F) for the WGS data and compared them with the published full SNP panel ($H_o = 0.144 \pm 0.01$, $F = 0.57 \pm 0.03$) and ancestral SNP panel ($H_o = 0.268 \pm 0.02$, $F = 0.35 \pm 0.05$).² Our WGS data showed higher heterozygosity ($H_o = 0.288 \pm 0.02$) and lower inbreeding coefficient ($F = 0.25 \pm 0.05$).

Genetic structure and differentiation

To explore the genetic relationships among HKF and other cattle populations, we performed principal-component analysis (PCA), neighbor-joining (NJ) tree, and ADMIXTURE analyses of the autosomal SNP genotype data. Principal component (PC)1 explained 14.03% of total variance and distinguished *B. taurus* from *B. indicus*, while the second and third PCs, which accounted for 2.64% and 3.80% of total variance, clustered HKF with EAI and distinguished HKF from SAI (Figures 1B, S2A, and S2B). Similar results were observed in the NJ tree (Figure 1C). To further investigate the ancestry of the populations, we used

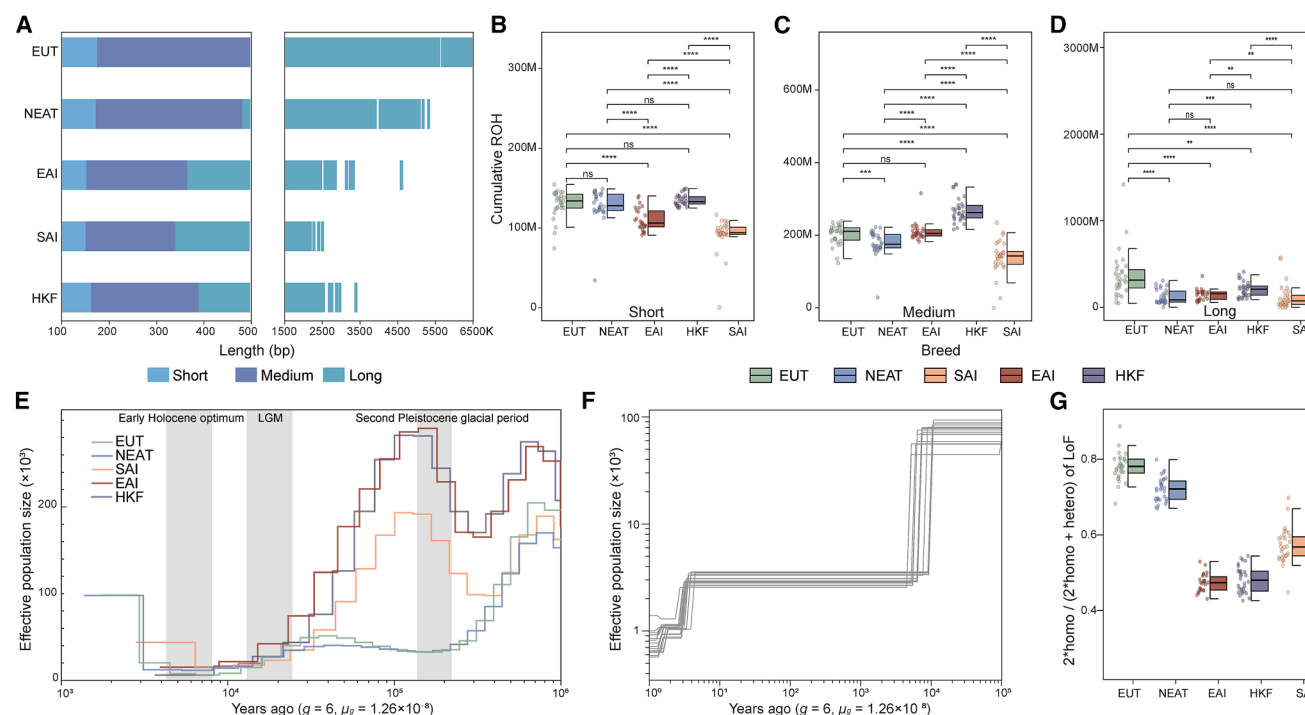


Figure 2. Characterization of inbreeding and demographic history in the HKF population

(A–D) Distribution of ROH segments ordered and colored by size class in all populations. Cumulative size of short (B), medium (C), and long ROH (D) segments across all cattle populations based on the Gaussian mixture model. Differences between populations were tested by Wilcoxon rank-sum tests. Asterisks denote * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

(E) Demographic history inferred from the MSMC2 method. The large gray-shaded boxes illustrate the early Holocene Optimum, the last glacial maximum (LGM), and the second Pleistocene glacial period.

(F) Effective population sizes for the HKF population were inferred using SMC++,¹⁵ computed on 15 groups of 10 randomly selected samples.

(G) The ratio of homozygous LoF alleles in all populations.

ADMIXTURE to calculate delta K for each K from 1 to 4, where K represents the assumed number of ancestral populations. $K = 2$ divided *B. taurus* and *B. indicus* ancestry; when $K = 3$, HKF shared genome ancestry with EAI; and when $K = 4$, HKF showed distinct genetic patterns compared to the other four populations (Figure 1D). In addition to autosomal analysis, we examined the Y chromosome to investigate the male histories of HKF cattle. The Y3 haplogroup belongs to *B. indicus*, and the HKF populations shared the Y3A sub-haplogroup with the EAI populations, whereas the sub-haplogroup Y3B was predominantly found in SAI (Figure S3). Consistent with SNP array results showing HKF as a distinct cluster with highly divergent SAI ancestry, our WGS analysis confirmed its close genetic relationship with EAI ancestry.

Fixation index (F_{ST}) also showed that HKF cattle had the greatest similarity to EAI compared with the other populations (Figure S4). We calculated nucleotide diversity (π) of the populations to measure the degree of polymorphism (Figure 1E). HKF cattle showed the highest level of nucleotide diversity, followed closely by EAI.

Inbreeding and demographic history of HKF cattle

Inbreeding in HKF populations was evaluated by screening runs of homozygosity (ROH) across the genome. In terms of population structure, short, medium, and long ROH reflect ancient

haplotypes predating continental migration, background relatedness, and recent parental relatedness, respectively.^{9,10} Among the three populations with indicine ancestry (EAI, SAI, and HKF), the HKF cattle had the most ROH, regardless of the length of ROH analyzed (Figures 2A–2D). The European taurine (EUT), which has experienced reproductive barriers and intensive artificial selection, had the highest number of cumulative long ROH among all five populations (Figure 2D). The number of long ROH in HKF was second only to the inbred EUT, indicating that HKF has a greater inbreeding risk than other indicine populations. Analysis of long ROH specific to HKF revealed a total of 270 genes, including a region on BTA13 (22.66–23.16 Mb) harboring genes associated with energy balance (*SKIDA1*¹¹), growth traits (*MLLT10*¹²), and heat resistance (*DNAJC1*^{13,14}). Haplotype analysis revealed distinct patterns relative to other groups, indicating potential genetic differentiation and selective pressure in HKF cattle (Figure S5). Cumulative ROH length was highly correlated with ROH number in all three ROH length classes (short, medium, and long; $R > 0.9$; $p < 0.01$) (Figure S6).

The HKF had the highest medium ROH among the populations studied (Figure 2C), which suggests that its effective population size is small. To further investigate the changes in effective population size over time in the HKF population, we estimated the historical N_e using the multiple sequentially Markovian coalescent (MSMC2) method,¹⁶ which revealed that the HKF

experienced a great population decline at 20–30 kya (Figure 2E). The population size history of HKF was similar to that of EAI but differed from that of the taurine populations. The recent effective population size of HKF was estimated using SMC++,¹⁵ which revealed that HKF N_e was lowest 7–9 kya, after which the population size remained constant (Figure 2F).

As the HKF cattle have a high level of inbreeding, loss-of-function (LoF) mutations may have accumulated at high frequency in this population, which is concerning for conservation genetics. The LoF mutations were screened across the entire genome of the five populations to investigate putatively derived deleterious mutations. The HKF cattle had a lower homozygous deleterious mutation load than SAI (Figure 2G).

