






















The genetic diversity of Indonesian cattle has been shaped by multiple introductions and adaptive introgression

Received: 10 October 2024

Accepted: 22 July 2025

Published online: 25 September 2025

 Check for updates

Xi Wang ^{1,25}, Casia Nursyifa^{1,25}, Sabhrina Gita Aninta ¹, Genís Garcia-Erill ^{1,2}, Laura D. Bertola ^{1,3}, Anubhab Khan ^{1,4}, Josiah Kuja ¹, Kristian Hanghøj ¹, Jonas Meisner ^{1,5,6}, Thomas Bøggild ¹, Corey J. A. Bradshaw ^{7,8}, Amal Al-Chaer¹, Alam Putra Persada ⁹, Dwi Sendi Priyono ¹⁰, Yuli A. Tribudi¹¹, Pita Sudrajad ¹², Cynthia Dewi Gaina ¹³, Yu Jiang¹⁴, Johannes A. Lenstra ¹⁵, Reagan Cauble-Sims¹⁶, Benjamin D. Rosen ¹⁷, Darren E. Hagen ¹⁸, Michael P. Heaton ¹⁹, Timothy P. L. Smith ¹⁹, Laurent Frantz^{20,21}, Greger Larson ²², Mikkel-Holger S. Sinding ¹, Dedy Duryadi Solihin^{9,23}, Muhammad Agil²⁴, Bambang Purwantara²⁴ & Rasmus Heller ¹ ✉

Genetic diversity is a crucial resource in livestock, determining their traits and ability to respond to selection. Indonesian cattle are unique due to their history of admixture involving both zebu (*Bos indicus*) and banteng (*B. javanicus*), and may therefore contain novel cattle genetic resources. We generated whole genome sequences from 126 Indonesian cattle, 51 domesticated banteng and three captive banteng. We show that Indonesian cattle have very high genetic diversity, especially the Madura breed due to introgression from banteng and possibly other *Bos* species, contributing up to 36.6% of the Madura's genome. We find that Indonesian zebu ancestry can be traced to at least three distinct ancestral populations, two of which were introduced more than 1345 years ago from mainland Southeast or eastern Asia. Peaks and valleys in banteng ancestry across the genome in admixed breeds suggest that both negative and positive selection act on introgressed haplotypes. Despite adaptive introgression being mainly breed-specific, we found evidence that some phenotypes, such as coat color, have experienced convergent adaptive introgression. Overall, our results provide insights into the historical movement of cattle in Asia, and showcase the potential for genetic improvement of cattle by identifying ~3.5 million novel SNPs introgressed into Indonesian cattle.

Global genetic diversity in livestock determines their ability to respond to selection and climate change, and their climatic footprint¹. A large proportion of the global genetic diversity in cattle is contained in indigenous breeds, which are generally adapted to their abiotic environment, fodder, parasites, and pathogens. Yet, they are understudied and under pressure from genetic erosion occurring as a result

of cross-breeding with higher-yielding breeds, leading to a loss of valuable genetic resources in cattle as a whole^{2,3}. On the other hand, historical admixture between different cattle breeds, as well as between cattle and other members of the *Bos* genus, has introduced new genetic variation that has helped populations adapt to novel environments, such as in eastern Asia^{4,5} and Africa^{6,7}.

A full list of affiliations appears at the end of the paper. ✉ e-mail: rheller@bio.ku.dk

While recent studies using whole genome sequencing have substantially increased our understanding of the overall evolutionary history of cattle in Asia^{4,5,8–10}, Southeast Asian cattle have not been studied, constituting a large area where the history of cattle is almost unknown. Yet, Southeast Asia is an important region in cattle history, constituting the main dispersal route from India to China^{4,5,11}, and is the likely scene of introgression from endemic wild bovine species into cattle^{4,5,12}. A resolution of the historical movement of cattle in this region would also provide information regarding the early trade routes, political and cultural links within Southeast Asia¹³.

Indonesian cattle are unique examples of stabilized hybrid cattle with substantial contributions from two parental *Bos* species with limited cross-fertility. They contain both zebu (*Bos indicus*) and banteng (*B. javanicus*) ancestry, with different breeds representing potentially different proportions of zebu and banteng^{14–16}. Banteng were probably domesticated as early as 3500 years ago on Java^{17,18}. These domesticated banteng, referred to as Bali cattle, have historically been one of the most abundant and important livestock breeds in the country and remain so today¹⁹. Zebu cattle were introduced to the archipelago later, but the sources and date of introduction are obscure. More recently, taurine cattle was imported to Indonesia from Europe and Australia from the 19th century with the aim of improving local cattle by crossbreeding²⁰. It is commonly assumed that indigenous Indonesian breeds have acquired adaptations to the Indonesian climate, diseases, and to the smallholder husbandry practiced throughout the archipelago^{21,22}. The mixing of indigenous banteng and introduced zebu cattle may have conferred banteng adaptations to the admixed Indonesian cattle by the process of adaptive introgression¹⁶. The unique history and mosaic of ancestry components make Indonesian cattle globally important because they contain genetic variation and adaptations that are not found in any other cattle population.

Admixture is common between species in the *Bos* genus²³, and cases of adaptive introgression of exotic *Bos* ancestry have been identified in East Asian cattle^{4,5,12}. However, in general, introgression from other species is thought to be predominantly under negative selection due to the presence of reproductive barriers between species²⁴. Such barriers can be in the shape of Bateson-Dobzhansky-Muller incompatibilities (BDMIs) or other genetic incompatibility mechanisms, whereby alleles are harmful outside their native genomic context²⁵. As a consequence, minor parent ancestry (i.e., the lesser of two ancestry components in a hybrid population) is usually under negative selection^{26,27}. These studies also highlight that the nature of the selection pressure may differ from case to case depending on the degree of evolutionary divergence between hybridizing populations, the amount of genetic load in each parental source population, and on the environment in which the hybrid population evolved²⁸. Furthermore, in admixed livestock such as Indonesian cattle, the selection regime on introgressing haplotypes might differ from those found in wild species. Indonesian cattle therefore represent an interesting model system to investigate how selection works on introgressed haplotypes in the *Bos* genus, which is a question of fundamental evolutionary importance and potential value in cattle breeding.

In order to elucidate the genomic resources in Indonesian cattle and how they have been shaped by their complex and poorly resolved history, we whole-genome sequenced 233 Indonesian cattle and banteng samples. These were analyzed alongside publicly available data representing global cattle diversity. This allowed us to quantify the population structure, genetic diversity, and admixture processes in Indonesian cattle, as well as the genetic contribution of banteng to present day cattle. We also investigated the genomic landscape of ancestry in the admixed breeds and identified candidates of adaptive introgression, and compared it to the parallel case of wild *Bos* introgression into East Asian cattle^{4,5,12}. Our analyses provide the detailed insights into the origins, history, introgression and selection in a unique and globally important admixture system of cattle.

Results

We generated 233 new whole genomes, out of which 179 are from Indonesian cattle breeds, including Aceh (31), Pesisir (37), Pasundan (26), Jabres (33), Madura (34) and Sumba Ongole (18). A further 51 whole genomes are of Bali cattle from Bali (19), Kupang (17) and the feral population in northern Australia (15); and three captive Javan banteng from Texas, USA. The samples were sequenced to a mean depth of ~13.1X (range 9.62X – 34.9X; Fig. 1a; Supplementary Fig. 1; Supplementary Data 1; Supplementary Data 2). We combined them with 81 publicly available genomes from 2 Javan banteng, 8 Bali cattle, 42 zebu, 27 taurine, and 2 gaur (*B. gaurus*), resulting in 314 samples for mapping and subsequent data filtering (Supplementary Fig. 1; Supplementary Data 1; Supplementary Data 2). After mapping to two reference genomes, taurine cattle (*BosTau9*, GenBank: [GCA_002263795.2](https://doi.org/10.1038/s41467-025-62692-z), ARS-UCD1.2) and water buffalo (*WaterBuffalo*, GenBank: [GCA_003121395.1](https://doi.org/10.1038/s41467-025-62692-z), UOA_WB_1), and sample-quality filtering, we removed 81 samples due to suspected duplication (54 samples) and relatedness (27 samples of first and second degree). Hence, the final dataset consisted of 233 samples, including 157 newly sequenced individuals and 76 downloaded individuals passing sample filtering (Supplementary Data 1; Supplementary Data 2). Rigorous filtering to keep only sites with high confidence genotype calls resulted in 1,486,371,715 (53.9%) remaining sites in the *BosTau9*-mapped datasets, and 1,462,140,103 (55.1%) sites in the *Waterbuffalo*-mapped datasets (Supplementary Data 3). The number of SNPs and reference genome used in each downstream analyses were shown in Supplementary Data 4.

Population structure

We used HaploNet²⁹ to visualize relative genetic distances using principal component analysis (PCA) and to estimate admixture proportions of the individuals (Fig. 1b; Fig. 1d; Supplementary Fig. 2; Supplementary Fig. 3). The first two principal components revealed three main clusters of ancestry, corresponding to banteng, zebu, and taurine cattle, with a gradient of admixed samples between zebu and taurine cattle also evident in the admixture analysis using three ancestry components ($K=3$; Fig. 1d). The Indonesian cattle (Madura, Pasundan, Pesisir, Jabres, Aceh, Sumba Ongole) were located in the zebu cluster in the PCA; however, there was a tendency towards the banteng cluster, especially in Madura cattle, which is also observed in the admixture proportions at $K=3$ (Fig. 1b; Fig. 1d). East Asian zebu also showed a tendency towards the banteng cluster in the PCA (Fig. 1b), but banteng admixture in this population was inferred to be only 1.26% in the admixture analysis at $K=3$ (Fig. 1d; Supplementary Data 5), which is much lower than previously reported⁴. This apparent underestimation of banteng ancestry in East Asian zebu and in some of the Indonesian breeds at $K=3$ could be due to an unknown ('ghost') source of admixture, which is not well represented by the Javan banteng ancestry component represented by our samples, consistent with Dai et al.¹². Three Pasundan individuals had substantial admixture from the taurine cluster (Fig. 1b; Fig. 1d), suggesting recent cross-breeding. All Bali cattle clustered with banteng except for two individuals from Kupang (N_31B and N101B), one of which had equal admixture from banteng and taurine cattle (Fig. 1b; Fig. 1d), suggesting it is a hybrid between these two lineages.

The admixture analysis at $K=7$, the highest value of K that converged, assigned each Indonesian breed to a separate genetic cluster, except for Jabres and Pasundan, both of which we inferred to be admixed between Madura and Sumba Ongole (Fig. 1d), consistent with the recent formation of Pasundan in 1907 by crossbreeding Madura with Ongole-type cattle³⁰. Pasundan was more heterogeneous than other breeds, which is consistent with reports that Pasundan consists of individuals with more variable morphological features than Bali cattle, Madura, or Sumba Ongole³¹. Similarly, Aceh (1) and Pesisir (5) contained individuals with uncharacteristic admixture proportions,

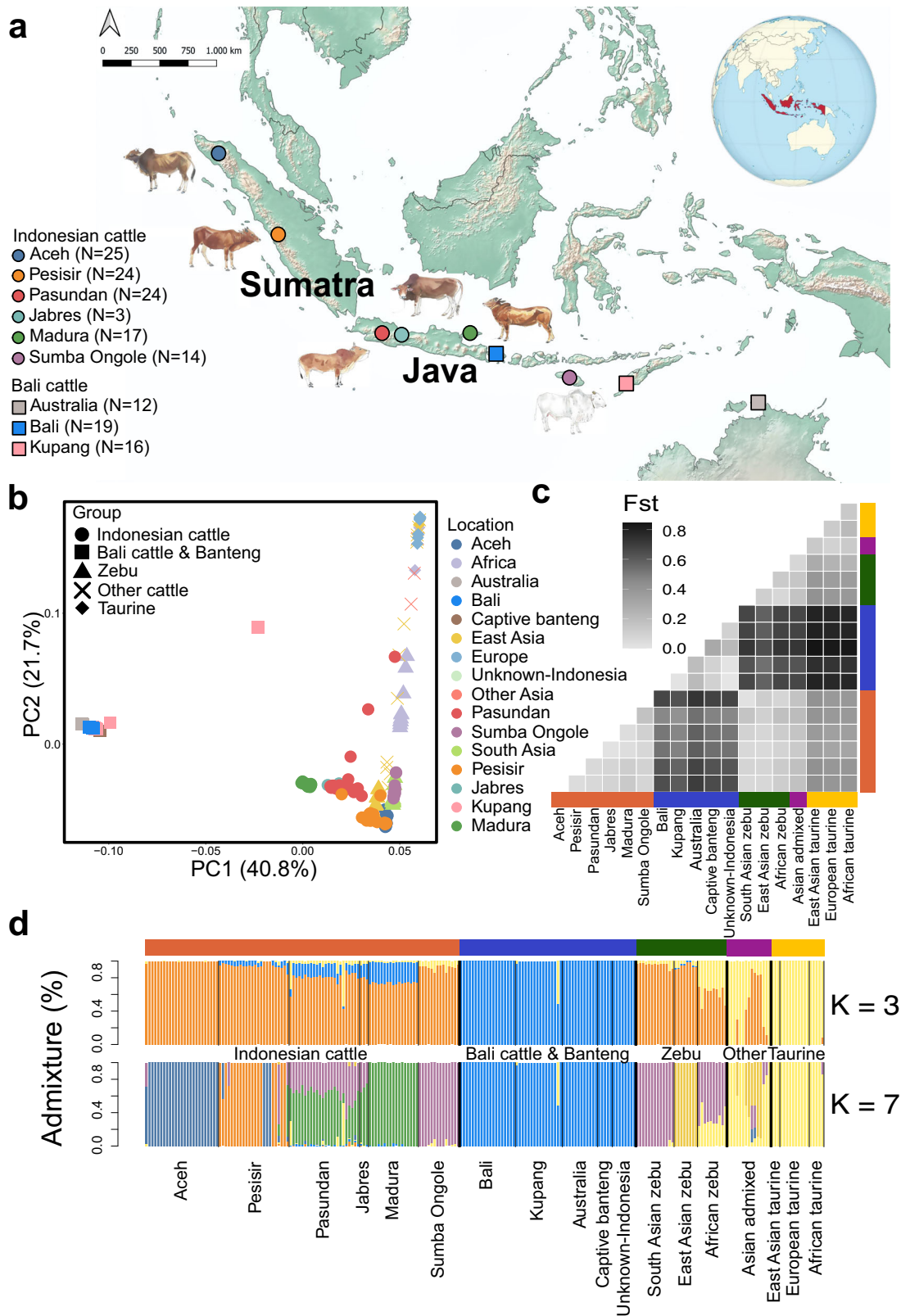


Fig. 1 | Sampling and population structure. **a** Sampling locations of the sequenced Indonesian cattle and Bali cattle. **b** Visualization of 231 samples using two principal components based on genetic distance inferred by HaploNet. One individual (N_31B) had equal admixture from banteng and taurine cattle, referred to

as the putative F1 hybrid. **c** Global genetic differentiation measured as F_{ST} values using Hudson's estimator between all population pairs. **d** Individual ancestry proportions inferred by HaploNet for $K = 3$ and $K = 7$.