Historical introgression

Previous studies have reported introgression from banteng and gaur into HKF cattle,⁸ and suggested admixture from banteng and gaur in EAI cattle.⁵ These findings indicated a significant contribution of wild cattle to the indicine populations in East Asia. In this study, SNPs derived from high-coverage WGS were used to calculate D statistics¹⁷ and RFMix¹⁸ to verify and quantify introgression from banteng and gaur into HKF cattle. The results showed that the proportions of banteng and gaur ancestries ranged from 10.74% to 11.81% and from 9.70% to 11.08%, respectively, in the HKF genomes. The proportions of banteng and gaur ancestries in the HKF genome were greater than those in the EAI genome (Figures 3A and 3B; Tables S4 and S5). We further explored the historical gene flow and introgression events using qpgraph and TreeMix.^{17,19} When testing without admixture, the topologies based on qpgraph and TreeMix were consistent with the topology of the NJ tree (Figures S7 and S8), showing that the common ancestors of HKF and EAI cattle diverged from SAI cattle. When modeling one admixture event, qpgraph and TreeMix analysis indicated an introgression from banteng and gaur into the common ancestor of HKF and EAI cattle (Figures 3C and S8). We therefore explored the possible introgression times from banteng and gaur into HKF cattle using fastsimcoal.²⁰ The estimated banteng introgression occurred ~8.9 kya (95% confidence interval [CI]: 6.3–22.2 kya) with an admixture fraction of ~10% in HKF cattle, while gaur introgression was estimated to have occurred ~9.9 kya (95% CI: 5.8–22.7 kya) with an admixture fraction of ~9% (Figure 3D). This is consistent with the qpgraph, suggesting that banteng and gaur introgression occurred prior to the divergence of HKF from EAI cattle.

Effects of introgression on adaptation

To evaluate the effects of introgression on adaptation of HKF cattle, we used the $U50$ statistic to detect adaptive introgressed genes. We found 1,403 genes in the HKF genomes to be of banteng origin and 1,553 genes to be of gaur origin, with 1,234 genes shared by both banteng and gaur (Figures S9A and S9B). These overlapping genes were highly enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) pathways (corrected p value <0.05) of oxygen binding, transport and carrier activity (*HBM*, *MB*, *HBQ1*, *HBA*, *HBZ*, and *HBA1*), and bitter taste receptor activity (*TAS2R7*, *TAS2R16*, *T2R12*, *TAS2R8*, and *TAS2R38*) (Table S6). In

addition, we observed that the most introgressed region on BTA9 (*FAM135A*) harbored genes relevant to coat color,²¹ lipase activity, and intramuscular fat deposition.^{22,23} The geographic distribution of haplotypes in global cattle populations revealed that the introgressed gaur and banteng haplotypes had the highest frequency in HKF and EAI cattle. This finding was further supported by the F_{ST} analysis (Figure S9C), phylogenetic analysis (Figure S9E), and the extent to which the haplotypes were shared across populations (Figure S9D).

Adaptive selection of HKF cattle to environmental stresses

Our genome-wide analysis showed that HKF cattle exhibit a genetic pattern that is distinct from other cattle. To identify signatures of positive selection in the feral population, we compared the genomes of HKF cattle with other populations of cattle using the D_i statistic.²⁴ A total of 653,861 SNPs (0.1%) had significant allele frequency differences, which may be targets of local adaptation (Figure 4A). Among these, strong signals of differentiation were obtained in the regions containing genes related to neurological development (*TRIO*),²⁵ thyroid development (*PDE8B*),²⁶ and autoimmune regulation (*WDR41*).²⁷ A significant selection signature was observed for *ADRA1A* gene on BTA8 (Figure S10), which is associated with efficient recovery from heat stress.²⁸ The *ADRA1A* gene had the strongest differentiation between HKF and EAI cattle for the F_{ST} and θ_{π} ratios (Figure 4B). Evidence of selection on this gene in HKF cattle was provided by high F_{ST} values and low genetic diversity (Figure 4C). In addition, we detected 12 D_i -SNPs in the *ADRA1A* gene, which are putative functional polymorphisms (Figure 4D; Table S7). These variants affect the enhancer and promoter activity of *ADRA1A*, thereby potentially influencing its expression. Notably, the SNP rs517553622 (C > T) located in the promoter of *ADRA1A* was predicted to enhance the binding of KLF17 transcription factor, which is involved in the regulation of immune responses (Table S8).²⁹ Another strong selection signature was located on BTA5, covering the *B4GALNT3* gene, which contributes to bone health and supports survival in the wild (Figure 4A).³⁰ It had high F_{ST} between HKF and other indicine cattle, as well as a likely recent selection signal in HKF cattle seen from the θ_{π} analysis (Figure 4E). The haplotypes of the *B4GALNT3* gene in HKF cattle are distinct from other cattle populations (Figures 4F and 4G).

Copy number variation calling and identification of copy number variation regions

Our mapped sequence data had an average depth of 29.89× per sample, which was sufficient for detecting copy number variations (CNVs) based on read depth.³¹ After CNV calling and filtering, a total of 8,120 CNV regions (CNVRs) were identified on autosomes (Table S9), which consisted of 6,371 deletions and 1,749 duplications compared with the ARS-UCD 1.2 reference genome (Figure 5A). KEGG and GO analyses revealed enrichment of genes involved in biological processes relevant to olfactory receptor activity, the immune system, and response to autoimmune thyroid processes (corrected p value <0.05) (Figure 5B; Table S10). Among these, a set of genes associated with the immune response, including *BLA-DQB*, *TNFSF10*,

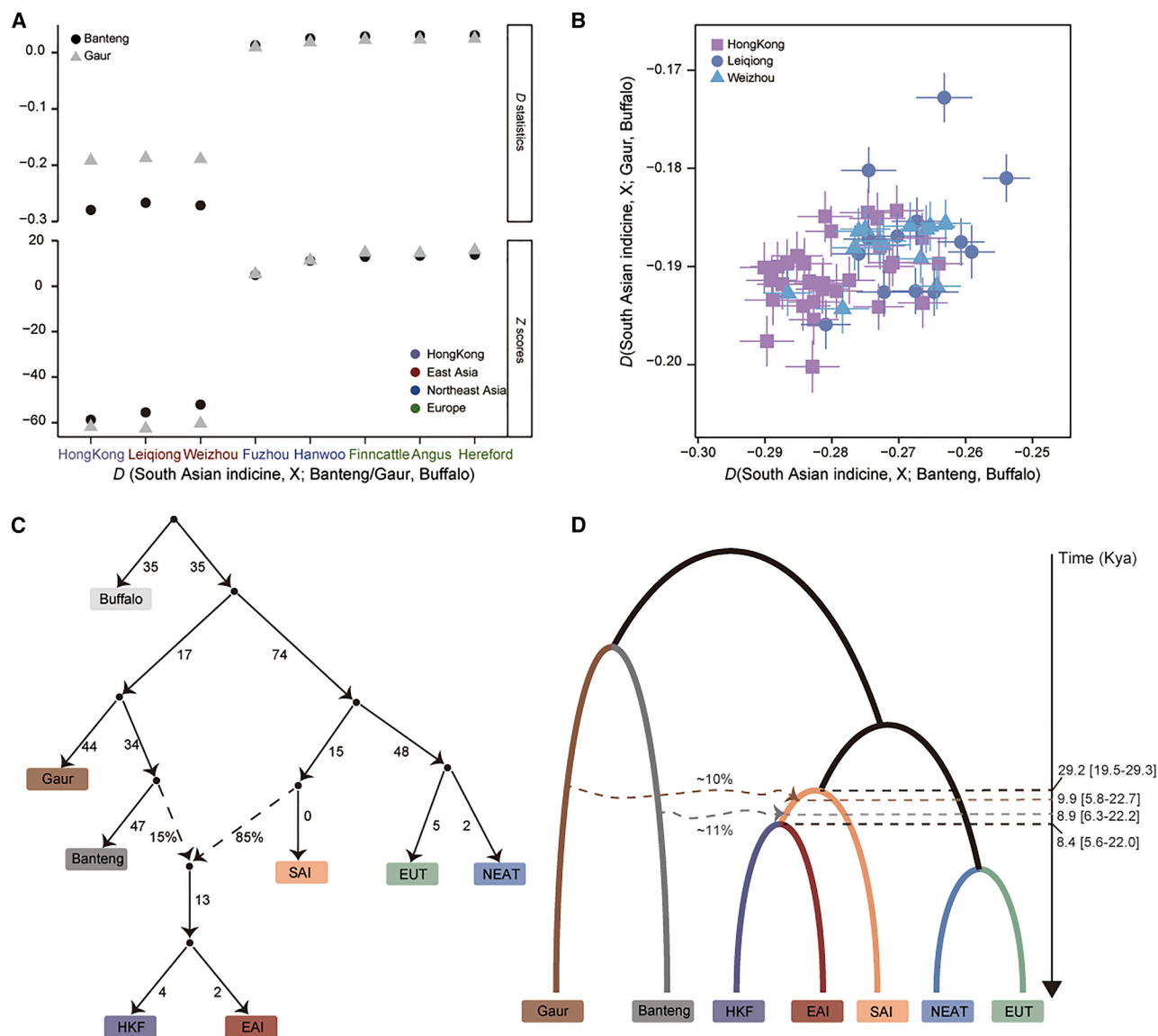


Figure 3. Introgression ratio and history of HKF cattle

(A) D statistics with the form $D(W, X; Y, Z)$ for banteng/gaur and five populations.

(B) Pairwise D statistics assessing allele sharing between HKF cattle and banteng/gaur, and comparisons of allele sharing between EAI breed cattle and banteng/gaur.

(C) Origin history with 1 admixture event inferred by qpgraph.

(D) Demographic model evaluated by fastsimcoal. The 95% confidence intervals are shown in parentheses.

RAET1G, *BOLA-DQB*, *CD1A*, *MIC1*, *IL32*, *BOLA-DQA5*, *BOLA-DQA2*, and *BOLA-NC1*, may affect the response of HKF cattle to parasitic pathogens.

To investigate the impact of candidate CNVRs on coat color of HKF cattle, we calculated the V_{ST} between individuals with black and yellow coats. A total of 143 CNVRs were found to have a V_{ST} value greater than or equal to the highest 1% of values (Figures 5C and 5D). Among these, a total of 51 genes were annotated as differentiated CNVRs. Four of these CNVRs were present in three genes that were associated with coat color (*NINL*, *DEXI*, and *CIITA*). We compared the copy numbers of

overlapping CNVRs for black and yellow HKF cattle individuals. In black individuals, CNVR-*NINL*, CNVR-*DEXI*, and CNVR-*CIITA* had higher copy numbers than in yellow individuals (Figures 5E and 5F). We hypothesize that these genes may play important roles in the coat color phenotype of HKF cattle.

DISCUSSION

HKF cattle exist primarily as feral animals in Hong Kong, China, and the population size has been declining.³² Our study used high-depth WGS data to investigate the genetic structure,

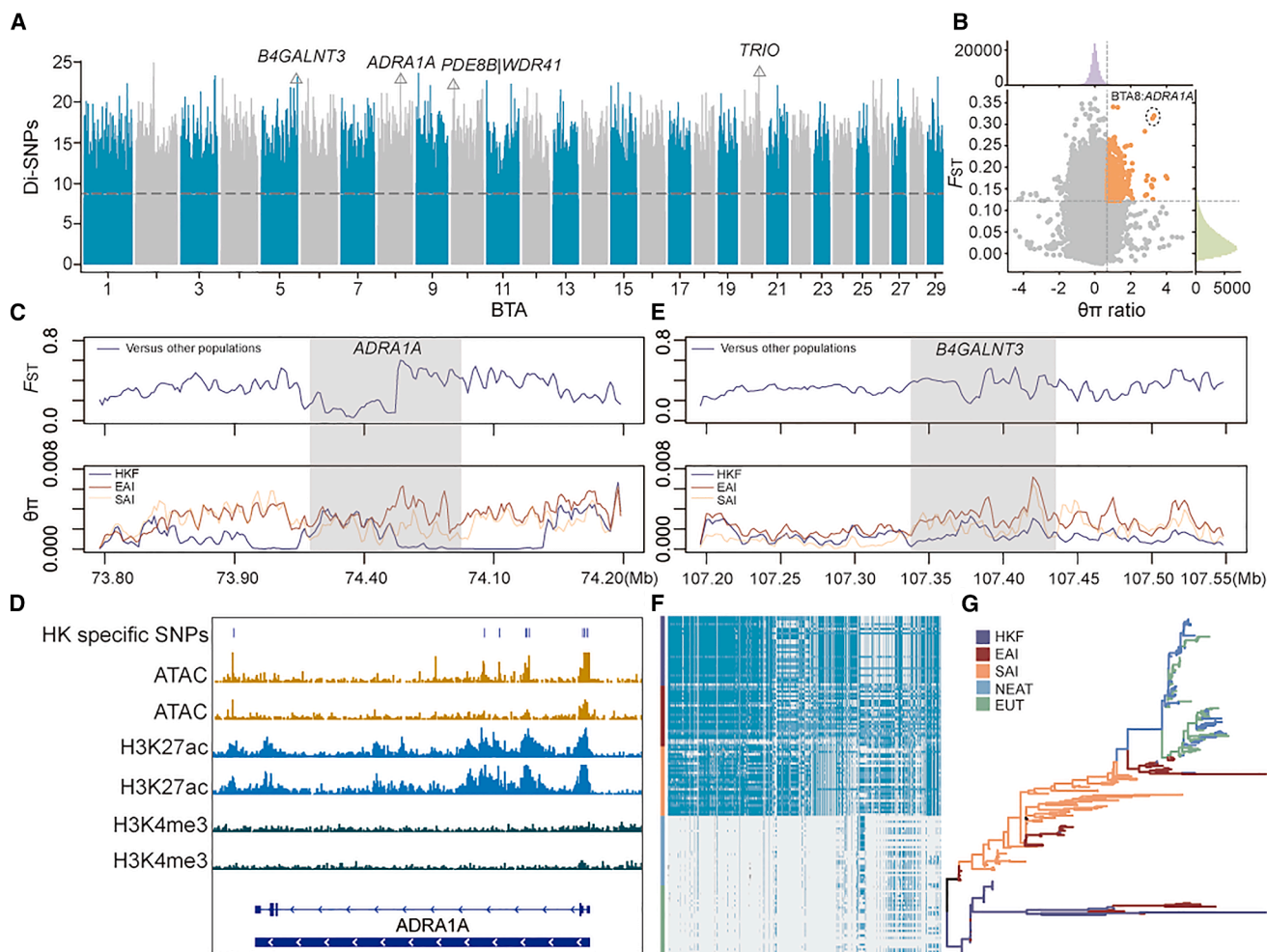


Figure 4. Candidate selection signature variants in HKF cattle

(A) Di-HKF are the 0.1% most significant Di-SNPs identified when comparing HKF versus other populations.
 (B) Distribution of the F_{ST} (x axis) and $\theta\pi$ ratio (y axis) between HKF and EAI cattle based on 50 kb windows with 20 kb steps. The loci with the highest values are highlighted by a dashed circle on BTA8.
 (C) The F_{ST} and $\theta\pi$ estimates for SNPs in the selected region on BTA8 between HKF and other indicine cattle based on 5 kb windows with 2 kb steps.
 (D) Overlap of HKF SNPs in ADRA1A with open chromatin, enhancer, and promoter markers.
 (E) The F_{ST} and $\theta\pi$ estimates for SNPs in B4GALNT3 gene on BTA5 between HKF and other indicine cattle based on 5 kb windows with 2 kb steps.
 (F) SNPs with MAF > 0.05 were used to construct haplotype patterns. The major allele at each SNP position in HKF cattle is colored magenta.
 (G) Haplotypes of the relationships among all populations on the B4GALNT3 genes. Phylogenetic tree constructed from SNPs in B4GALNT3 genes.

demographic history, introgression time frames, and genetic architecture of adaptive traits of HKF cattle. The use of WGS data is an advantage over our previous work on HKF cattle that relied on BovineSNP50 SNP chip⁸ because it reduces ascertainment bias. The BovineSNP50 chip^{33,34} is designed based on *B. taurus* and, as such, is not representative of the genetic diversity of HKF cattle that have substantial indicine genetics. Moreover, our dataset allowed for the detection of CNVs that were missed in previous studies relying primarily on genotyping arrays.⁸

Here, we showed that the ancestral contributions of HKF cattle came from EAI cattle, a pattern that was supported by the distribution of Y chromosome uniparental lineages. We also found that the HKF cattle exhibit distinct genetic patterns and internal

uniformity, likely shaped by prolonged geographic isolation. We quantified inbreeding in HKF cattle by identifying ROH. Our analysis revealed that ROH length and number in HKF cattle were significantly higher than in other indicine cattle. The high degree of inbreeding in HKF cattle may be partly explained by limited gene flow with mainland populations during Hong Kong's 151 years as a British colony (1842–1941 and 1945–1997). Interestingly, we observed fewer homozygous mutations that were predicted to be deleterious (LoF variants) in HKF cattle compared to the other cattle groups included in this study. Many predicted deleterious alleles were found in the heterozygous state. When combined with the small effective population size of HKF cattle, they pose a significant threat to future population's viability. It is possible that inbreeding in HKF cattle has