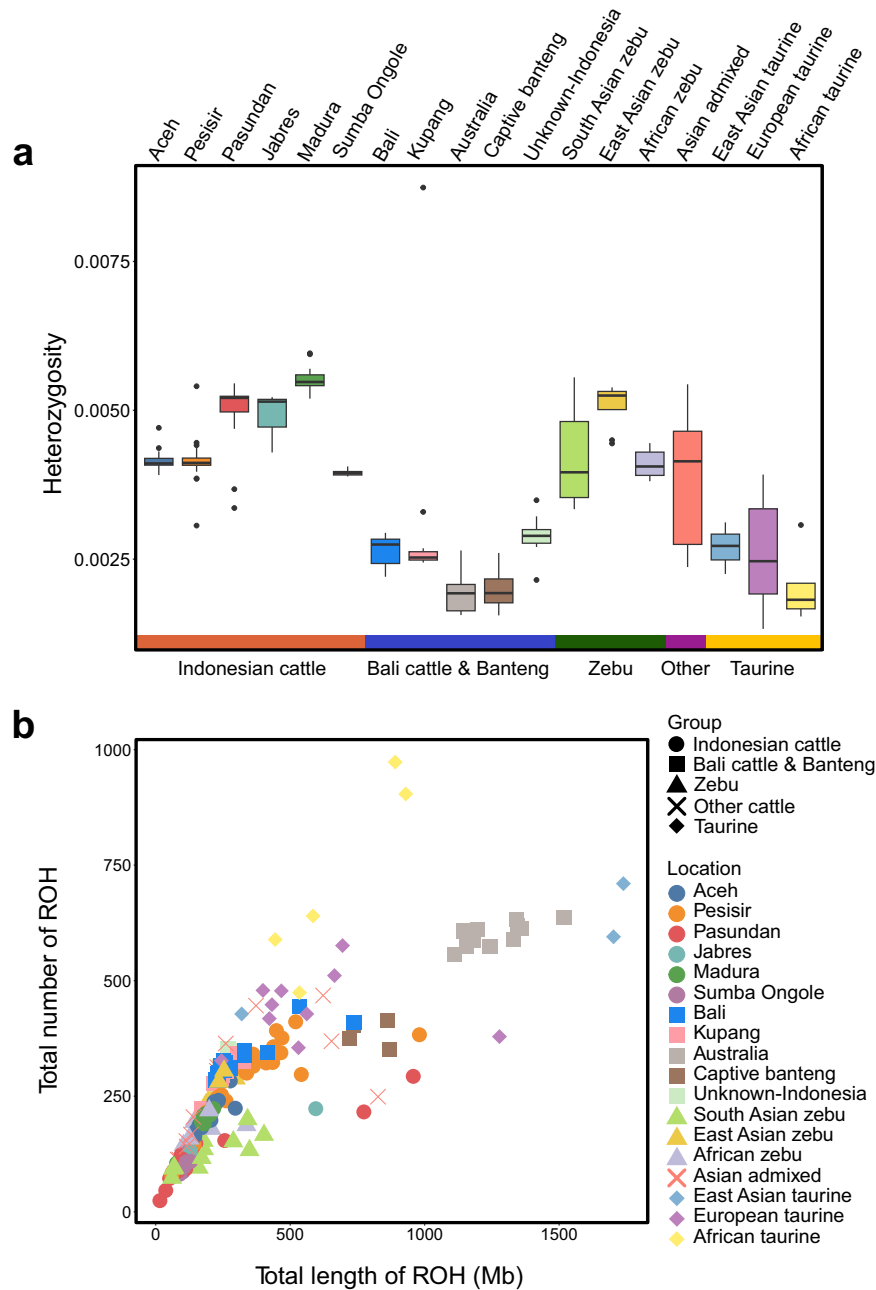


Fig. 2 | Genetic diversity and inbreeding. a Heterozygosity of all Indonesian cattle and other cattle based on genotype data with non-variable sites included (See Methods). Boxplots indicate median (centre line), the 25th and 75th percentiles (box), and the highest and lowest values within the upper and lower quartiles $\pm 1.5^*$ interquartile range, respectively (whiskers). **b** The total number of ROH segments (y axis) and the total length (Mb) of the genome in ROH (x axis) for all 231 samples.

Each dot represents an individual. Sample sizes of the populations included in (a, b) are as follows: Aceh ($n = 25$), Pesisir ($n = 24$), Pasundan ($n = 24$), Jabres ($n = 3$), Madura ($n = 17$), Sumba Ongole ($n = 14$), Bali ($n = 19$), Kupang ($n = 16$), Australia ($n = 12$), Captive banteng ($n = 5$), Unknown-Indonesia ($n = 8$), South Asian zebu ($n = 13$), East Asian zebu ($n = 8$), African zebu ($n = 10$), Asian admixed ($n = 15$), East Asian taurine ($n = 3$), European taurine ($n = 10$), and African taurine ($n = 5$).

suggesting that some recent cross-breeding has taken place. Even at $K = 7$, we found no substructure within the banteng-Bali cattle group.

We found low genetic differentiation among three Javan breeds (Pasundan, Jabres, and Madura; $F_{ST} = 0.01$ – 0.04), and between the two Sumatran breeds (Aceh and Pesisir; $F_{ST} = 0.04$) (Fig. 1c; Supplementary Fig. 4). Similarly, we found almost no genetic differentiation ($F_{ST} = 0.00$ – 0.01) among the Bali cattle populations, with the exception of the Australian Bali cattle, consistent with its recent strong founder event³². Differentiation was generally high between banteng and unadmixed cattle populations ($F_{ST} = 0.62$ – 0.85), but several Indonesian breeds had reduced F_{ST} with banteng, especially Madura ($F_{ST} = 0.42$ – 0.54). A NeighborNet tree based on the pairwise global F_{ST}

distances conform with the PCA and admixture analyses, and suggests reticulation in the evolutionary relation between the populations (Supplementary Fig. 5).

Genetic diversity and inbreeding

We estimated genome-wide heterozygosity in all 231 samples (Fig. 2a). We found the highest median heterozygosity in Madura cattle (0.00548), followed by other admixed or unadmixed zebu cattle: East Asian zebu (0.00525), Pasundan (0.00521), Jabres (0.00515), Aceh (0.00411), Pesisir (0.00412), African zebu (0.00406), and South Asian zebu (0.00396). Madura cattle have even higher genetic diversity than East Asian zebu, previously considered the most genetically diverse

cattle in the world due to introgression from other *Bos* species^{4,33}. Taurine cattle had lower heterozygosities, consistent with previous findings^{34–37}, whereas banteng and Bali cattle had the lowest heterozygosity of all populations.

We also analyzed runs of homozygosity (ROH) as a measure of genomic autozygosity and recent inbreeding. A visualization approach similar to Liu et al.³⁸ was taken to assess different settings for the ROH calling and to validate that our final ROH calling approach was accurate (Supplementary Fig. 6). While we emphasize that ROH calling is highly sensitive to even small errors in genotype calling, our approach was able to detect most of the apparent true ROHs, and evidently able to identify meaningful differences in ROH length and abundance between different samples. Most of the inferred ROHs were relatively short (0.5–2 Mb), except for captive banteng and Australian Bali cattle, which were dominated by long ROHs (>5 Mb), indicative of recent inbreeding in these individuals (Fig. 2b; Supplementary Fig. 7). We also observed a few individuals in Pasundan and Jabres with large fractions of long ROHs (>10 Mb), indicating occasional, rather than prevalent, cases of close-kin mating in Indonesian cattle (Supplementary Fig. 7). In banteng-Bali cattle, heterozygosity across populations was more uniform when removing ROH regions from the heterozygosity estimation (Supplementary Fig. 8), confirming that inbreeding and recent demographic events play a large role in explaining the variation in genetic diversity across these populations. In contrast, genetic diversity in the cattle breeds is more driven by other factors, such as older demographic history or potentially the amount of admixture.

Admixture and origin of Indonesian cattle

To assess the origin and history of admixture events in Indonesian cattle, we performed a series of analyses after removing individuals with evidence of recent admixture. We first used TreeMix on the Indonesian cattle, banteng, and representative zebu and taurine cattle populations to infer the population tree assuming different numbers of admixture events. This analysis showed that at least five admixture events are required to explain the inter-population genetic variation (Supplementary Fig. 9), and we inferred extensive admixture events from the banteng branch into Indonesian and East Asian cattle (Fig. 3a). TreeMix also suggested three distinct lineages of Indonesian cattle, with the Sumatran cattle breeds (Aceh and Pesisir) being separated from the Javan breeds (Madura, Jabres, and Pasundan) and Sumba Ongole. D-statistics analyses confirmed that all Indonesian cattle and East Asian cattle have excess allele sharing with banteng relative to South Asian zebu (Fig. 3b). Conversely, the different Bali cattle populations had similar amounts of allele sharing with zebu cattle as did banteng, with the exception of Bali cattle from Kupang, which showed a weak but significant signal of excess allele sharing with zebu, indicating introgression from zebu cattle into this population after its divergence from other Bali cattle populations.

To quantify the amount of banteng-like introgression in admixed cattle populations, we estimated genome-wide admixture proportions using local ancestry inference with LOTER³⁹ and F4-derived admixture ratio⁴⁰. The pattern was consistent with D-statistics and other results, and suggested that Madura individuals have very high banteng ancestry proportions of 0.33–0.37 (Fig. 3c; Supplementary Data 6). Banteng ancestry proportions were generally homogeneous within breeds, suggesting that the admixture is relatively old and evenly distributed across individuals, except in Pasundan where we found 0.13–0.26 banteng ancestry (Fig. 3c; Supplementary Data 6), consistent with the recent formation of this breed and the variable proportions of Madura and Ongole ancestry in Pasundan individuals (Fig. 1d). Overall, we obtained similar proportions using both F4 ratio and LOTER. However, the F4 ratio was lower than the LOTER-based results in Madura (LOTER_{median} = 0.35, F4_{median} = 0.29), Pasundan (LOTER_{median} = 0.23, F4_{median} = 0.19) and Jabres (LOTER_{median} = 0.24, F4_{median} = 0.20), possibly due to small amounts of taurine

introgression into these populations, or other violations of the F4 assumptions. Correspondingly, the F1 hybrid offspring of a Bali cattle and a taurine cattle from Kupang showed a much-reduced F4 ratio as expected (Fig. 3c; Supplementary Data 6).

To disentangle the history of introgression from other *Bos* sources, we used three complementary approaches. First, we used Hmfix⁴¹ to identify regions in the genome that were introgressed from highly divergent source populations into the six admixed Asian cattle breeds. We then calculated the genetic similarity of each introgressed region with two different potential source lineages, Javan banteng, and gaur. We found that the average genetic similarity with these two source species in introgressed regions differed substantially among breeds, with Javan breeds having higher similarity with the Javan banteng source, and the Sumatran breeds and East Asian zebu having higher similarity with gaur (Fig. 4a). Aceh cattle and East Asian zebu had almost identical average genetic similarity patterns, whereas Pesisir cattle had patterns intermediate between the Javan breeds and Aceh. Furthermore, the genetic distance calculated within introgressed regions shared between pairs of individuals showed a similar pattern (Fig. 4b; Supplementary Fig. 10), suggesting that the three Javan breeds have a distinct introgression source from the two Sumatran breeds, which are in turn more similar to East Asian zebu.

Next, we inferred the introgression time using AncestryHMM⁴² (Fig. 4c) and again found distinct differences between the Javan breeds and the other admixed Asian cattle, with a younger inferred admixture time of 269 generations for Madura, corresponding to 1345–1883 years ago assuming a generation time of 5–7 years^{4,43,44}. Pesisir and especially Aceh showed substantially older introgression times (407 and 1051 generations, respectively), again suggesting that the introgression into Sumatran breeds was a separate event from the introgression into Javan breeds. We note that AncestryHMM tends to overestimate the age of the introgression if the source is a ghost population not well represented by the proxy donor population data used in the analysis⁴⁵, as already indicated by the HaploNet (Fig. 1d) and Hmfix (Fig. 4a) results. Therefore, the older inferred dating of introgression in the Sumatran breeds, and even more so in the East Asian zebu, could be explained if the *Bos* source deviated significantly from that represented by our Javan banteng data – e.g., one or more of the *Bos* lineages found in mainland Southeast Asia, as suggested before for East Asian zebu³³. Complicated introgression involving more than one species of wild *Bos* has been suggested for East Asian zebu before¹².