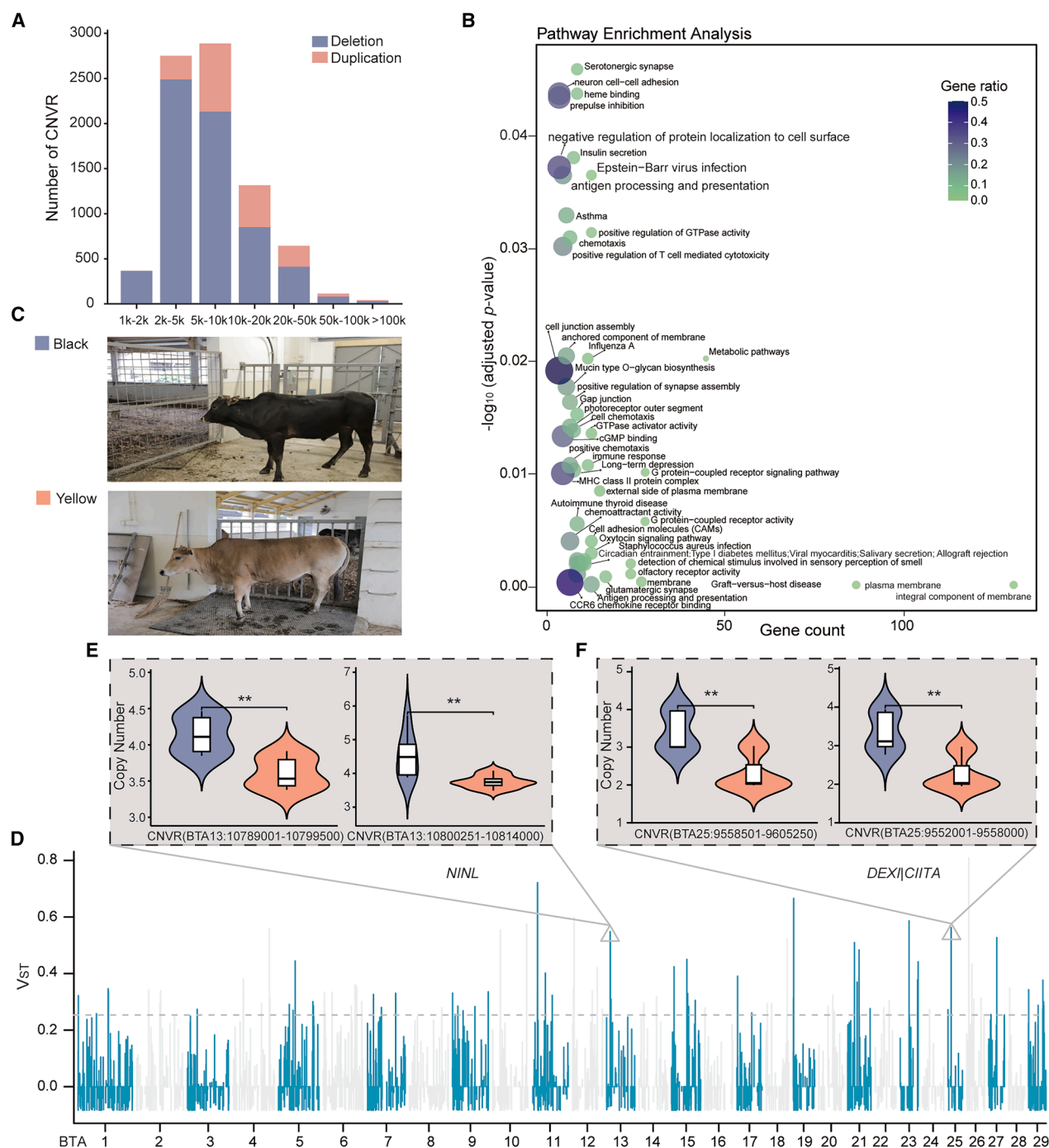


Figure 5. Detection and functional annotation of CNVRs

- (A) Counts of each CNVR type in HKF cattle, with purple and pink bars representing deletion and duplication, respectively.
- (B) Enrichment analysis based on the CNVR-related genes.
- (C) Photographs of example HKF cattle used in this study, showing the black and yellow color phenotypes (adapted from Barbato et al.⁹).
- (D) Manhattan plot of V_{ST} statistics between black and yellow color phenotypes in HKF cattle.
- (E) Estimated copy numbers of CNVRs on BTA13 (CNVR-*NINL*) between yellow and black color phenotypes.
- (F) Estimated copy numbers of CNVRs on BTA15 (CNVR-*DEXI* and CNVR-*CIITA*) between yellow and black color phenotypes.

resulted in the purging of deleterious alleles through purifying selection. Moreover, associative overdominance, whereby heterozygous loci mask deleterious effects, may also contribute to this pattern. Both mechanisms warrant further investigation to clarify their roles.

We found evidence of adaptive introgression of banteng and gaur ancestry in HKF cattle, consistent with EAI cattle.⁵ The banteng and gaur introgression occurred before the divergence between HKF and EAI cattle. The high proportion of the banteng/gaur ancestry in these populations may be due to their geographic location, which served as the primary entry point of indicine cattle into East Asia.⁵ The banteng and gaur genes introgressed into HKF were enriched in biological pathways of oxygen binding, transport and carrier activity, and bitter taste receptor activity. In particular, the bitter taste receptor activity pathway included several target genes (*TAS2R7*, *TAS2R16*, *T2R12*, *TAS2R8*, and *TAS2R38*), which are associated with local adaptation to different food sources. Preference for different food sources due to taste is considered a strong driving force in ecological speciation.^{35–37} The oxygen binding pathway has been shown to be associated with disease and parasite tolerance in EAI cattle.^{5,38} Generally, energy homeostasis and color variation can influence the adaptability and survival of species. In the HKF population, *FAM135A*, that is introgressed from banteng, was found to significantly influence coat color,²¹ potentially contributing to the adaptation of HKF cattle to tropical and subtropical environments. *FAM135A* has been found to be associated with the lipid metabolism pathway and maintenance of homeostasis.^{22,23}

Genomic adaptations to diverse environments are increasingly recognized as vital for conservation and food security, with rapid evolution shaping population productivity and viability in the face of climate change.³⁹ Our analyses identified unique and clear signals of recent selection in HKF cattle, which reflect adaptation to a free-ranging environment. One of these regions contained the gene *ADRA1A*, which is associated with heat stress recovery.⁴⁰ *ADRA1A* is a member of the G protein-coupled receptor superfamily that enhances adipocyte thermogenesis⁴¹ and changes feeding behavior due to heat stress.⁴² Variants in *ADRA1A* identified in this study revealed SNPs in predicted promoter and enhancer regions from assay for transposase-accessible chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) data. Notably, the SNP in the promoter was predicted to enhance the binding of KLF17, a transcription factor involved in immune regulation.²⁹ These non-coding variants may play a role in regulating its expression and facilitating the adaptation of HKF cattle to hot tropical ecosystems in the absence of human heat-stress mitigation management. Although the open chromatin region data came from Hereford cattle, they provide a basis for the tentative identification of conserved regulatory regions.⁴³ Further studies are needed to account for potential differences in HKF cattle for more accurate inference of regulatory regions. We also identified a selection signal for *B4GALNT3* located at BTA5 in HKF cattle, with a haplotype that was different from the other four populations analyzed. We observed a mutation in *B4GALNT3* that affects protein structure. *B4GALNT3* is associated with circulating sclerostin in humans,⁴⁴ and it may

increase bone mass, reducing fracture risk.³⁰ The positive selection signals in the genes *ADRA1A* and *B4GALNT3* may reflect the adaptation of HKF cattle to heat stress and environmental stress.

CNVs can affect gene function by altering gene dosage and structure and thus may play a role in phenotypic evolution.^{45–48} Our analysis of the number of CNVs in HKF cattle revealed that compared to the reference genome, deletions significantly exceeded the expansions. Pathway annotation of the expanded CNVRs identified in HKF cattle revealed a significant enrichment of genes related to the immune system. This suggests that HKF cattle may possess genetic features related to adaptive immune function that are potentially associated with survival and reproduction in their specific environments. The genes *BOLA-DQB*, *BOLA-DQA5*, *BOLA-DQA2*, and *BOLA-NC1* located on BTA 23, which are part of the bovine major histocompatibility complex (MHC) class II gene family, are in regions under selection and play key roles in immune response and adaptation to tropical conditions.⁴⁹ Importantly, HKF cattle have not been subjected to recent artificial selection for production and morphological traits, which has led to phenotypic diversity in populations. In this study, we observed that the *NINL* gene differed between the two coat colors of HKF cattle. *NINL* has been shown to be involved in microtubule organization, leading to the dispersion of lysosomes toward the cell periphery, thereby affecting the distribution of melanosomes.^{50,51} The expression of *NINL* is higher in Asian skin, which may regulate the distribution of melanosomes, contributing to the distinct skin color differences compared to Caucasians.⁵² We hypothesize that *NINL* may have a role in the coat color phenotypic variations seen in HKF cattle. *DEXI* and *CIITA* have been shown to be associated with black color in camels.⁵³

In this study, we explored the unique interactions between the demography and genetics of HKF cattle to better understand the evolutionary history of this population. We provided evidence that the ancestral contributions of HKF cattle came from EAI cattle. We identified wild introgression into HKF cattle, likely a result of their geographic position as a gateway for indicine migration into East Asia. At the genomic level, we were able to find several genetic variations that may be adaptations to the environment, potentially related to heat tolerance, survival, and disease resistance. Practically, HKF cattle have developed into a genetically distinct population clearly differentiated from their domestic ancestors. These results indicate how animals free from human management have to reshape their genome to cope with a harsher environment as they are unprotected from biotic and abiotic stress. This adaptation is at the expense of diversity caused by purifying selection and increased inbreeding due to mating of relatives from families having higher fitness under natural conditions.

Limitations of the study

This study provides insights into the genetic origin and unique adaptation of HKF cattle. One limitation lies in the lack of breed-specific open chromatin data, which restricts the direct assessment of selection on non-coding regulatory regions. Without such data, inferences about regulatory adaptation are indirect and may not fully reflect the breed's unique regulatory

landscape. Generating open chromatin datasets in relevant tissues would allow a more precise characterization of regulatory selection. Another limitation is that the functional effects of candidate loci, particularly those associated with coat color, remain untested. Targeted *in vivo* and *in vitro* studies will be necessary to validate these loci and test the hypotheses proposed here. Additionally, potential sex-specific effects cannot be excluded, and future studies should incorporate sex as an important variable.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Wai Yee Low (wai.low@adelaide.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The newly generated whole-genome sequence data from this study have been deposited in the National Genomics Data Center (NGDC), China National Center for Bioinformation, under the Bioproject accession number PRJCA040422. The details of aforementioned data and other downloaded publicly available data used in this study are provided in [Table S1](#).
- All analysis code has been deposited in the GitHub repository and is publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

This work was supported by National Key R&D Programs of China (grant no. 2021YFF1001000), the National Natural Science Foundation of China (32341054 and 32522096), and the China Agriculture Research System of MOF and MARA (CARS-37). W.Y.L., J.L.W., and R.T. were supported by the Davies Bequest through the Davies Livestock Research Center of the University of Adelaide. We thank the High-Performance Computing (HPC) of Northwest A&F University (NWAUFU) for providing the computing resources. We gratefully acknowledge all participants who contributed to this study. The cow illustration used in the graphical abstract was provided by [BioGDP.com](https://biogdp.com/) (<https://biogdp.com/>).⁵⁴

AUTHOR CONTRIBUTIONS

W.Y.L. and N.C. designed and supervised the project. X. Luo and X. Lu performed the majority of analysis. Y.R., X. Liu, and Y.C. were involved in data visualization. X. Luo, N.C., Y.R., W.Y.L., and C.L. wrote, reviewed, and edited the manuscript with input from all authors. M.P.R. and R.A.L.B. collected the samples and revised the manuscript. J.L.W. initiated the study, coordinated sample collection, and revised the manuscript. P.A.-M., M.B., and R.T. contributed ideas for analysis and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
- [METHOD DETAILS](#)
 - DNA re-sequencing data
 - Variant calling
 - Comparison of findings between SNP array and WGS
 - Population structure and genetic differentiation analysis
 - Runs of homozygosity
 - Screening of loss-of-function mutations
 - Estimation of effective population size
 - Selective sweep analysis
 - ATAC-seq, ChIP-seq and transcription factor footprinting analyses
 - Introgression analysis
 - Copy number variations calling and detection
 - Annotation and functional enrichment analysis
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.115465>.

Received: September 19, 2025

Revised: October 26, 2025

Accepted: March 20, 2026

Published: March 24, 2026

REFERENCES

1. Felius, M., Beerling, M.-L., Buchanan, D., Theunissen, B., Koolmees, P., and Lenstra, J. (2014). On the History of Cattle Genetic Resources. *Diversity* 6, 705–750.
2. Pitt, D., Bruford, M.W., Barbato, M., Orozco-terWengel, P., Martínez, R., and Sevane, N. (2019). Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evol. Appl.* 12, 105–122. <https://doi.org/10.1111/eva.12641>.
3. AFCD (2022). Stray Cattle Management Plan. Agriculture, Fisheries and Conservation Department. <https://www.afcd.gov.hk/english/quarantine/cattlebuffalo.html>.
4. Wang, C. (2024). From Isolated Fishing Village to Cosmopolitan City: Geographical Changes of Hong Kong in Modern Times. *Journal of Education, Humanities and Social Sciences* 35, 131–136. <https://doi.org/10.54097/sm9q6r74>.
5. Chen, N., Xia, X., Hanif, Q., Zhang, F., Dang, R., Huang, B., Lyu, Y., Luo, X., Zhang, H., Yan, H., et al. (2023). Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. *Nat. Commun.* 14, 7803.
6. Wu, D.D., Ding, X.D., Wang, S., Wójcik, J.M., Zhang, Y., Tokarska, M., Li, Y., Wang, M.S., Faruque, O., Nielsen, R., et al. (2018). Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* 2, 1139–1145.
7. Moran, B.M., Payne, C., Langdon, Q., Powell, D.L., Brandvain, Y., and Schumer, M. (2021). The genomic consequences of hybridization. *eLife* 10, e69016.
8. Barbato, M., Reichel, M.P., Passamonti, M., Low, W.Y., Colli, L., Tearle, R., Williams, J.L., and Ajmone Marsan, P. (2020). A genetically unique Chinese cattle population shows evidence of common ancestry with wild species when analysed with a reduced ascertainment bias SNP panel. *PLoS One* 15, e0231162.
9. Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* 91, 275–292.
10. Razali, R.M., Rodriguez-Flores, J., Ghorbani, M., Naeem, H., Aamer, W., Aliyev, E., Jubran, A., Qatar Genome Program Research Consortium; Clark, A.G., Fakhro, K.A., and Mokrab, Y. (2021). Thousands of Qatari

genomes inform human migration history and improve imputation of Arab haplotypes. *Nat. Commun.* **12**, 5929.