Last, to disentangle the origin of the zebu and banteng ancestry components of the admixed Asian breeds, we used local ancestry inference to mask sequentially one of these two ancestries in the individual genomes, and performed PCA using a method targeted for datasets with very high missingness on the resulting genotype matrices. We checked how well the local ancestry masking works by redoing the D-statistic analysis applying the masking, and confirmed that this masking removes all significant traces of banteng-like ancestry (Supplementary Fig. 11a), bringing the D-statistics close to zero for all admixed populations. The masked EMU-PCA analyses showed that the banteng ancestry groups into three different clusters: (1) extant Javan banteng and Bali cattle, (2) Javan cattle, and (3) Sumatran cattle and East Asian zebu (Fig. 4e; Supplementary Fig. 12; Supplementary Fig. 13). The zebu ancestry groups into five main clusters represented by (i) South Asian zebu and Sumba Ongole, (ii) Madura, (iii) Pesisir, (iv) Aceh, and (v) East Asian zebu, respectively (Fig. 4e; Supplementary Fig. 12; Supplementary Fig. 13). The other Javan breeds – Pasundan and Jabres – fall on a cline between Madura and Ongole cattle, consistent with them being a recent admixture of these two sources. The addition of new downloaded samples from northernmost Southeast Asia and East Asia (e.g. Myanmar and southern China, Supplementary Data 7) did not resolve the origins of Indonesian zebu ancestry, but showed a cline of zebu ancestry consistent with their geographical placement on the shortest route between India and China (Supplementary Fig. 11b).

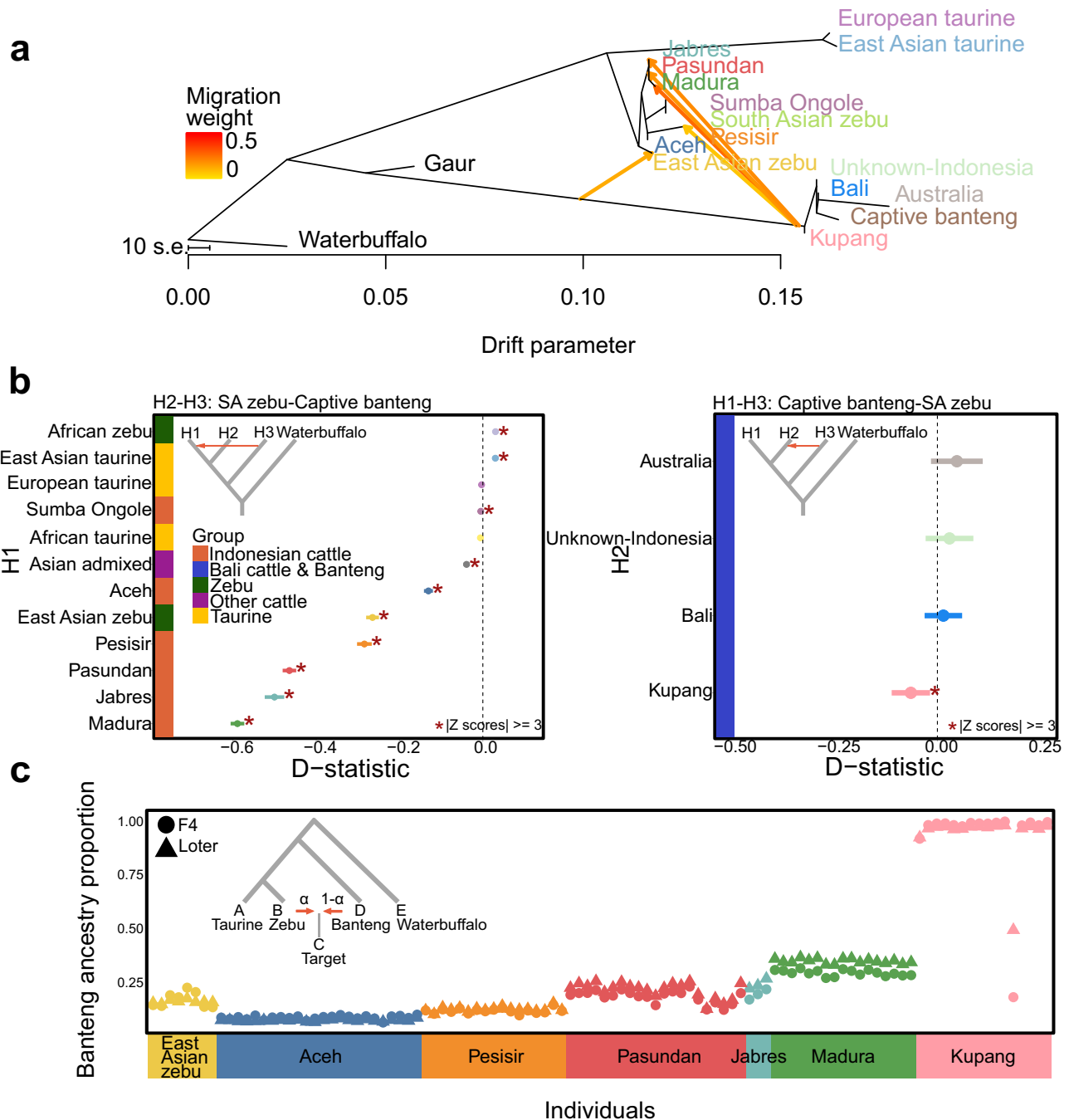


Fig. 3 | Population history and introgression. **a** Population tree inferred with TreeMix assuming five migration events. We used a water buffalo as an outgroup. Arrows represent migration edges, with colour indicating the migration weight (proportion of the admixed population estimated to derive from the source population). **b** D-statistics calculated by ADMIXTOOLS2 when using water buffalo as outgroup H4, various cattle populations as H1/H2. Left panel: South Asian zebu and captive banteng as H2 and H3; right panel: captive banteng and South Asian zebu as H1 and H3. A significant non-zero value, as depicted by the red arrow in the graphic for each panel, provides evidence for gene flow between H3 and H1 (Left panel), and between H3 and H2 (Right panel). Data are presented as the estimated

D-statistic \pm 3 standard errors. Star represents significant allele sharing for each combination. Sample sizes for each population are as follows: Aceh ($n = 24$), Pesisir ($n = 17$), Pasundan ($n = 21$), Jabres ($n = 3$), Madura ($n = 17$), Sumba Ongole ($n = 14$), Bali ($n = 19$), Kupang ($n = 14$), Australia ($n = 12$), Captive banteng ($n = 5$), Unknown-Indonesia ($n = 8$), South Asian zebu ($n = 11$), East Asian zebu ($n = 8$), African zebu ($n = 10$), Asian admixed ($n = 15$), East Asian taurine ($n = 3$), European taurine ($n = 10$), and African taurine ($n = 5$). **c** Individual admixture proportion from LOTER and F4. We used unadmixed zebu and banteng as the two ancestry source references for LOTER. We used water buffalo and taurine as outgroups for the F4 ratio admixture inference.

As a measure of polymorphism within and between the source populations of admixed individuals, we calculated heterozygosity in each of the three ancestry states along the genome in admixed breeds: tracts homozygous for zebu ancestry, tracts heterozygous for zebu/banteng ancestry, and tracts homozygous for banteng ancestry (Fig. 4d; Supplementary Fig. 14). The results highlight how

introgression elevates the overall genetic diversity, but also showed that the banteng source contributing to Madura had much higher genetic diversity (median heterozygosity: 0.0043) than in any extant banteng populations (median heterozygosity in captive banteng: 0.0019; in Bali cattle: 0.0027). This suggests that the formation of the Madura breed did not entail a major bottleneck, and that the banteng

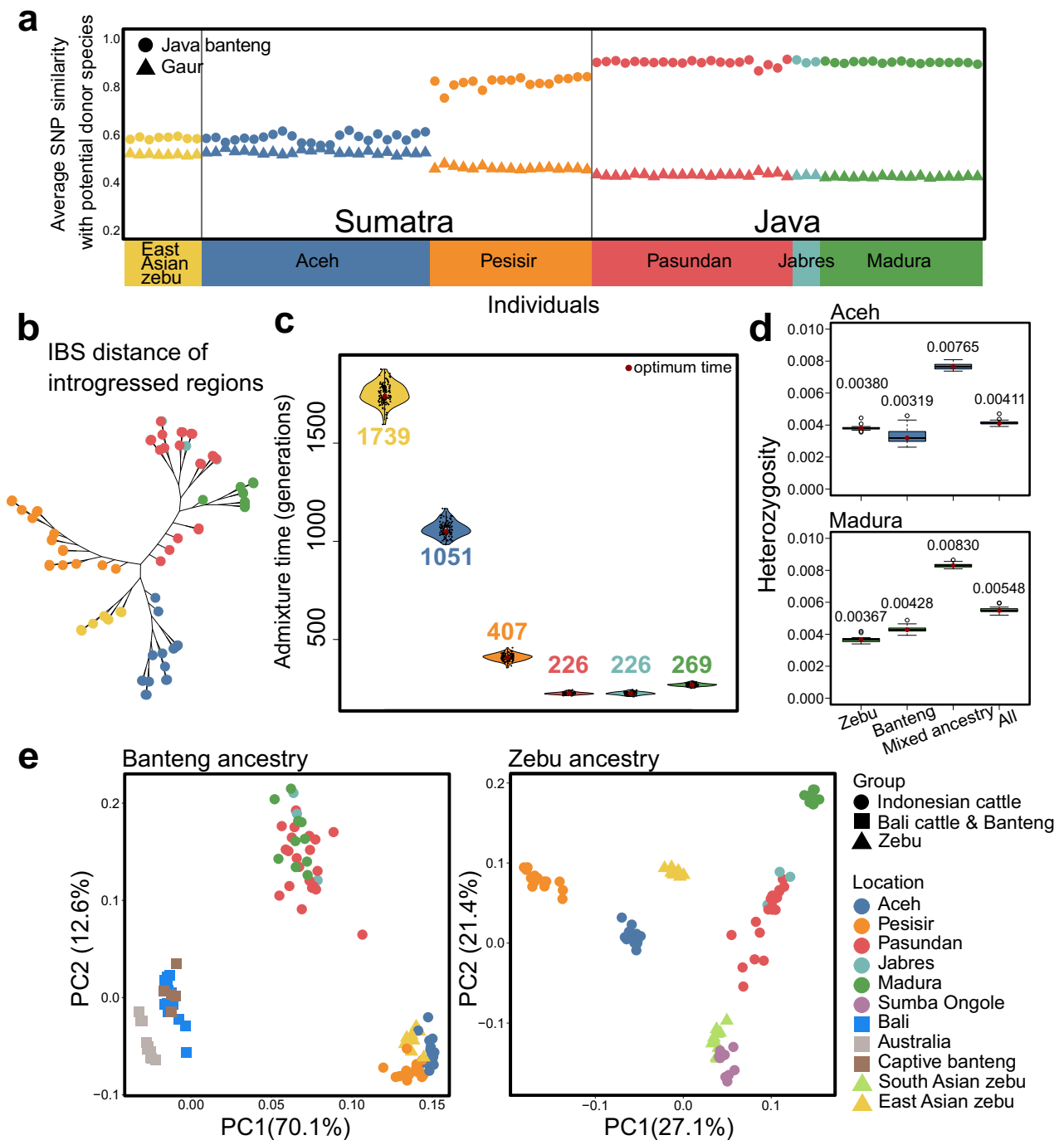


Fig. 4 | Origin of introgressed fragments in Indonesian cattle. **a** The mean ratio of SNPs shared with two outgroups, Javan banteng and gaur, in inferred introgressed regions of each admixed cattle population, based on Hmfix. **b** Pairwise identity by state matrix (mibs) calculated from overlapping archaic regions (probability > 0.9) between two individuals using a 10 Kb window size. Insert: neighbor-joining tree of the genetic distance (1-mibs) in pairwise overlapping archaic regions. **c** Inference of admixture time using AdmixtureHMM with one pulse model. Numbers shown are average generation times across 100 bootstraps. **d** Genome-wide heterozygosity in two admixed populations, Aceh ($n = 24$) and Madura ($n = 17$), when stratified by the three possible ancestry states: homozygous

for zebu ancestry (“Zebu” on x-axis), homozygous for banteng ancestry (“Banteng”), heterozygous for zebu/banteng ancestry (“Mixed ancestry”) and finally across the whole genome (“All”). Results for all admixed populations and East Asian zebu are in Supplementary Fig. 14. Boxplots indicate median (centre line), the 25th and 75th percentiles (box), and the highest and lowest values within the upper and lower quartiles $\pm 1.5^*$ interquartile range, respectively (whiskers). **e** PCA analysis on genome segments of banteng ancestry origin and zebu ancestry origin respectively, as inferred by LOTER. To remove any distant relatedness among samples, we removed one of each pair of individuals with $K1 > 0.2$ identified by ngsRelate for (b, d) (Supplementary Fig. 31).

population contributing to Madura had a higher genetic diversity than the ones found today. Aceh and Pesisir showed lower diversity in the banteng ancestry tracts compared with those in Madura and Pasundan, further supporting the different introgression history between Sumatran breeds and Javan breeds (Fig. 4d; Supplementary Fig. 14).

In summary, our analyses suggest three distinct origins of the Indonesian cattle: (1) Sumatran breeds Aceh and Pesisir, which have similar zebu and banteng ancestries as East Asian zebu and show signs of either old and/or ghost *Bos* introgression from mainland Southeast Asia; (2) Javan breeds Madura, Pasundan, and Jabres, which have

introgression that occurred more recently and from a source more similar to extant Javan banteng; and (3) Sumba Ongole, which does not have appreciable *Bos* introgression and is genetically similar to South Asian zebu, consistent with its recent introduction²⁰.

Genomic landscape of ancestry

mtDNA haplotypes in Indonesian breeds were found both within banteng and cattle haplogroups, whereas Y-chromosomal haplotypes were exclusively of cattle origin (Supplementary Fig. 15; Supplementary Fig. 16). This suggests that interbreeding did not involve male banteng, whereas both male and female cattle were involved, consistent with previous evidence¹⁵. In addition, the mtDNA haplotype network corroborated a closer affinity of the Indonesian breeds, except Sumba Ongole, with East Asian than with South Asian zebu (Supplementary Fig. 15).