11. Gulisija, D., Gonzalez-Reymundez, A., Fenton, J.I., de Los Campos, G., Bray, M.S., and Vazquez, A.I. (2025). Uncovering covariance patterns across energy balance traits enables the discovery of new obesity-related genes. *Obesity* **33**, 1184–1194. <https://doi.org/10.1002/oby.24291>.
12. Yang, P., Zhang, Z., Xu, J., Qu, K., Lyv, S., Wang, X., Cai, C., Li, Z., Wang, E., Xie, J., et al. (2020). The Association of the Copy Number Variation of the MLLT10 Gene with Growth Traits of Chinese Cattle. *Animals* **10**, 250. <https://doi.org/10.3390/ani10020250>.
13. Chevalier, M., Rhee, H., Elguindi, E.C., and Blond, S.Y. (2000). Interaction of murine BIP/GRP78 with the DnaJ homologue MTJ1. *J. Biol. Chem.* **275**, 19620–19627. <https://doi.org/10.1074/jbc.M001333200>.
14. Sahi, C., Kominek, J., Ziegelhoffer, T., Yu, H.Y., Baranowski, M., Marszałek, J., and Craig, E.A. (2013). Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. *Mol. Biol. Evol.* **30**, 985–998. <https://doi.org/10.1093/molbev/mst008>.
15. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309.
16. Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166.
17. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* **192**, 1065–1093.
18. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288.
19. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
20. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905.
21. Baltazar-Soares, M., Karell, P., Wright, D., Nilsson, J.Å., and Brommer, J.E. (2024). Genomic basis of melanin-associated phenotypes suggests colour-specific environmental adaptations in tawny owls. *Mol. Ecol.* **33**, e17247.
22. Poleti, M.D., Regitano, L.C.A., Souza, G.H.M.F., Cesar, A.S.M., Simas, R.C., Silva-Vignato, B., Oliveira, G.B., Andrade, S.C.S., Cameron, L.C., and Coutinho, L.L. (2018). Longissimus dorsi muscle label-free quantitative proteomic reveals biological mechanisms associated with intramuscular fat deposition. *J. Proteomics* **179**, 30–41.
23. Zhang, W., Wu, C., Ni, R., Yang, Q., Luo, L., and He, J. (2021). Formimidoyltransferase cyclodeaminase prevents the starvation-induced liver hepatomegaly and dysfunction through downregulating mTORC1. *PLoS Genet.* **17**, e1009980.
24. Akey, J.M., Ruhe, A.L., Akey, D.T., Wong, A.K., Connelly, C.F., Madeoy, J., Nicholas, T.J., and Neff, M.W. (2010). Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. USA* **107**, 1160–1165.
25. Barbosa, S., Greville-Heygate, S., Bonnet, M., Godwin, A., Fagotto-Kaufmann, C., Kajava, A.V., Laouteouet, D., Mawby, R., Wai, H.A., Dingemans, A.J.M., et al. (2020). Opposite Modulation of RAC1 by Mutations in TRIO Is Associated with Distinct, Domain-Specific Neurodevelopmental Disorders. *Am. J. Hum. Genet.* **106**, 338–355.
26. Taylor, P.N., Porcu, E., Chew, S., Campbell, P.J., Traglia, M., Brown, S.J., Mullin, B.H., Shihab, H.A., Min, J., Walter, K., et al. (2015). Whole-genome sequence-based analysis of thyroid function. *Nat. Commun.* **6**, 5681.
27. McAlpine, W., Sun, L., Wang, K.W., Liu, A., Jain, R., San Miguel, M., Wang, J., Zhang, Z., Hayse, B., McAlpine, S.G., et al. (2018). Excessive endosomal TLR signaling causes inflammatory disease in mice with defective SMCR8-WDR41-C9ORF72 complex function. *Proc. Natl. Acad. Sci. USA* **115**, E11523–E11531.
28. de Freitas, A.C., Reolon, H.G., Abduch, N.G., Baldi, F., Silva, R.M.O., Lourenco, D., Fragomeni, B.O., Paz, C.C.P., and Stafuzza, N.B. (2024). Proteomic identification of potential biomarkers for heat tolerance in Caracu beef cattle using high and low thermotolerant groups. *BMC Genom.* **25**, 1079.
29. Pérez-Stuardo, D., Frazão, M., Ibaceta, V., Brianson, B., Sánchez, E., Rivas-Pardo, J.A., Vallejos-Vidal, E., Reyes-López, F.E., Toro-Ascuy, D., Vidal, E.A., and Reyes-Cerpa, S. (2023). KLF17 is an important regulatory component of the transcriptomic response of Atlantic salmon macrophages to *Piscirickettsia salmonis* infection. *Front. Immunol.* **14**, 1264599. <https://doi.org/10.3389/fimmu.2023.1264599>.
30. Movérare-Skrtic, S., Voelkl, J., Nilsson, K.H., Nethander, M., Luong, T.T.D., Alesutan, I., Li, L., Wu, J., Horkeby, K., Lagerquist, M.K., et al. (2023). B4GALNT3 regulates glycosylation of sclerostin and bone mass. *EBioMedicine* **91**, 104546.
31. Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., Bruhn, L., Shendure, J., and 1000 Genomes Project; and Eichler, E.E. (2010). Diversity of human copy number variation and multicopy genes. *Science* **330**, 641–646.
32. Massei, G., Koon, K.K., Benton, S., Brown, R., Gomm, M., Orahood, D.S., Pietravalle, S., and Eckery, D.C. (2015). Immunoneutralization for managing feral cattle in Hong Kong. *PLoS One* **10**, e0121598.
33. Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilki, J., Seabury, C.M., Caetano, A.R., et al. (2009). Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl. Acad. Sci. USA* **106**, 18644–18649. <https://doi.org/10.1073/pnas.0904691106>.
34. Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., and Van Tassell, C.P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* **4**, e5350. <https://doi.org/10.1371/journal.pone.0005350>.
35. Berner, D., and Thibert-Plante, X. (2015). How mechanisms of habitat preference evolve and promote divergence with gene flow. *J. Evol. Biol.* **28**, 1641–1655.
36. Kopp, M., Servedio, M.R., Mendelson, T.C., Safran, R.J., Rodríguez, R.L., Hauber, M.E., Scordato, E.C., Symes, L.B., Balakrishnan, C.N., Zonana, D.M., and van Doorn, G.S. (2018). Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. *Am. Nat.* **191**, 1–20.
37. Barbato, M., Hailer, F., Orozco-terWengel, P., Kijas, J., Mereu, P., Cabras, P., Mazza, R., Pirastru, M., and Bruford, M.W. (2017). The genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci. Rep.* **7**, 7623. <https://doi.org/10.1038/s41598-017-07382-7>.
38. Brown, C.G. (1997). Dynamics and impact of tick-borne diseases of cattle. *Trop. Anim. Health Prod.* **29**, 1S–3S.
39. Stockwell, C.A., Hendry, A.P., and Kinnison, M.T. (2003). Contemporary evolution meets conservation biology. *Trends Ecol. Evol.* **18**, 94–101. [https://doi.org/10.1016/S0169-5347\(02\)00044-7](https://doi.org/10.1016/S0169-5347(02)00044-7).
40. Reolon, H.G., Abduch, N.G., de Freitas, A.C., Silva, R.M.d.O., Fragomeni, B.d.O., Lourenco, D., Baldi, F., de Paz, C.C.P., and Stafuzza, N.B. (2024). Proteomic changes of the bovine blood plasma in response to heat stress in a tropically adapted cattle breed. *Front. Genet.* **15**, 1392670.
41. Rahbani, J.F., Scholtes, C., Lagarde, D.M., Hussain, M.F., Roesler, A., Dykstra, C.B., Bunk, J., Samborska, B., O'Brien, S.L., Tripp, E., et al. (2022). ADRA1A-Galpa(q) signalling potentiates adipocyte thermogenesis through CKB and TNAP. *Nat. Metab.* **4**, 1459–1473.
42. Cross, A.J., Keel, B.N., Brown-Brandl, T.M., Cassady, J.P., and Rohrer, G.A. (2018). Genome-wide association of changes in swine feeding behaviour due to heat stress. *Genet. Sel. Evol.* **50**, 11.

43. Kern, C., Wang, Y., Xu, X., Pan, Z., Halstead, M., Chanthavixay, G., Saelao, P., Waters, S., Xiang, R., Chamberlain, A., et al. (2021). Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat. Commun.* 12, 1821.
44. Zheng, J., Maerz, W., Gergei, I., Kleber, M., Drechsler, C., Wanner, C., Brandenburg, V., Reppe, S., Gautvik, K.M., Medina-Gomez, C., et al. (2019). Mendelian Randomization Analysis Reveals a Causal Influence of Circulating Sclerostin Levels on Bone Mineral Density and Fractures. *J. Bone Miner. Res.* 34, 1824–1836.
45. Harris, R.M., and Hofmann, H.A. (2015). Seeing is believing: Dynamic evolution of gene families. *Proc. Natl. Acad. Sci. USA* 112, 1252–1253.
46. Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22, 778–790.
47. Choi, J.W., Lee, K.T., Liao, X., Stothard, P., An, H.S., Ahn, S., Lee, S., Lee, S.Y., Moore, S.S., and Kim, T.H. (2013). Genome-wide copy number variation in Hanwoo, Black Angus, and Holstein cattle. *Mamm. Genome* 24, 151–163.
48. Low, W.Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D.M., Rosen, B.D., Kronenberg, Z.N., Kingan, S.B., Tseng, E., et al. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat. Commun.* 11, 2071. <https://doi.org/10.1038/s41467-020-15848-y>.
49. Maiorano, A.M., Cardoso, D.F., Carvalheiro, R., Júnior, G.A.F., de Albuquerque, L.G., and de Oliveira, H.N. (2022). Signatures of selection in Nelore cattle revealed by whole-genome sequencing data. *Genomics* 114, 110304.
50. Casenghi, M., Meraldi, P., Weinhart, U., Duncan, P.I., Körner, R., and Nigg, E.A. (2003). Polo-like kinase 1 regulates Nlp, a centrosome protein involved in microtubule nucleation. *Dev. Cell* 5, 113–125.
51. Wu, X., Sakamoto, T., Zhang, F., Sellers, J.R., and Hammer, J.A., 3rd. (2006). In vitro reconstitution of a transport complex containing Rab27a, melanophilin and myosin Va. *FEBS Lett.* 580, 5863–5868.
52. Yin, L., Coelho, S.G., Ebsen, D., Smuda, C., Mahns, A., Miller, S.A., Beer, J.Z., Kolbe, L., and Hearing, V.J. (2014). Epidermal gene expression and ethnic pigmentation variations among individuals of Asian, European and African ancestry. *Exp. Dermatol.* 23, 731–735.
53. Bitaraf, S.M., Zare, H.J., Banabazi, M.H., Faraz, A., Esmailkhanian, S., Naderi, A.S., Salim, N., Teimoori, A., Bitaraf, A., Zadehrahmani, M., et al. (2022). Identification of Candidate Genes for Pigmentation in Camels Using Genotyping-by-Sequencing. *Animals* 12, 1095.
54. Jiang, S., Li, H., Zhang, L., Mu, W., Zhang, Y., Chen, T., Wu, J., Tang, H., Zheng, S., Liu, Y., et al. (2025). Generic Diagramming Platform (GDP): a comprehensive database of high-quality biomedical graphics. *Nucleic Acids Res.* 53, D1670–D1676. <https://doi.org/10.1093/nar/gkae973>.
55. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
56. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
57. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
58. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.
59. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
60. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
61. Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinf.* 12, 246.
62. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 38, 3022–3027.
63. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
64. Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *Rom. Jahrb.* 8, 289–317.
65. Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A., and Li, J.Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* 93, 90–102. <https://doi.org/10.1016/j.ajhg.2013.05.003>.
66. Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* 47, 11.12.1–11.12.34.
67. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
68. Feng, Y., Xie, N., Inoue, F., Fan, S., Saskin, J., Zhang, C., Zhang, F., Hansen, M.E.B., Nyambo, T., Mpoloka, S.W., et al. (2024). Integrative functional genomic analyses identify genetic variants influencing skin pigmentation in Africans. *Nat. Genet.* 56, 258–272. <https://doi.org/10.1038/s41588-023-01626-1>.
69. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
70. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
71. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
72. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 52, D174–D182.
73. Hou, J., Guan, X., Xia, X., Lyu, Y., Liu, X., Mazei, Y., Xie, P., Chang, F., Zhang, X., Chen, J., et al. (2024). Evolution and legacy of East Asian aurochs. *Sci. Bull.* 69, 3425–3433.
74. Chen, S., Ren, L., Gao, Y., Dong, G., Sheng, G., Han, J., Liu, X., Chen, N., and Chen, F. (2024). Evidence of hybridization of cattle and aurochs on the Tibetan Plateau ~3750 years ago. *Sci. Bull.* 69, 2825–2828.
75. Yan, H., Chen, J., Lei, C., Zhang, F., He, S., Song, W., Ma, H., Li, S., Li, G., Yongcuo, L., et al. (2026). Genetic Diversity, Adaptation, Wild Introgression, and Coat Color Mutation of Golden Yak. *Animal Research and One Health*, 1–13. <https://doi.org/10.1002/aro2.70049>.
76. Liu, Y.-K., Fu, W.-W., Wang, Z.-Y., Pei, S.-W., Li, K.-H., Wu, W.-W., Le, M.-Z., and Yue, X.-P. (2025). Genomic insights into the genetic diversity, lateral gaits and high-altitude adaptation of Chakouyi (CKY) horses. *J. Genet. Genomics* 52, 1001–1010.
77. Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., et al. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325.
78. Zhao, Q., Xu, J., Shi, Z., Zhang, Y., Du, X., Zhai, Y., Xu, J., Liu, F., and Zhang, Q. (2025). Genome-wide Pleiotropy Analysis Reveals Shared Genetic Associations between Type 2 Diabetes Mellitus and Subcortical Brain Volumes. *Research* 8, 0688.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Hong Kong feral cattle (HKF)	Hong Kong Special Administrative Region of China	N/A
Deposited data		
Resequencing data of Hong Kong feral cattle (HKF)	This study; National Genomics Data Center (NGDC)	PRJCA040422
Resequencing data of East Asian indicine cattle (EAI)	National Center for Biotechnology Information search database (NCBI)	PRJNA658727; PRJNA379859; PRJNA283480
Resequencing data of South Asian indicine cattle (SAI)	National Center for Biotechnology Information search database (NCBI)	PRJNA658727; PRJNA379859
Resequencing data of European taurine cattle (EUT)	National Center for Biotechnology Information search database (NCBI); European Bioinformatics Institute (EBI)	PRJNA176557; PRJEB28185
Resequencing data of Northeast Asian taurine cattle (NEAT)	National Center for Biotechnology Information search database (NCBI)	PRJNA210519; PRJNA1085861
Resequencing data of Buffalo cattle	National Center for Biotechnology Information search database (NCBI)	PRJNA547460
Resequencing data of Banteng and Gaur	National Center for Biotechnology Information search database (NCBI); European Bioinformatics Institute (EBI)	PRJNA325061; PRJNA658727; PRJEB31621; PRJNA427536
Software and algorithms		
Burrow-Wheeler Aligner	https://github.com/lh3/bwa	v.0.7.13
SAMtools	https://github.com/samtools/	v.1.9
Genome Analysis Toolkit (GATK)	https://github.com/broadinstitute/gatk/releases	v.3.8.1.0
ANNOVAR	https://github.com/WGLab/doc-ANNOVAR	v2016-02-01
PLINK	https://www.cog-genomics.org/plink/	v.1.9
EIGENSOFT	https://github.com/argriffing/eigensoft	v4.2
ADMIXTURE	https://dalexander.github.io/admixture/	v1.3.0
MEGA	https://www.megasoftware.net/	v11.0
VCFtools	https://vcftools.github.io/index.html	v0.1.16
Mclust R package	https://mclust-org.github.io/mclust/	v6.1.1
BEDTools	https://github.com/arq5x/bedtools2	v.2.25.0
SnEff	http://pcingola.github.io/SnpEff/	v4.3
MSMC2	https://github.com/stschiff/msmc2	V2.1.0
SMC++	https://github.com/popgenmethods/smcpp	v1.15.4
RepeatMasker	http://www.repeatmasker.org	v.4.0.5
Trim Galore	https://github.com/FelixKrueger/TrimGalore	v.0.6.10
Bowtie2	https://bowtie-bio.sourceforge.net/bowtie2	v.2.4.5
Picard	http://broadinstitute.github.io/picard	v.2.20.2
Mac2	https://github.com/mac3-project/MACS	v.2.2.7.1
MEME	https://github.com/cinquin/MEME	v5.5.2
RFMix	https://github.com/slowkoni/rfmix	v2.02

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ADMIXTOOLS	https://github.com/DReichLab/AdmixTools	v7.0.2
TreeMix	https://github.com/carolindahms/TreeMix	v1.13
Fastsimcoal	https://speciationgenomics.github.io/fastsimcoal2/	v2.6
CNVnator	https://github.com/abyzovlab/CNVnator	v.0.4.1
Kobas	http://bioinfo.org/kobas	v3.0
Genome analysis pipeline	https://github.com/xxiaoyull/cattle-genome-analysis	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Whole-blood samples were collected from 30 cattle in Hong Kong, China. The animals sampled are part of a feral cattle population of approximately 1,000 individuals, with samples mainly collected from Sai Kung, Lantau Island, and the East and North New Territories of Hong Kong.⁸ Both male and female animals were included, and sex information was recorded for each individual (Table S1). All animal work was done in compliance with Animals (Control of Experiments) Ordinance Chapter 340, the Laws of Hong Kong (License Nos 16–47 to 50, DH/HA&P/8/2/5 Pt 5), and with City University of Hong Kong Ethical Committee approval. Animals were maintained under standard local husbandry conditions prior to sampling.

METHOD DETAILS**DNA re-sequencing data**

Genomic DNA was extracted from whole-blood samples of 30 feral cattle (see Experimental Model and Study Participant Details). DNA libraries were prepared and sequenced according to the standard protocols of BGI using a BGI-SEQ500 sequencer. Paired-end sequencing (2x150bp) was done with the target coverage of 30X. We checked the sequencing quality with FastQC v0.11.3 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). In addition, we collected 124 WGS data from the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra/>), for 30 European taurine (Hereford, Angus, and Finncattle), 30 Northeast Asian taurine (Hanwoo and Fuzhou), 26 East Asian indicine (Weizhou and Leiqong) and 30 South Asian indicine (Sahiwal, Thawalam and Dhanni), four banteng, two of gaur, and two swamp buffalo (Table S1).