To investigate the distribution of different ancestries across the genome in admixed cattle, we used the local ancestry inference results from the five admixed Asian cattle breeds with sufficient sample sizes (Aceh, Pesisir, Pasundan, Madura, and East Asian zebu). Preliminary analyses concluded that the local ancestry inference was robust to the composition of the reference panels used in LOTER (Supplementary Fig. 17) and to the choice of reference genome (Supplementary Fig. 18). Based on the local ancestry inference from each individual, we calculated the mean banteng ancestry in non-overlapping 50 Kb windows for each population. We visualized the genome-wide local ancestry landscape for each admixed breed (Fig. 5a; Supplementary Fig. 19). We found noticeable peaks and valleys, suggesting that selection (positive as well as negative) might have been acting on introgressed haplotypes. Some large-scale features are visible; for example, there are two large-scale ‘deserts’ of banteng ancestry on chromosome 7 (-5–25 Mb and 40–55 Mb), and another one on chromosome 16 (-2–12 Mb). These decrease the average banteng ancestry markedly on the entire chromosomes 7 and 16 relative to other autosomes (Supplementary Fig. 20).

In addition, banteng ancestry was severely depleted on the X chromosome across breeds (Supplementary Fig. 21), consistent with sex chromosomes harboring more incompatibility loci than autosomes^{28,46}. To investigate the overall type of selection acting on haplotypes of banteng (minor) ancestry in a zebu (major) ancestry background, we calculated the correlation between the mean banteng ancestry and three genomic features: recombination rate, conservation score, and coding region density in Madura cattle (Supplementary Fig. 22). These analyses revealed a very weak positive correlation between recombination rate and banteng ancestry, a very weak negative correlation between conserved sites and banteng ancestry, and a very weak positive correlation between banteng ancestry and coding sites density, collectively indicating a lack of linked purifying selection removing banteng ancestry overall, contrasting with the results from hybridizing swordtail (*Xiphophorus* spp.) and hominin populations²⁴.

We observed an overall high correlation in banteng ancestry proportions per 50 Kb window even between breeds that we inferred to have independent introgression events – e.g., East Asian zebu and Madura (Spearman’s $\rho = 0.30$; Fig. 5c). The correlation was markedly higher between Madura and Pasundan ($\rho = 0.58$; Fig. 5c), and between Aceh and Pesisir ($\rho = 0.49$; Fig. 5c), both inferred to share introgression history.

Regions of high banteng ancestry in admixed cattle

To investigate the functional consequences of introgression, we identified regions of high banteng ancestry (top 5%) within breeds using 50 Kb non-overlapping windows across the genome. We used three complementary methods of local ancestry inference (LOTER, Hmix, and U_x), which showed consistent but complementary results (Supplementary Fig. 23, see also Methods). In total, the top 5% proportion of population-wise banteng ancestry as inferred from LOTER overlaps with 839 genes in Madura, 813 genes in Pasundan, 892 genes

in Pesisir, 1041 genes in Aceh, and 914 genes in East Asian zebu (Fig. 5d), all significantly lower gene density than expected by chance (Supplementary Fig. 24). These results suggest that elevated banteng ancestry tends to occur in relatively gene-poor regions. Gene ontology enrichment tests identified multiple immune system-related enrichment categories across breeds, and developmental processes (Supplementary Data 8). In addition, the outlier regions overlap with many production relevant quantitative trait loci that differed among populations, notably milk-related (e.g., ‘milk fat percentage’, ‘milk yield’), fertility-related (‘calving ease’, ‘conception rate’), and body size or growth-related traits (‘body weight’, ‘average daily gain’) (Fig. 5e). We additionally reported the gene list targeted by top 5% outliers that were overlapping between any two out of three methods (LOTER, Hmix, and U_x) and a Venn diagram plot showing the number of overlapping genes for each Indonesian breed (Supplementary Fig. 25; Supplementary Data 9–11).

Shared outlier regions and convergent adaptive introgression

Identifying regions of convergent negative or positive selection acting on introgressing haplotypes in admixed cattle can be valuable for assessing the potential for genetic improvement of cattle through gene editing or controlled crossing. To investigate signals of shared or convergent adaptive introgression from banteng, we tabulated the overlap between genes in the top 5% windows with banteng ancestry across the five admixed breeds with sufficient sample sizes (Fig. 5d; Supplementary Data 12). The magnitude of gene sharing between populations followed the same trends as the overall genomic correlations in banteng ancestry (Fig. 5c), with the exception that the Pesisir and Aceh (Sumatran) breeds shared more outlier genes (195) than any other population pair (Fig. 5d).

Among the genes that have elevated (top 5%) banteng ancestry across multiple breeds, several are well-known for their associations to particular cattle phenotypes (Supplementary Fig. 19). These include the well-known coat-color genes *ASIP* (Pasundan, Pesisir, Aceh, and East Asian zebu), *KITLG* (Madura, Pesisir, Aceh, and East Asian zebu), *TYR* (Pesisir, Aceh and East Asian zebu), and *KIT* (Pesisir, Aceh, and East Asian zebu). Lactase *LCT* and myostatin *MSTN* were also in outlier windows in four populations, (Madura, Pesisir, Aceh, and East Asian zebu), although the same windows also contain other genes that are plausible targets of selection, e.g., *MCM6* which is also associated with coat color⁴⁷ and is positioned in the *LCT* window. *LCT* and *MSTN* have well-known roles in lactose digestion^{48,49} and the growth of muscle fibre⁵⁰, respectively. In addition, the bovine MHC region (BoLA), which is strongly implicated in the immune system of all vertebrates⁵¹, was also among the regions with highest summed normalized banteng ancestry score (Fig. 5a; Supplementary Fig. 19). Of these regions, *ASIP* and *KIT* regions were also clearly identified by the U_x statistic (Fig. 5b; Supplementary Fig. 26), corroborating that these two color coding genes are examples of convergent adaptive introgression. Both genes are within extensive genomic regions that have elevated banteng ancestry across all admixed cattle populations (Fig. 5b; Supplementary Fig. 19). The window containing *ASIP* and the one immediately downstream had the two highest summed normalized banteng ancestry across the genome, with East Asian zebu, Aceh, and Pesisir having particularly elevated banteng ancestry in this region (Fig. 5a; Supplementary Fig. 19). When investigating the haplotype structure in *ASIP*, we identified several banteng-like haplotypes segregating among the admixed cattle populations (Supplementary Fig. 27), suggesting introgression of different *ASIP*-carrying haplotypes, some of which are not common in extant Javan banteng and Bali cattle. We discovered nine non-synonymous mutations (Supplementary Data 13) unique to the banteng-derived *ASIP*. Two of these are classified as potentially high impact with regards to protein structure.

The remaining highlighted gene regions identified with LOTER have less straightforward interpretations, as they could represent

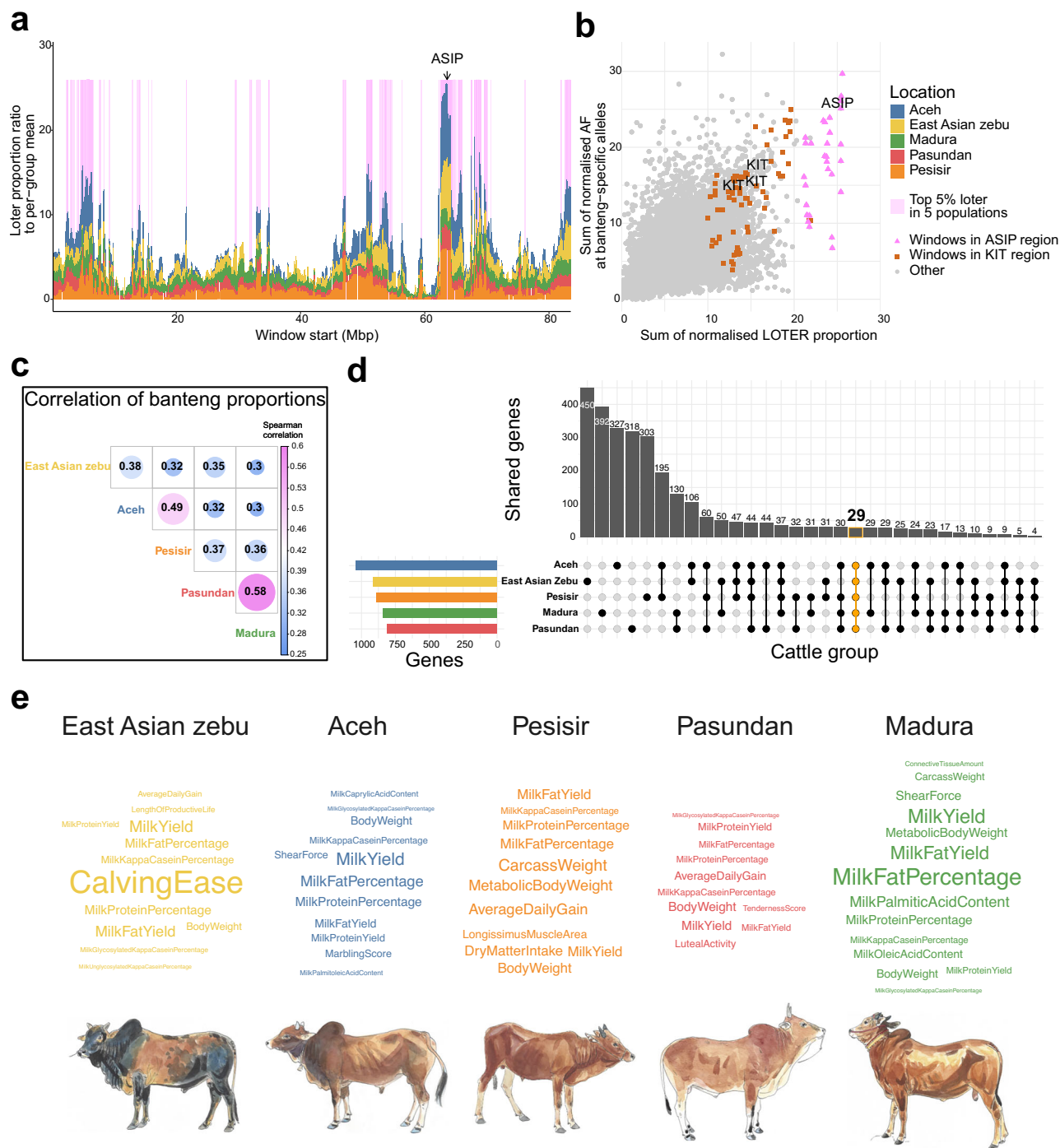


Fig. 5 | Ancestry landscape of Indonesia cattle. **a** Banteng ancestry across chromosome 13 in five cattle groups (Aceh, East Asian zebu, Madura, Pasundan, and Pesisir). Window-based scan of regions with extreme banteng ancestry using the proportion of inferred banteng SNPs from LOTER for each cattle group divided by the mean proportion per group. Pink shade marks regions in the genome-wide top 5% of the normalized LOTER summed across all groups. Similar plots for other chromosomes are in Supplementary Fig. 19. **b** LOTER and U_x results across cattle groups, showing windows in *ASIP* (chr 13, 62700000–64250000) and *KIT* region (chr 6, 64900000–71100000). The windows containing the actual coding region

of *ASIP* and *KIT* are highlighted with the gene name. **c** Spearman correlations of banteng proportion in all windows of 50 Kb. **d** Upset plot of genes within top 5% highest banteng proportion windows for five admixed groups (Aceh, Pesisir, Pasundan, Madura, East Asian zebu). The number in each bar represents the number of genes between different breed comparisons. **e** Word clouds of the QTL terms contained within the top 5% of LOTER inferred banteng ancestry for each group. We included only the top 5% of terms overlapping most frequently with regions of high banteng ancestry for each cattle group.

regions with incomplete lineage sorting or with more complex haplotype structure complicating the assignment of regions to either a banteng or zebu origin. This does not exclude that haplotypes in these gene regions are adaptively introgressed as they are consistently identified as closer to existing banteng than zebu haplotypes across admixed breeds.

Discussion

We generated the whole-genome data set from Indonesian cattle, a group of diverse breeds of interest for the history of cattle and their unique genetic variation caused by wild *Bos* introgression. Although admixture between zebu and banteng in Indonesian cattle is not a

novel finding^{14,15,23,52,53}, the admixture proportions have so far not been accurately estimated, and the evolutionary history of different breeds has not been investigated, nor has the genomic landscape of ancestries. Our findings uncover several novel and relevant aspects for livestock genetics – we identify 3,454,058 novel introgressed SNP variants in cattle (Supplementary Data 14), and we have provided a genomic map of banteng introgression both within and across a number of cattle breeds with unique introgression histories, potentially useful for genetic improvement in cattle.

The magnitude of wild *Bos* introgression in Indonesian cattle is unprecedented in any other cattle populations, and it makes these populations of global importance for cattle genetic resources. Similar to East Asian cattle, banteng-like introgression has elevated the genetic diversity in Indonesian cattle relative to non-admixed zebu breeds^{4,12}, and in Madura we found the highest genetic diversity measured to date in any cattle population. We found that the ancestor of Madura cattle was the result of a pulse-like admixture between an introduced zebu and a genetically diverse endemic *Bos javanicus* population around 1345–1883 years ago. Our genetic dating of this event is remarkably consistent with the historical record⁵⁴ and coincides with the earliest archaeological evidence of cattle on Java²⁰. The high diversity of the banteng-derived introgressed regions, combined with the fertility barriers between banteng and zebu requiring several generations of backcrossing to achieve full fertility in hybrids¹⁹, both point towards a large-scale and possibly human-controlled practice of interbreeding. We also demonstrated that these admixed archaic Javan cattle were subsequently mixed with an Ongole-like zebu population to form the two West Javan breeds, Pasundan and Jabres, which therefore inherited significant amounts of banteng ancestry from Madura-like ancestors.

In contrast, we found uniformly low genetic diversity in banteng and Bali cattle. The captive banteng and the feral Bali cattle in Australia had particularly low genetic diversity and up to 61% of their genome in runs of homozygosity. Interestingly, the banteng ancestry component in the Javan cattle breeds was different from that in extant *Bos javanicus*, with higher genetic diversity than the extant individuals. This indicates that Javan banteng might have experienced a reduction in genetic diversity since the ancestral population introgressed into the admixed Javan cattle about 1345–1883 years ago, and that extant Indonesian cattle therefore carry otherwise lost *B. javanicus* genetic variation, potentially available for genetic rescue. Besides Bali cattle in Kupang, where we identified a weak signal of historical introgression, as well as an F1 hybrid and one other recently admixed individual, we found no indication of introgression of cattle alleles into other Bali cattle and banteng individuals. This alleviates previous concerns that Bali cattle are experiencing genetic swamping from cattle²². However, the reduced genetic diversity and high inbreeding coefficients in Bali cattle do warrant concerns about breeding practices⁵⁵, and suggest that more focus is needed to prevent inbreeding depression in this important local livestock. The feral population of Australian Bali cattle has been proposed as a possible back-up population for wild Javan banteng, considering that it is the largest existing wild population of banteng^{32,56}. However, we corroborated that this population has strongly reduced genetic diversity and increased inbreeding⁵⁶ as a result of its small founder population of just 20 individuals^{57,58}, casting doubt on its utility as a back-up population without deliberate intervention using genetic rescue⁵⁹.

We identify three separate origins and introductions of cattle to Indonesia: two early introductions (>1345 years ago) giving rise to the breeds now on Java and Sumatra, respectively, and one recent and well-documented introduction of Ongole cattle from India²⁰. Due to the absence of genomic data from most of mainland Southeast Asia, we could not infer the geographical source of the zebu founders of the Javan and Sumatran cattle, but our analyses show that the Sumatran zebu ancestry is more similar, but not identical, to East Asian zebu, and that Sumatran cattle share features of *Bos* introgression with East Asian

zebu. On the other hand, the Javan zebu ancestry is more distant from any other known Asian cattle, including South Asian and East Asian, and Javan breeds show a distinct age and source of *Bos* introgression. In conjunction with mtDNA similarities (Supplementary Fig. 15) and previously inferred SNP genotype similarities among Sumatran, Thai, and Chinese breeds^{52,53,60}, we conclude that the most likely source of the early introduction of cattle to Sumatra was mainland Southeast or eastern Asia, in contrast to previous hypotheses based on strong cultural and trade links between the Indian subcontinent and Indonesia in this period^{15,20}. Hence, our results support the presence of extensive connectivity between mainland and maritime Southeast Asia in the period before 700 AD, possibly through Indianized empires in Southeast Asia with extensive maritime trading networks⁶¹. In contrast, our ability to trace the source of Javan zebu cattle is limited, and we simply note that Javan cattle are highly unique compared to all other Asian cattle with genomic data, both with respect to their zebu ancestry and their introgression signature. Interestingly, recent linguistic research established that the ancient Javanese word for zebu cattle derives from a Mon-Khmer language, consistent with cattle being introduced to Indonesia through Indianized cultures in mainland Southeast Asia⁶².

Consistent with two distinct, early cattle introductions, Javan and Sumatran breeds also have distinct signatures of introgression from banteng-like (wild *Bos*) populations. We hypothesize that at least part of the introgressed component in the Sumatran breeds Aceh and Pesisir comes from an alternative *Bos* source not represented in this study. Mainland or Bornean banteng, gaur, gayal (*B. frontalis*) and kouprey (*B. sauveli*) could all be candidates, and the latter three have possibly introgressed into zebu in East Asia^{12,23}. The challenges of establishing the species source of wild *Bos* introgression have been highlighted before^{12,33}, and are exacerbated by the paucity of whole-genome data from these endangered species.

There are both theoretical arguments and empirical evidence supporting that negative selection on introgressing haplotypes is the norm^{28,46,63}, even though introgression can also be adaptive^{64–66}. In cattle, introgression is thought to have helped particular breeds adapt to local conditions when introduced to new areas of the world^{5,23,67}. We found several pathways and quantitative trait loci plausibly related to local adaptations, or to the individual agro-ecologies of the different breeds, that are enriched in the regions of high banteng ancestry, including some related to the immune system, growth and development, and fertility. We found only weak evidence of genome-wide linked negative selection acting on banteng haplotypes in the admixed breeds, in contrast to *Homo sapiens/neanderthalensis*^{68,69} and swordtails³⁴. Previous studies suggested that overall linked negative selection on introgression is expected to be correlated with divergence time⁷⁰, but our results suggest that introgression is overall much better tolerated among *Bos* species than among hominins and swordtails, despite *Bos* divergence times being intermediate between these two other examples (~500,000 to 700,000 years ago estimated for banteng-zebu; Supplementary Fig. 28). This is an important finding, because it suggests that genetic variation contained in wild *Bos* may be suitable to genetically improve cattle worldwide. Limited genetic incompatibility among *Bos* could potentially explain the highly reticulated evolution of the *Bos* genus²³, although more studies are required to test this hypothesis. Moreover, it is possible that introgression into a domesticated species can follow distinct selection trajectories compared to non-domesticated species due to the fundamentally different pressures imposed by artificial as opposed to natural selection⁷¹, an intriguing possibility that warrants further investigation.

Correlated genomic distributions of banteng ancestry across breeds with presumed distinct admixture histories, e.g., Madura and East Asian zebu, suggests the presence of convergent selection, consistent with parallelism found in replicate populations of hybrid

swordtail fishes²⁷. Regions almost devoid of banteng ancestry in all breeds, including on large parts of the X chromosome, could contain Bateson-Dobzhansky-Muller incompatibilities, resulting in consistent negative selection on introgressing haplotypes²⁸, or regions where banteng haplotypes have convergently been removed by artificial selection. Incompatibility in specific regions of the genome is not inconsistent with a lack of genome-wide linked selection against the minor parent ancestry. We found lower cross-breed correlations of banteng ancestry in the candidate regions for positive selection than genome-wide (Fig. 5c; Supplementary Fig. 29), suggesting that positive selection on introgression is less convergent than negative selection. This pattern is expected if negative selection is caused by ubiquitous species-wide features, such as genetic incompatibilities or genome architecture, whereas adaptive introgression is generally idiosyncratic in each breed, consistent with their diverse morphology and agroecology²⁰.

Despite predominantly breed-specific adaptive introgression, we did find cases of likely convergent adaptive introgression, such as in the coat-color genes *ASIP*, *KIT*, and possibly *KITLG*, and *TYR*, suggesting that coat color is a trait that has recurrently been under adaptive introgression in cattle receiving gene flow from banteng-like sources. A region around *ASIP* in particular had a strong signal of convergent adaptive introgression, with multiple different haplotypes introgressing in different admixture events. This is consistent with coat color being among the most generic domestication traits⁷², and possibly one of the earliest targets of artificial selection in the domestication process⁷³. If coat color is generally under artificial selection in livestock, introgressing genetic variants affecting this trait would provide a new substrate for such selection⁶⁴. Interestingly, the Indonesian cattle breeds examined here share a tan coat color with banteng, supporting that introgressing coat color genes may have been under positive selection. These results are consistent with previous research identifying *ASIP* and other coat-color genes as candidates of both positive selection on standing variation in zebu cattle⁵ and adaptive introgression in Chinese cattle^{4,5,12}.

We identified many SNP variants in Indonesian cattle that are not found in cattle breeds lacking wild *Bos* introgression, providing a rich and novel source of genetic variation of global importance for managing the genetic resources of cattle. We show that the introgressed regions in high frequency across admixed breeds have important functional associations in cattle, making them potentially relevant for targeted cattle breeding and genetic improvement. As many of the introgressed genetic variants are previously unknown and therefore have not been included in previous cattle genome-wide association studies, they could provide new substrate for production-related selection. In addition, their presence across multiple admixed cattle populations provides a catalog of variants that have already passed through a sustained filter of selection in a zebu-genomic background, proving their genomic compatibility in a cattle background. We anticipate that future studies will be able to test the potential phenotypic associations of some of the regions we have identified here, enhancing the utility of this catalog of introgressed variants for cattle breeding. Such introgressed diversity in tropical cattle breeds may prove vital for future adaptation of other cattle breeds to increasing global temperatures as predicted over the coming decades.

Methods

Sample collection and laboratory protocol

The research presented in this study complies with all relevant ethical regulations and was conducted in accordance with the Code of Conduct for Responsible Research of the University of Copenhagen. We collected 233 samples from 6 Indonesian cattle breeds (Aceh, Pesisir, Pasundan, Jabres, Madura, and Sumba Ongole), 3 Bali cattle populations (Bali, Kupang, Australia), and three individuals of Javan banteng from captivity in Texas, USA (Fig. 1a; Supplementary Data 1;

Supplementary Data 2; Supplementary Fig. 1). The Bali cattle from Australia come from a feral population in Garig Gunak Barlu National Park in northern Australia, descended from 20 individuals that were released from an abandoned British outpost in 1849^{32,56}. Samples consisting of blood were kept in an EDTA buffer in the field, stored at -196°C in dry shipper as soon as possible for transferring from the field to the laboratory in Bogor, and were further transferred to a -80°C freezer for long-term storage. We then followed the manufacturer's protocol instructions of the QIAGEN Blood and cell culture Kit to extract DNA. Before we did the default protocol, we added three treatment steps: (1) adding 500 μl of ice cold water to the blood samples, (2) centrifuge the diluted blood samples for 20 min with the speed of 17,900 $\times g$ in 4°C , (3) discard the supernatant without disturbing the pellet. These extra steps were required to do the default kit protocol because of the humid climate in the Indonesian lab. Before using gel electrophoresis to check the quality of the genomic DNA, we further measured the DNA concentrations with a Qubit 2.0 Fluorometer and a Nanodrop. After DNA extraction, 1 mg genomic DNA was fragmented by Covaris (350 base pairs on average), followed by purification by AxyPrep Mag PCR clean-up kit. The fragments were end-repaired by End Repair Mix and then purified. The repaired DNA was combined with A-Tailing Mix, then the Illumina adaptors were ligated to the DNA adenylate 3' ends, followed by product purification. Size selection was performed targeting insert sizes of 350 base pairs (bp). Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the adaptor-ligated DNA fragments. After purification, the size and quality of libraries was assessed by the Agilent Technologies 2100 Bioanalyzer and ABI StepOnePlus Realtime PCR System.

Additionally, we downloaded 81 publicly available, whole-genome sequencing datasets: 8 samples from Bali cattle from an unknown locality in Indonesia, 42 individuals of *Bos indicus* spreading from East Asia, South Asia, Latin America, and Africa, 27 individuals of *Bos taurus* from Asia, Middle East, Europe, and Africa, and two gaur (*Bos gaurus*), two Javan banteng (*Bos javanicus*) from zoological gardens (Supplementary Data 2; Supplementary Fig. 1).

Sequencing and mapping

All samples were sequenced using illumina paired-end 2×150 bp reads. This includes 230 samples sequenced to depth of 9.62X–17.0X coverage on Illumina NovaSeq platform and 3 samples from captive Javan banteng sequenced to depth of 15.9X–34.9X on the Illumina HiSeq2500 platform (Illumina Inc., San Diego, CA, USA). We assessed the quality of the raw reads using FastQC (bioinformatics.babraham.ac.uk/projects/fastqc) and MultiQC⁷⁴ before mapping.

For mapping, we used a modified version of PALEOMIX BAM pipeline⁷⁵ (github.com/xiqtcacf/IndonesianCattle-Scripts), which is a pipeline designed for the processing of demultiplexed, high-throughput, short-read sequencing data. We first trimmed Illumina universal adapters using AdapterRemoval v2.3.2⁷⁶. We merged read pairs with overlapping sequences of at least 11 bp to improve the fidelity of the overlapping region by selecting the highest-quality base when mismatches are observed. Mismatching positions in the alignment, where both read bases had the same quality, were set to 'N' via the '--collapse-conservatively' option. We did not trim Ns or low-quality bases and only empty reads resulting from primer-dimers were discarded. We then mapped all trimmed reads using BWA-mem v0.7.17-r118870⁷⁷ to two chromosome-level reference genomes: (1) *BosTau9* (GenBank: GCA_002263795.2, ARS-UCD1.2), a female taurine from Hereford breed, and (2) *Waterbuffalo* (GenBank: GCA_003121395.1, UOA_WB_1), a female water buffalo from the Mediterranean breed. PCR duplicates were flagged using samtools v1.11 'markup' for paired reads and PALEOMIX 'rmdup_collapsed' for merged reads.

We merged the resulting BAM alignments from collapsed and paired reads for each individual, and filtered them based on standard

BAM flags to exclude unmapped reads, reads with unmapped mate reads, secondary alignments, reads that failed QC, PCR duplicates, and supplementary alignments. We further excluded reads in alignments with inferred insert sizes <50 bp or >1000 bp, reads where <50 bp or <50% of the reads were aligned, and read pairs in which mates mapped to different contigs or not in the expected orientation. We finally generated statistics of the filtered BAM files by samtools 'stats' and 'idxstats'⁷⁸.

Sample filtering

Heterozygosity. We excluded samples with extraordinarily high heterozygosity, because these samples likely suffer from DNA contamination or considerable sequencing errors. We calculated heterozygosity per individual based on site frequency spectrum (SFS) using genotype likelihood with the GATK model in ANGSD. The analysis revealed six individuals with excessively high heterozygosity (≥ 0.00620 ; Supplementary Data 2) and excluded five out of six for downstream analyses. We kept the sample with highest heterozygosity from Kupang (N_31B) as preliminary analyses suggested it might be a potential F1 hybrid.

Relatedness filtering. We removed duplicates and closely related samples using the methodology described in Waples et al.⁷⁹ We first computed the two-dimensional site-frequency spectrum (2d-SFS) for each pair of samples and then calculated three statistics from the 2d-SFS: RO, RI, and the KING-robust kinship coefficient. We found 54 duplicated pairs with KING-robust kinship > 0.460, of which most were from the Jabres breed, and 27 pairs of up to approximately second-degree relatives (KING-robust > 0.150). We excluded all but one sample with lower coverage from identified duplicated and related pairs, leading to 81 samples discarded (Supplementary Data 2).