Variant calling

All reads were mapped to the *Bos taurus* reference genome ARS-UCD1.2 using BWA-MEM v0.7.13-r1126.⁵⁵ SAMtools v1.9⁵⁶ and Picard tools v2.20.2 (<http://broadinstitute.github.io/picard>) were used to sort the BAM file and filter duplicate reads. All genomic data were uniformly processed following the approach of a previous study.⁵ Variants were identified using ‘HaplotypeCaller’, ‘GenotypeGVCFs’, ‘SelectVariants’ and ‘VariantFiltration’ in the Genome Analysis Toolkit (GATK v3.8-1-0-gf15c1c3ef).⁵⁷ Only biallelic alleles were retained for downstream analyses. Additionally, we performed functional annotation on all autosomes using ANNOVAR v2016-02-01.⁵⁸

Comparison of findings between SNP array and WGS

Previously, the genetic composition of 21 HKF cattle was investigated with Illumina BovineSNP50v2 BeadChip.⁸ We downloaded the SNP data of these 21 animals to compare allele frequency spectrum and overlapping SNPs with our WGS data. There were only two animals in common between the two datasets. The genomic coordinates of the SNP array markers were determined according to the ARS-UCD1.2 genome assembly (liftover coordinates available at https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/). Minor allele frequency, heterozygosity and inbreeding coefficients were computed using PLINK v1.9.⁵⁹

Population structure and genetic differentiation analysis

For the population structure analysis, we filtered the SNPs with MAF < 0.01 and high linkage disequilibrium (LD) using PLINK v1.9.⁵⁹ The LD filtering parameter was ‘-indep-pairwise 50 10 0.1’, corresponding to an r^2 threshold of 0.1. This gave us 3,504,555 SNPs for principal component analysis (PCA) using EIGENSOFT v4.2.⁶⁰ Using the same dataset, genetic admixture analysis was performed with ADMIXTURE v1.3.0.⁶¹ For the phylogenetic analysis, we selected SNPs with MAF < 0.01, resulting in 46,167,659 SNPs for tree construction. We calculated a matrix of pairwise genetic distances and constructed the neighbor-joining (NJ) phylogenetic tree with MEGA v.11.0.⁶²

For 84 male samples, we obtained Y chromosomes with an average sequencing depth of 8.39x. All clean reads were aligned to the Btau_5.0.1 Y chromosome reference sequence (GCF_000003205.7), and SNPs with more than 10% missing genotypes were filtered out. A total of 2823 SNPs passed the quality control. Phylogenetic trees were then inferred using ML methods performed with MEGA v.11.0.⁶²

F_{ST} were calculated using the ‘smartpca’ command from EIGENSOFT v4.2⁶⁰ with parameters ‘fstonly’. The nucleotide diversity of each breed was assessed by VCFtools v0.1.16⁶³ with parameters: ‘-window-pi 50000 -window-pi-step 20000’.

Runs of homozygosity

ROH were calculated using PLINK v1.9 based on phased and imputed SNPs. To characterize the distribution of ROH lengths, we applied a Gaussian mixture model implemented in the ‘mclust’ R package per population.⁶⁴ The ROH segments were classified into short, medium and long size classes following the recommended approach.^{9,65} The boundaries of ROH were calculated to compare distribution patterns across different populations.

To evaluate specific ROH in HKF cattle, we first defined common ROH within a population as those present in more than 50% of individuals. We then used BEDTools v.2.25.0⁶⁶ to compare ROH across multiple populations, retaining regions present in HKF but absent in other populations as specific ROH to HKF cattle.

Screening of loss-of-function mutations

We estimated the LoF mutations in all population genomes using SnpEff v4.3.⁶⁷ We used the swamp buffalo genomes to infer ancestral alleles in the HKF genome to screen derived alleles. Here, we considered ‘stop_gained’, ‘stop_lost’, ‘start_lost’, ‘splice_donor_variant’, and ‘splice_acceptor_variant’ as LoF mutations. We counted the number of LoF mutations per individual in the homozygous and heterozygous states and calculated the proportion of homozygous-derived SNPs with the following formula: $2 \times \text{homozygous sites} / (2 \times \text{homozygous sites} + \text{heterozygous sites})$.

Estimation of effective population size

To infer effective population sizes (N_e), we applied both MSMC2¹⁶ and SMC++¹⁵ analyses. MSMC2 was applied to all populations using two deep-coverage individuals per group, while SMC++ was computed on 15 groups of randomly selected 10 samples in the HKF population. In accordance with the recommendation, we masked low-complexity regions of the genome identified by RepeatMasker v.4.0.5 (<http://www.repeatmasker.org>), and focused on biallelic SNPs. For demographic inference, we used a generation time of six years and a per-generation mutation rate of 1.26×10^{-8} .

Selective sweep analysis

We calculated the D_i statistic^{24,68} to identify SNPs with highly differentiated allele frequencies between HKF and other cattle populations. For each SNP, D_i measures the standardized deviation of pairwise F_{ST} values from the mean across populations, with higher values indicating greater population differentiation. We then selected the top 0.1% of D_i values for further downstream functional analysis. We also used a F_{ST} and θ_{π} ratio to detect selection sweeps within 50 kb windows and 20 kb steps in HKF cattle. The genomic selection regions considered as important were those identified using both methods.

ATAC-seq, ChIP-seq and transcription factor footprinting analyses

A total of ATAC-seq and ChIP-seq data from seven tissues of two animals ($n=2$) were obtained from NCBI. The raw reads were filtered using Trim Galore v.0.6.10 (<https://github.com/FelixKrueger/TrimGalore>) and mapped to the ARS-UCD1.2 reference genome using Bowtie2 v.2.4.5.⁶⁹ Duplicates and low-quality alignments were removed with Picard v.2.20.2 (<http://broadinstitute.github.io/picard>). The ATAC-seq peaks were called using macs2 v.2.2.7.1.⁷⁰ The two duplicates of the same tissue were merged using BEDTools v.2.25.0.⁶⁶ The same pipeline was used for processing the ChIP-seq data. Transcription factor binding sites were predicted using FIMO from the MEME suite,^{68,71} with motifs obtained from the JASPAR CORE 2024 database.⁷²

Introgession analysis

We used the D statistics¹⁷ and RFMix v2.02¹⁸ methods to characterize potential banteng or gaur introgression into HKF cattle. To avoid bias in the introgression ratio, we selected five pure South Asian indicine cattle (SAI), retrieved from a previous publication,⁵ for subsequent D statistics calculations using ADMIXTOOLS v7.0.2.¹⁷ RFMix v2.02¹⁸ was used to determine introgressed ratios based on phased data. In addition, we inferred a population-level phylogeny and migration events using qpgraph in admixtools2 (<https://uqrmaie1.github.io/admixtools/articles/admixtools.html>) and TreeMix.¹⁹ We used the ‘find_graphs’ function to find the best-fitting admixture graph, with the swamp buffalo set as the outgroup.^{73,74} We further explored demographic models and estimated the timing of introgression based on the qpgraph results using fastsimcoal v2.6.²⁰ For the demographic model, we performed 100 optimization runs with ‘-n100000’ and ‘-L20’ to obtain the best-fitting parameters yielding the highest likelihood.

To detect regions of adaptive introgression, we used U_{50SAI} , HKF cattle, banteng/gaur (1%, 50%, and 100%) based on 50 kb windows with 20 kb steps. For each window, we counted the number of SNPs where a specific allele was fixed (100% frequency) in banteng/gaur, its frequency in SAI cattle was less than 1% and greater than 50% in HKF cattle. Windows in the top 5% of such allele counts that met these criteria were considered as candidate introgressed regions.

Copy number variations calling and detection

The CNVnator v.0.4.1 software was used to call CNVs relative to the ARS-UCD1.2 reference assembly. The bin size was set to 250bp for each individual. All detected events had a length >1 kb, a p-value based on t-test statistics <0.001 and a fraction of reads with zero mapping quality (q0) < 0.5. CNVRs were defined using TCAG-WGS-CNV-workflow (<https://github.com/bjtrost/TCAG-WGS-CNV-workflow>). We considered CNVs with at least 50% reciprocal overlap as CNVRs.

To detect genetic variants involved in the coat color phenotype of HKF cattle, we calculated V_{ST} by comparing black and yellow HKF cattle. CNVRs with V_{ST} values in the top 1% of the empirical distribution were identified as putative selective sweeps associated with coat color.^{75,76}

Annotation and functional enrichment analysis

We obtained candidate genes by annotating candidate selective sweep regions, introgressed regions and CNVRs. These candidate genes were analyzed using kobas 3.0⁷⁷ to explain their biological functions (corrected P value < 0.05). The statistical significance of enriched terms was assessed using the hypergeometric test or Fisher's exact test, and the Benjamini and Hochberg method was applied to control the false discovery rate (FDR).⁷⁸

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details are provided in the corresponding figure legends. For comparisons of cumulative ROH length among populations, pairwise statistical analyses were performed using the Wilcoxon rank-sum test. Statistical significance is indicated by asterisks: * P < 0.05, ** P < 0.01, *** P < 0.001, and **** P < 0.0001. Demographic models were inferred using fastsimcoal, and parameter uncertainty was assessed using 95% confidence intervals.