Site filtering

Reference genome filtering. We implemented reference genome filtering based on different criteria. We used GenMap v1.2.0⁸⁰ to calculate the mappability score of each site of both the *BosTau9* and the *Waterbuffalo* reference genomes, conservatively using 100 bp k-mers with up to two mismatches allowed (-K 100 -E 2), and default remaining settings. We removed all sites with a mappability score <1 for downstream analyses. We used RepeatMasker v.4.1.1 (repeatmasker.org) to identify repeat regions in both reference genomes, using 'rmbblast' as the search engine and 'mammal' as the query species with default settings. We also excluded repeat regions identified by RepeatMasker, annotated sex chromosomes and scaffolds that were not assembled into chromosomes (Supplementary Data 3). Additionally, we inferred the sample sex using SATC⁸¹, based on the normalized sequencing depth on sex-linked scaffolds for each sample (Supplementary Data 2).

Global depth filtering. For each of the two mapping datasets, we estimated the global depth (read count) per site across all samples using the ANGSD command '-minMapQ 25 -minQ 30 -doCounts 1 -doDepth 1 -dumpCounts 1 -maxdepth 4000' and then estimated the per-site median depth. We excluded sites with a global depth <0.5 times the median ($0.5 \times 1717 = 858.5$) and >1.5 times the median ($1.5 \times 1717 = 2575.5$) from all analyses (Supplementary Data 3).

Excess heterozygosity filtering. We removed regions with excessive heterozygosity, which is likely caused by problematic mapping due to repetitive or paralogous regions. We first generated a preliminary file of genotype likelihoods using ANGSD with the GATK model (-GL 2) from common polymorphic sites ($MAF \geq 0.05$ and $SNP p < 0.000001$), base quality at least 25 (-minQ 25), and minimum mapping quality of 30 (-minMapQ 30). Using these genotype likelihoods as input to PCAngsd v0.985^{82,83}, we then calculated the per-site inbreeding coefficients (F), ranging from -1 where all samples are heterozygous to 1 where all

samples are homozygous, and performed a Hardy-Weinberg equilibrium likelihood ratio test accounting for population structure. The optimal number of principal components to model the population structure was inferred based on Velicer's minimum average partial test⁸⁴ implemented in PCAngsd. Finally, we removed windows of 10 Kb around sites with significant excessive heterozygosity estimates ($F < -0.95$ and $p < 0.000001$) based on the per-site inbreeding coefficients for both *BosTau9* and *Waterbuffalo* reference (Supplementary Data 3).

Genotype calling and imputation

We performed genotype calling for both datasets mapped to *BosTau9* and *Waterbuffalo* reference genomes using bcftools v1.14⁸⁵. Genotype calling was only performed on the samples maintained after sample filtering and only on genomic regions retained after site filtering. We used the 'bcftools pileup' function based on reads with a minimum base quality of 25 and a minimum mapping quality of 30, enabling '-per-sample-mF' to increase calling sensitivity. We then did genotype calling by using '--multiallelic-caller'. Finally, we removed both multi-allelic sites and indels, and applied additional filtering using the setGT plugin of bcftools, imposing a minimum depth of coverage per site of 10 and only accepting heterozygous calls with at least 3 reads supporting each allele.

We did genotype imputation and phasing to remedy genotype missingness and refine the genotypes, because some samples had low depth for regular genotype calling. To prepare the input, we extracted bi-allelic SNPs from the genotype data mapped to both references: *BosTau9*, or 'internal' reference and *Waterbuffalo*, or 'external' reference. We did imputation and phasing using BEAGLE v3.3.2⁸⁶ separately for each chromosome. We visualized the distribution of genotype discordance between the original vcf and imputed vcf genotype files (Supplementary Fig. 30a). In order to evaluate the accuracy of imputation, we additionally conducted the analysis by downsampling the highest-depth banteng individual (34.86X, LIB112407_Banteng_85B_Texas) to depths 1X, 5X, 10X, then imputing them, and comparing the imputed genotypes with the high-quality genotype calls for the full data from this sample using an R package 'vcfppR'⁸⁷. The analysis showed a very high concordance between imputed genotype calls and true genotype calls, supporting the accuracy of imputation (Supplementary Fig. 30b).

PCA and admixture analyses

To investigate population structure, we used HaploNet²⁹, which implements a neural network on local clustering of phased data. We trained the HaploNet model using default settings and produced log-likelihoods that can be processed further to PCA and Admixture. We used ten eigenvectors to capture population structure between and within the filtered dataset of 231 individuals (without the two gaurs) mapped to *BosTau9*. For the admixture analysis, we set the number of ancestry (K) from 3 to 12, with 50 independent runs for each K . We used a convergence criterion of reaching within 5 log-likelihood units of the lowest log-likelihood in at least 3 independent replicates. We obtained convergence with K from 3 to 7. Based on the HaploNet results at $K = 7$ we then defined a subset of individuals as non-admixed representatives of each breed or population by removing samples that contained >10% of admixture from a different ancestry source than the population majority.

Genetic diversity, runs of homozygosity, and population differentiation

Heterozygosity. We assessed genetic diversity of cattle based on genome-wide individual heterozygosity, using filtered genotype-called data that included non-variable sites with a range of depth from 6 to two times of average depth of each sample, a minimum allelic support of 3, a minimum mapping quality of 30, and a base quality of 25. We

calculated the individual heterozygosity as the number of heterozygous genotypes relative to the number of total sites.

Runs of homozygosity. We first explored various different approaches and data filtering options to optimize the detection of runs of homozygosity (ROHs) in each individual. For assessment and validation, we visualized ROHs across the genome similar to the approach in Liu et al.³⁸ and checked visually the performance of ROH calling under different filtering and PLINK detection settings. The assessment criteria include looking for signs of long apparent ROHs being broken up, and the ability of smaller regions of consistently reduced heterozygosity to be called as ROHs in the analysis. Based on these analyses, we used the imputed datasets mapped to *BosTau9* to estimate ROHs using PLINK v1.90b6.24. In PLINK we applied maximum heterozygous calls of 5 (-homozyg-window-het 5), a minimum of 500 kilobases for a ROH (-homozyg-kb 500), and filtered out all variants with missing calls (-geno 0). We merged distinct ROHs within 100 Kb distance, and categorized all resulting ROHs into five length groups: 0.5–1 Mb, 1–2 Mb, 2–5 Mb, 5–10 Mb, and >10 Mb.

Global F_{ST} . To infer population differentiation, we calculated genome-wide global F_{ST} for each pair of populations (after removing admixed individuals) using ‘-fst’ implemented in PLINK2.0⁸⁸. This calculates Hudson’s F_{ST} estimator, which is robust to differences in sample size between populations⁸⁹. We further built a NeighborNet tree based on the pairwise global F_{ST} distances in order to show the relationships of the populations.

Population history and introgression analysis

Treemix. To investigate the history of population splits and historical admixture events, we performed a TreeMix analysis⁹⁰ using the imputed data mapped to the *Waterbuffalo* reference. Because relatedness can underestimate covariance and lead to spurious inferences of migration using Treemix, we did a more stringent sample filtering for the input datasets by applying a threshold of KING kinship coefficient of 0.1 to exclude potentially second degree of related samples. In the TreeMix analysis we included the Indonesian cattle breeds, Bali cattle, captive banteng, East Asian zebu, East Asian taurine and European taurine, two gaurs, and the *Waterbuffalo* to root the tree (Supplementary Data 2). We ran TreeMix assuming 0–10 migration events. For each number of migration events (m), we ran 100 iterations using bootstrap (-bootstrap) and a block size of 1000 SNPs (-k 1000). We inferred the final, optimal number of migration edges (m) from the second-order rate of change in likelihood (Δm) weighted by the standard deviation using the ‘Evanno’ method implemented in OptM⁹¹.

D-statistics. To infer the evolutionary relationships and ancient admixture events, we calculated D-statistics (Patterson’s *D*, also called ABBA-BABA) using the R package ADMIXTOOLS2⁹². We used datasets mapped to the *Waterbuffalo* reference to mitigate the effects of reference bias. We ran D-statistics analyses of the type (H1-H2-H3-H4) with both topology of captive banteng-cattle-South Asian zebu-water buffalo and cattle-South Asian zebu-captive banteng-water buffalo, using the function ‘qpdstat’ implemented in ADMIXTOOLS2.

mtDNA, Y-chromosome DNA analyses, and phylogenetic tree. To infer maternal and paternal phylogenetic relationship between cattle, we inferred the phylogenetic tree for both mitochondrial DNA (mtDNA) and Y-chromosome (Ychr). To identify the matriline in the cattle, we generated consensus mitochondrial sequences. Briefly, we used the -doFasta 1 option of ANGSD⁹³ with quality values specified as ‘-minMapQ 30 -minQ 30 -setMinDepth 5 -uniqueOnly 1 -remove_bads 1’ and also ‘-doCounts 1’ to generate consensus fasta from whole-genome sequencing data mapped to *BosTau9* mitochondrial scaffold (NC_006853). We aligned these fasta sequences using the FFT-NS-1

(fast) option of MAFFT⁹⁴. We imported the aligned sequences into Jalview alignment editor⁹⁵ and removed the regions in the alignment with high ‘N’, and exported the edited sequence in fasta format. We then aligned these edited sequences using the G-INS-i (accurate) option of MAFFT and wrote the output in fasta format. For creating a haplotype network, we converted the fasta files to nexus format and imported to POPART⁹⁶ to create a minimum spanning haplotype network with an epsilon 0. For creating a neighbor-joining tree, we imported the aligned sequences to TreeViewer⁹⁷ and used a Hamming distance model with 5000 bootstrap replicates.

We built the phylogenetic tree for the Y chromosome (Ychr) using BEAST v1.10.4⁹⁸. We first generated the consensus Ychr (only in males) sequences using ANGSD from whole-genome sequencing data mapped to taurine Ychr (GenBank: CM001061.2), with the same settings as for mtDNA consensus sequences. We removed heteroplasmic sites by masking (as ‘N’) any Y chromosome site where <95% of reads carried the same base. After consensus calling, we then performed phylogenetic analyses using the GTR + G + I substitution model and a coalescent Extended Bayesian Skyline Plot prior to avoid restricting the tree by imposing a confining demographic prior. We then ran the Markov chain-Monte Carlo (MCMC) chain for 10^7 steps, sampling trees and parameters every 1000 steps. We assessed convergence and proper mixing by visual inspection and by estimating parameter effective sample sizes using TRACER⁹⁹. We used TreeAnnotator¹⁰⁰ to make a maximum clade credibility tree, discarding the first 1000 trees. We used Figtree v1.4.4 (tree.bio.ed.ac.uk/software/figtree) to visualize the maximum clade credibility tree.

Divergence time. We used MSMC2¹⁰¹ to estimate the divergence time between zebu (*Bos indicus*), banteng (*B. javanicus*) and Bali cattle. We used the phased callable regions of two individuals per population, randomly sampling 10 million SNPs from the genome. We scaled the results for visualization by assuming a generation time of 5–7 years^{4,43,44}, and a mutation rate of 1.26×10^{-8} generation⁴.

Local ancestry inference in admixed populations

LOTER. After having identified cattle populations with signs of ancestral admixture, we did local ancestry inference using LOTER³⁹. LOTER has been used for a wide range of species such as humans^{102,103}, primates^{104,105}, cattle⁷, and rapeseed¹⁰⁶, and does not require prior knowledge such as recombination maps to be implemented. We used imputed and phased datasets with a total of 22,158,517 SNPs mapped to *BosTau9* as input for LOTER. We used unadmixed zebu and banteng individuals as two ancestry source references: (1) the zebu reference population consisted of 14 Sumba Ongole and 11 South Asian zebu, and (2) the banteng reference population consisted of 19 Bali cattle from Bali, 12 Bali cattle from Australia, and five captive banteng. In the reference sets we considered that a low probability of being admixed was more important than having a larger reference set of individuals with more uncertain admixture profiles. We used a total of 90 individuals as target admixed individuals from the following populations with introgression from banteng or another *Bos* source: Aceh, Pesisir, Pasundan, Jabres, Madura, and East Asian zebu. In addition, we included Bali cattle from Kupang, because this population showed signs of introgression from cattle. We performed all analyses using the ‘lc.loster_smooth’ function, which enables a phase-correction module. We estimated the overall proportion of banteng ancestry in each individual by calculating the number of SNPs inferred to be derived from banteng and dividing by the total number of SNP sites. Afterwards, we used non-overlapping sliding windows of 50 Kb to consolidate the raw output from LOTER across individuals in each population. For each admixed population, we calculated the proportion of banteng ancestry in each 50 Kb window by calculating the proportion of SNPs inferred to be of banteng ancestry in each individual (two haplotypes per individual), and taking the mean of this value across all haplotypes in the population.

As local ancestry inference can potentially be affected by the choice of reference genome we assessed whether mapping to a banteng reference would impact the LOTER analyses. We downloaded a recently available banteng reference genome (RefSeq: [GCF_032452875.1](https://doi.org/10.1038/s41467-025-62692-z), ARS-OSU_banteng_1.0) and performed mapping of all the raw data as described above to this reference genome. We then redid all steps described above in the “Site filtering” and “Genotype calling and imputation” sections separately for this alternative mapping, and performed a LOTER analysis as described above. Finally, we calculated the genome-wide proportion of inferred banteng ancestry based on this alternative mapping and plotted the count of SNPs inferred to be of banteng ancestry in each 50 Kb genomic window for each of the two mapped data sets, for two example individuals from Madura (N_911 and N_935). Comparability between the two mapped data sets was ensured by exploiting a liftover of genomic coordinates between the banteng and cattle reference genome, available on NCBI. The analyses found almost identical genome-wide banteng proportions using either mapped data set (Madura population as example shown in Supplementary Fig. 18a), and that the correlation between banteng-inferred SNPs across individual windows was also very high (Supplementary Fig. 18b). We therefore conclude that our findings are robust to the choice of reference genome.

F4 ratios. To estimate the ancestry proportions in each admixed individual, we calculated F4 admixture ratios using ‘qpadm’ implemented in ADMIXTOOLS²². This models a target population as a mixture of two source populations given a set of outgroup populations⁴⁰. We used the same ancestry source references and target admixed populations as in the LOTER analysis. We then estimated F4 ratios in the form of $\alpha = f4$ (taurine, water buffalo; banteng source, target) / $f4$ (taurine, water buffalo; banteng source, zebu source). We used 5×10^6 as the SNP block size for jackknifing.

Hmmix. Additionally, we detected the segments of individual genomes of archaic introgression on Indonesian cattle (Aceh, Pesisir, Pasundan, Jabres, Madura) and East Asian zebu population, with South Asian zebu and Sumba Ongole as outgroups using Hmix v0.6.9⁴¹. This approach is based on a hidden Markov model that identifies genomic regions with a high density of single nucleotide variants not seen in outgroup populations (non-admixed); therefore, it can be used without relying on ancestry reference sources. The rationale behind this approach is to identify regions of high SNP density after removing variation found in outgroup populations, because introgressed regions with higher SNP density have spent more time accumulating variation that is not found in the outgroup compared to non-introgressed regions. We first prepared the input files for this method from imputed datasets mapped to *BosTau9* using the sites retained after site filtering as weights, local mutation rates, and individual observation files using scripts provided with the repository for Hmix (github.com/LauritsSkov/Introgression-detection). We then applied the method to Indonesian cattle and East Asian zebu populations using the following different prior parameters as model training to detect the best-fitting hidden Markov model parameters: Aceh (starting_probabilities = 0.93, 0.07, transitions = 0.98, 0.02 and 0.25, 0.75, emissions = 2, 25); Pesisir (starting_probabilities = 0.88, 0.12, transitions = 0.99, 0.01 and 0.09, 0.91, emissions = 2, 25); Pasundan (starting_probabilities = 0.80, 0.20, transitions = 0.99, 0.01 and 0.05, 0.95, emissions = 2, 25); Jabres (starting_probabilities = 0.76, 0.24, transitions = 0.99, 0.01 and 0.04, 0.96, emissions = 2, 25); Madura (starting_probabilities = 0.63, 0.37, transitions = 0.98, 0.02 and 0.04, 0.94, emissions = 2, 25); East Asian zebu (starting_probabilities = 0.84, 0.16, transitions = 0.93, 0.07 and 0.36, 0.64, emissions = 2, 25). Subsequently, we decoded the data with the best hidden Markov model parameters that maximized the likelihood and identified the archaic introgressed segments. We annotated the archaic introgressed regions by potential source populations (or

species) by calculating the ratio of inferred archaic SNPs in each archaic fragment that was shared with each of two possible source populations: 1) a banteng population consisting of five Javan banteng, 19 Bali cattle from Bali, and 12 Bali cattle from Australia, and 2) two gaur individuals. Moreover, we calculated the identity-by-state matrix for all pairs of admixed cattle individuals based on their overlapping archaic regions. We did all analyses using a 10 Kb window and retaining only the archaic regions with probability >0.9.

U_x and related analyses. We also explored a metric proposed by Racimo et al.¹⁰⁷ to identify putatively adaptively introgressed regions. The metric tabulates the sites that are nearly fixed for different alleles in cattle and banteng (banteng-specific alleles), and where the banteng-specific allele occurs at a high frequency in the admixed population. Known as the $U_{A,B,C}(w,x,y)$ ¹⁰⁷ statistic, where w = the maximum allele frequency in unadmixed cattle, x = minimum allele frequency in the target admixed population, and y = the minimum allele frequency in banteng. We calculated both $U_{A,B,C}(0.05,0.50,0.95)$ and $U_{A,B,C}(0.05,0.20,0.95)$ in non-overlapping, 50 Kb windows in each admixed population. We used the same reference populations to represent unadmixed zebu and banteng as in the LOTER analysis. For practicality, we refer to these statistics as *Uabc50* and *Uabc20*, respectively. However, it is challenging to decide on the value of minimum allele frequency x in population B that gives the best discriminative power for adaptive introgression¹⁰⁷, and we found that this problem is exacerbated when the original admixture proportions as well as sample sizes differ among populations. In addition, this count statistic is influenced by the genome-wide variation in absolute sequence divergence between populations A and C , leading to a potential decoupling of the $U_{A,B,C}$ count from the proportion of local ancestry. Therefore, we also calculated another statistic (U_x) by calculating x in $U_{A,B,C}(0.05,x,0.95)$ for each admixed population in 50 Kb bins. In other words, we calculated the mean allele frequency of population B in sites that had <0.05 derived allele frequency in unadmixed cattle, and >0.95 across Bali cattle and banteng. This mean allele frequency in banteng-diagnostic sites U_x has a more continuous distribution than *Uabc20* and *Uabc50*, and had a higher correlation with local LOTER and Hmix ancestry proportions than any of the $U_{A,B,C}$ we examined, but still constitutes an independent approach for inferring regions of high banteng ancestry in admixed cattle.

Correlation between banteng ancestry and genomic features: recombination rates, coding region density, and conservation score

To investigate correlations between genomic features and banteng ancestry in admixed breeds, we compared the estimated proportion of banteng ancestry in each 50 Kb window with the mean recombination rate, the mean coding region density, and the number of conserved sites in the same window for the Madura population. For recombination rates, we obtained a sex-specific cattle recombination map¹⁰⁸ and did linear interpolation for each SNP by using the ‘approx’ function in R. We then aggregated the recombination map to the 50 Kb window for both sexes. We also obtained information regarding the number of sites in coding regions and the number of conserved sites (phast-Cons30way) from the Ruminant Genome Database (RGDv2¹⁰⁹) binned to the same window size. The latter was based on the *BosTau9* version of the *Bos taurus* reference genome collated with the Y chromosome of the *Btau5.0.1* version (ARS-UCD1.2_Btau5.0.1). For each comparison, we split the windows into 10 quantiles and then calculated the mean proportion of banteng ancestry for each quantile along with its standard error as a measure of uncertainty. We plotted the distributions across quantiles as both scatter plots overlaid with the means, as well as kernel density plots. We also applied a genome-wide Spearman rank correlation test of banteng ancestry proportions among each pair of admixed breeds across all windows and those in the top 5%.

Ancestry-specific population structure in admixed populations

To investigate each of the ancestry sources in admixed populations, we inferred population structure for zebu-specific and banteng-specific ancestry regions, using EMU, a method designed to be robust to both random and non-random missingness¹¹⁰. We used imputed datasets mapped to *BosTau9* and performed a stringent-sample filtering by removing one of each pair of $K1 > 0.2$ identified by NgsRelate^{111,112} within each population (Supplementary Fig. 31). For inferring zebu ancestry structure, we first extracted SNPs inferred to be of zebu ancestry by LOTER per individual, and regions identified as ‘cattle’ by Hmmix per individual. We then merged all admixed individuals with LOTER-zebu and with Hmmix-zebu ancestry with the unadmixed zebu population (Sumba Ongole and South Asian Zebu) separately as inputs for EMU analyses. To infer population structure with EMU, we applied seven eigenvectors ($-n_eig\ 7$), maximum iterations of 1000 ($-iter\ 1000$), and a threshold for minor allele frequencies of 0.05 ($-f\ 0.05$). We estimated PCA for banteng ancestry structure in a similar way to inferring zebu ancestry by extracting SNP positions with banteng ancestry using LOTER and regions with ratio > 0.8 (number of banteng population SNPs \div derived number of SNPs) of Hmmix per admixed sample, respectively. EMU-PCA was then performed on this set of banteng annotated ancestry regions in admixed individuals and the whole genome of unadmixed banteng populations (Bali cattle from Bali, Bali cattle from Australia, and captive banteng) using the same settings as above, except for eight eigenvectors ($-n_eig\ 8$).

As some additional whole-genome sequenced samples from Asian zebu became available during the preparation of the study that could potentially be of relevance to place the masked zebu ancestry of the Indonesian cattle into a geographical context, we downloaded data from Chen et al.¹¹³ and Chen et al.⁵ (Supplementary Data 7) and performed mapping, imputation and LOTER analysis as described for our original data set. We then included them in an additional EMU-PCA. These samples were placed on a cline towards the East Asian zebu samples in the zebu-specific EMU-PCA, consistent with their geographical placement in northernmost Southeast Asia along the shortest route between India and China (Supplementary Fig. 11b). While supporting the credibility of the EMU-PCA as a method to detect historical dispersals, this result does not enable us to further disentangle the origin of zebu cattle introduced to Indonesia. Only samples from further south in mainland Southeast Asia would have enabled this.

To check the performance of masking based on LOTER results, we performed a check by recalculating D-statistics after masking all SNPs inferred to be introgressed in each sample. This plot demonstrates the ability of the LOTER based masking to remove all or by far most of the introgression from banteng-like *Bos* sources (Supplementary Fig. 11a).

Estimation of admixture time

To infer the timing of admixture events, we traced the ancestry of discrete genomic segments for all of Indonesian cattle populations and East Asian zebu using Ancestry_HMM, a hidden Markov model-based method^{42,45}. We fitted a single-pulse admixture model to the genome-wide variation data and used the mean ancestry proportions estimated by LOTER as the assumed admixture proportion ($-p\ 1\ 100000\ -0.6\ -p\ 0\ -500\ -0.4$). We also tried a two-pulse admixture model for the Javan breeds with the settings: ‘ $-p\ 1\ 100000\ -0.6\ -p\ 0\ -1000\ -0.2\ -p\ 0\ -200\ -0.2$ ’. However, the higher admixture time and lower admixture ($\sim 0.01\%$) estimated for the first pulse when using the two-pulse model suggests that the two-pulse model is a poor fit to the data⁴⁵ (Supplementary Fig. 32). We quantified uncertainties by doing 100 bootstrap replicates for each population using a block size of 5000 SNPs ($-b\ 100\ 5000$).

Overlapping introgressed segments among LOTER, Hmmix, and U_X

To obtain the top 5% windows of highest-inferred banteng ancestry in each of the five cattle groups (Aceh, Pesisir, Pasundan, Madura, and

East Asian Zebu), we merged the results from LOTER, Hmmix, and U_X using non-overlapping 50 Kb windows made according to the taurine autosomal chromosomes (*BosTau9*) and annotated with genes coming from *BosTau9* reference-genome annotation (GFF, Ensembl version 106) using bedtools intersect¹¹⁴. For each of the 50 Kb windows, we counted the proportion of banteng SNPs coming from LOTER (anc1), number of SNPs in U_X , and the mean proportion of archaic regions with probability ≥ 0.9 from Hmmix. For each statistic, we determined the top 5% quantile from each cattle group to filter out windows that do not contain the highest banteng ancestry. Due to the high number of windows with *Uabc50* and *Uabc20* = 0, the top 5% quantile from U_X statistics are 0 and hence not used for this filtering step. Before obtaining the top 5% windows of highest banteng ancestry, we first fitted a linear model to explore the predictive power of Hmmix for LOTER and found that Hmmix can explain more than 50% of the variation in LOTER values (Supplementary Fig. 23). However, consistently higher proportion of introgressed ancestry inferred by Hmmix than by LOTER and U_X , could potentially be due to either Hmmix finding introgression from a wider array of sources than LOTER (e.g. other bovines present in SEA), or to a higher tendency of false positives in Hmmix, or a mixture of the two. We thus only kept windows that passed the top 5% quantile of LOTER values in each cattle group. We then listed these regions for each cattle group and annotated them with a gene list of *BosTau9* from NCBI *Bos taurus* Annotation Release 106 (2019-12-18, GCF_002263795.1_ARS-UCD1.2_genomic.gff.gz¹¹⁵) and QTL information of *Bos taurus* ARS-UCD1.2 from QTLdb release 53¹¹⁶ to identify signs of adaptive introgression. We used values from Hmmix and U_X for further validation as an overall sanity check, or robustness analysis.

Gene enrichment analysis

To characterize the functional associations of these candidate regions for adaptive introgression, we performed GO-enrichment analyses for the outlier gene set from each breed. For each unique gene in the top 5% list, as well as genes in the zero-banteng regions, we performed an overrepresentation analysis for gene-enrichment by using the g:GOST feature of the web-based g:Profiler¹¹⁷. Genes overlapping the top 5% banteng ancestry regions of each cattle population (Aceh, Pesisir, Pasundan, Madura, and East Asian zebu) were used as input to g:GOST v110 using *Bos taurus* as the annotation set and a g:SCS significance threshold of 0.05. For each cattle population we collected the outcome of the gene enrichment based on molecular function, cellular component, biological process, as well as terms from KEGG, REAC, and HP if any (Supplementary Data 8). We then made an UpSet plot for genes in the top 5% using the ComplexUpset package¹¹⁸ in R v4.3.3¹¹⁹ to see how many outlier genes were shared across multiple cattle populations. To assess any overall overrepresentation or underrepresentation of genes contained in the top 5% banteng ancestry regions in each breed, we performed a simulation by randomly choosing 1000 times 5% of the genomic windows (2482 windows) out of the total of 49,624 genome-wide windows. We then counted the number of genes for each iteration and plotted the results in a histogram.

Haplotype structure

We visualized haplotype structure of the *ASIP* gene on chromosome 13 between 63.64 Mb and 63.67 Mb using Haplostrips¹²⁰. The software extracts the haplotype data from the phased genotypes, keeps the samples belonging to populations of interest and chooses only the most informative sites by eliminating variations with very low frequency in all the populations, and finally produces a plot that displays the haplotypes in rows while each column represents a SNP within a region of interest. Populations of captive banteng and Bali cattle were lumped as banteng and were treated as the reference population. We also built the haplotype network for *ASIP* gene using the same methods described above for inferring mtDNA phylogenetic tree.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequence data generated in this study have been deposited in the NCBI SRA database under accession code [PRJNA1108075](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002263795.1/). Chromosome-level assemblies of a taurine (*B. taurus*), a water buffalo (*B. bubalis*), and a banteng (*B. javanicus*) which are available on the NCBI database with GenBank code GCA_002263795.2 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002263795.1/], GenBank code GCA_003121395.1 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003121395.1/], and RefSeq code GCF_032452875.1 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_032452875.1/] respectively, were also used in this study. Additional Y chromosome of a taurine cattle is available in GenBank CM001061.2 [<https://www.ncbi.nlm.nih.gov/nucleotide/CM001061.2/>] was utilized for Y chromosome analyses. The data details and NCBI accession code for other downloaded data used in this study are in Supplementary Data 2 and Supplementary Data 7. Source data are provided as a Source Data file. Source data are provided with this paper.

Code availability

Scripts used to generate all analyses and plots are available at github.com/xiqtcacf/IndonesianCattle-Scripts.

References

- Strandén, I., Kantanen, J., Lidauer, M. H., Mehtiö, T. & Negussie, E. Animal board invited review: Genomic-based improvement of cattle in response to climate change. *Animal* **16**, 100673 (2022).
- Taberlet, P., Coissac, E., Pansu, J. & Pompanon, F. Conservation genetics of cattle, sheep, and goats. *C. R. Biol.* **334**, 247–254 (2011).
- The State of the World's Biodiversity for Food and Agriculture: FAO COMMISSION ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE ASSESSMENTS • *Food and Agriculture Organization of the United Nations Food & Agriculture Org* (2019).
- Chen, N. et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).
- Chen, N. et al. Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. *Nat. Commun.* **14**, 7803 (2023).
- Gebrehiwot, N. Z., Strucken, E. M., Aliloo, H., Marshall, K. & Gibson, J. P. The patterns of admixture, divergence, and ancestry of African cattle populations determined from genome-wide SNP data. *BMC Genomics* **21**, 869 (2020).
- Kim, K. et al. The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nat. Genet.* **52**, 1099–1110 (2020).
- Mei, C. et al. Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome Resequencing. *Mol. Biol. Evol.* **35**, 688–699 (2018).
- Jiang, L. et al. Whole-genome sequencing of endangered Zhoushan cattle suggests its origin and the association of MC1R with black coat colour. *Sci. Rep.* **11**, 17359 (2021).
- Sun, L. et al. Assessing genomic diversity and selective pressures in Bashan cattle by whole-genome sequencing data. *Anim. Biotechnol.* **34**, 835–846 (2023).
- Chen, S. et al. Zebu cattle are an exclusive legacy of the South Asia neolithic. *Mol. Biol. Evol.* **27**, 1–6 (2010).
- Dai, X. et al. A Chinese indicine pangenome reveals a wealth of novel structural variants introgressed from other Bos species. *Genome Res* **33**, 1284–1298 (2023).
- Manguin, P. Y. The archaeology of early maritime polities of Southeast Asia. *Southeast Asia* (2023).
- Nijman, I. J. et al. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity* **90**, 10–16 (2003).
- Mohamad, K. et al. On the origin of Indonesian cattle. *PLoS One* **4**, e5490 (2009).
- Mohamad, K. et al. The origin of Indonesian cattle and conservation genetics of the Bali cattle breed. *Reprod. Domest. Anim.* **47**, 18–20 (2012).
- Mason, I. L. *Evolution of Domesticated Animals*. (Longman Publishing Group, 1984).
- Felius, M. *Cattle Breeds: An Encyclopedia*. (2007).
- Purwantara, B., Noor, R. R., Andersson, G. & Rodriguez-Martinez, H. Banteng and Bali cattle in Indonesia: status and forecasts. *Reprod. Domest. Anim.* **47**, 2–6 (2012).
- Sutarno, S. & SETYAWAN, A. D. The diversity of local cattle in Indonesia and the efforts to develop superior indigenous cattle breeds. *Biodiversitas Journal of Biological Diversity* **17**, 275–295 (2016).
- McCool, C. Buffalo and Bali cattle-exploiting their reproductive behaviour and physiology. *Trop. Anim. Health Prod.* **24**, 165–172 (1992).
- Martojo, H. Indigenous Bali cattle is most suitable for sustainable small farming in Indonesia. *Reprod. Domest. Anim.* **47**, 10–14 (2012).
- Wu, D.-D. et al. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
- Schumer, M. et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656–660 (2018).
- Coyne, J. A. & Orr, H. A. *Speciation*. (Sinauer Associates Incorporated, 2004).
- Schumer, M. et al. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife* **3**, e02535 (2014).
- Langdon, Q. K. et al. Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLoS Genet* **18**, e1009914 (2022).
- Moran, B. M. et al. The genomic consequences of hybridization. *Elife* **10**, e69016 (2021).
- Meisner, J. & Albrechtsen, A. Haplotype and population structure inference using neural networks in whole-genome sequencing data. *Genome Res* **32**, 1542–1552 (2022).
- Sapi Pasundan, Plasma Nutfah Kebanggan Jawa Barat. *Direktorat Jenderal Peternakan dan Kesehatan Hewan, Kementerian Pertanian* <https://ditjenpkh.pertanian.go.id/berita/88-sapi-pasundan-plasma-nutfah-kebangganjawa-barat>.
- Sulamsi, G. A., Priyanto, R. and Sumantri, C. Karakteristik Sapi Pasundan Berdasarkan Studi Morfometrik Dan Kraniometrik. (Bogor Agricultural University (IPB), 2017).
- Bradshaw, C. J. A., Isagi, Y., Kaneko, S., Bowman, D. M. J. S. & Brook, B. W. Conservation value of non-native banteng in northern Australia. *Conserv. Biol.* **20**, 1306–1311 (2006).
- Sinding, M.-H. S. et al. Kouprey (*Capreolus*) genomes unveil polytomic origin of wild Asian. *iScience* **24**, 103226 (2021).
- Tapio, I. et al. Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data. *Conserv. Biol.* **20**, 1768–1779 (2006).
- Upadhyay, M. R. et al. Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. *Heredity* **118**, 169–176 (2017).

36. Utsunomiya, Y. T. et al. Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. *BMC Genet* **15**, 47 (2014).
37. Upadhyay, M. et al. Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Mol. Ecol.* **30**, 6701–6717 (2021).
38. Liu, X. et al. Introgression and disruption of migration routes have shaped the genetic integrity of wildebeest populations. *Nat. Commun.* **15**, 2921 (2024).
39. Dias-Alves, T., Mairal, J. & Blum, M. G. B. Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. *Mol. Biol. Evol.* **35**, 2318–2326 (2018).
40. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
41. Skov, L. et al. Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet* **14**, e1007641 (2018).
42. Corbett-Detig, R. & Nielsen, R. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genet* **13**, e1006529 (2017).
43. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 803–808 (2002).
44. Gautier, M. et al. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* **177**, 1059–1070 (2007).
45. Medina, P., Thornlow, B., Nielsen, R. & Corbett-Detig, R. Estimating the Timing of Multiple Admixture Pulses During Local Ancestry Inference. *Genetics* **210**, 1089–1107 (2018).
46. Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360 (2016).
47. Gao, Y. et al. Species composition and environmental adaptation of indigenous Chinese cattle. *Sci. Rep.* **7**, 16196 (2017).
48. Mattar, R., de Campos Mazo, D. F. & Carrilho, F. J. Lactose intolerance: diagnosis, genetic, and clinical factors. *Clin. Exp. Gastroenterol.* **5**, 113–121 (2012).
49. Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G. & Swallow, D. M. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124**, 579–591 (2009).
50. Bellinge, R. H. S., Liberles, D. A., Iaschi, S. P. A., O'Brien, P. A. & Tay, G. K. Myostatin and its implications on animal breeding: a review. *Anim. Genet.* **36**, 1–6 (2005).
51. Behl, J. D. et al. The major histocompatibility complex in bovines: a review. *ISRN Vet. Sci.* **2012**, 872710 (2012).
52. Decker, J. E. et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet* **10**, e1004254 (2014).
53. Sudrajat, P. et al. An insight into the evolutionary history of Indonesian cattle assessed by whole genome data analysis. *PLoS One* **15**, e0241038 (2020).
54. Payne, W. J. A. & Rollinson, D. H. L. Madura cattle. *Z. Tierzuecht. Zuechtungsbiol.* **93**, 89–100 (1976).
55. Putra, W. P. B. & Muzawar, M. The inbreeding case of Bali cattle (*Bos javanicus*) at breeding station. *Kocatepe Vet. J.* <https://doi.org/10.30607/kvj.733991> (2020).
56. Bradshaw, C. J. A. et al. Low genetic diversity in the bottlenecked population of endangered non-native banteng in northern Australia. *Mol. Ecol.* **16**, 2998–3008 (2007).
57. Calaby, J. H. Letter: Introduction of Bali cattle to Northern Australia. *Aust. Vet. J.* **51**, 108 (1975).
58. Corbett, L. Does dingo predation or buffalo competition regulate feral pig populations in the Australian wet-dry tropics? An experimental study. *Wildl. Res.* **22**, 65 (1995).
59. Bell, D. A. et al. The Exciting Potential and Remaining Uncertainties of Genetic Rescue. *Trends Ecol. Evol.* **34**, 1070–1079 (2019).
60. Wangkumhang, P. et al. Genetic analysis of Thai cattle reveals a Southeast Asian indicine ancestry. *PeerJ* **3**, e1318 (2015).
61. Hall, K. R. *A History of Early Southeast Asia: Maritime Trade and Societal Development, 100–1500*. (Rowman & Littlefield Publishers, 2010).
62. Hoogervorst, T. G. & Ják, J. Ruminant relations. *J. Am. Orient. Soc.* **144**, (2024).
63. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881–891 (2016).
64. Hedrick, P. W. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618 (2013).
65. Tigano, A. & Friesen, V. L. Genomics of local adaptation with gene flow. *Mol. Ecol.* **25**, 2144–2164 (2016).
66. Suarez-Gonzalez, A., Lexer, C. & Cronk, Q. C. B. Adaptive introgression: a plant perspective. *Biol. Lett.* **14**, 20170688 (2018).
67. Verdugo, M. P. et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176 (2019).
68. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
69. Juric, I., Aeschbacher, S. & Coop, G. The Strength of Selection against Neanderthal Introgression. *PLoS Genet* **12**, e1006340 (2016).
70. Dagilis, A. J. & Matute, D. R. The fitness of an introgressing haplotype changes over the course of divergence and depends on its size and genomic location. *PLoS Biol.* **21**, e3002185 (2023).
71. Gregory, T. R. Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. *Evolution* **2**, 5–27 (2009).
72. Cieslak, M., Reissmann, M., Hofreiter, M. & Ludwig, A. Colours of domestication. *Biol. Rev. Camb. Philos. Soc.* **86**, 885–899 (2011).
73. Ludwig, A. et al. Coat color variation at the beginning of horse domestication. *Science* **324**, 485 (2009).
74. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
75. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
76. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
77. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).
78. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. Waples, R. K., Albrechtsen, A. & Moltke, I. Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Mol. Ecol.* **28**, 35–48 (2019).
80. Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).
81. Nursyifa, C., Brüniche-Olsen, A., Garcia-Erill, G., Heller, R. & Albrechtsen, A. Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Mol. Ecol. Resour.* **22**, 458–467 (2022).
82. Meisner, J. & Albrechtsen, A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731 (2018).
83. Meisner, J. & Albrechtsen, A. Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Molecular Ecology Resources* **19**, 1144–1152 (2019).
84. Velicer, W. F. Determining the number of components from the matrix of partial correlations. *Psychometrika* **41**, 321–327 (1976).

85. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
86. Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
87. Li, Z. vcfpp: a C++ API for rapid processing of the variant call format. *Bioinformatics* **40**, btac049 (2024).
88. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
89. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res* **23**, 1514–1521 (2013).
90. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967 (2012).
91. Fitak, R. R. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* **6**, bpab017 (2021).
92. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
93. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinforma.* **15**, 356 (2014).
94. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
95. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
96. [No title]. https://www.researchgate.net/profile/Tabasum-Akhter/post/Any-advice-on-how-to-analyze-CO1-and-16s-data-for-my-population-genetic-studies/attachment/5c654071cfe4a781a57fc662/AS%3A726151869767685%401550139505725/download/Leigh_et_al-2015-Methods_in_Ecology_and_Evolution.pdf.
97. Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol. Evol.* **14**, e10873 (2024).
98. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
99. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
100. Helfrich, P., Riebel, E., Abrami, G., Lücking, A. & Mehler, A. TreeAnnotator: Versatile visual annotation of hierarchical text relations. in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (2018).
101. Schiffels, S. & Wang, K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
102. Jacobs, G. S. et al. Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell* **177**, 1010–1021.e32 (2019).
103. Cuadros-Espinoza, S., Laval, G., Quintana-Murci, L. & Patin, E. The genomic signatures of natural selection in admixed human populations. *Am. J. Hum. Genet.* **109**, 710–726 (2022).
104. Wu, H. et al. Hybrid origin of a primate, the gray snub-nosed monkey. *Science* **380**, eabl4997 (2023).
105. Zhang, B.-L. et al. Comparative genomics reveals the hybrid origin of a macaque group. *Sci. Adv.* **9**, eadd3580 (2023).
106. Zou, J. et al. Genome-wide selection footprints and deleterious variations in young Asian allotetraploid rapeseed. *Plant Biotechnol. J.* **17**, 1998–2010 (2019).
107. Racimo, F., Marnetto, D. & Huerta-Sánchez, E. Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Mol. Biol. Evol.* **34**, 296–317 (2017).
108. Ma, L. et al. Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. *PLoS Genet* **11**, e1005387 (2015).
109. Fu, W. et al. RGD v2.0: a major update of the ruminant functional and evolutionary genomics database. *Nucleic Acids Res* **50**, D1091–D1099 (2022).
110. Meisner, J., Liu, S., Huang, M. & Albrechtsen, A. Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics* **37**, 1868–1875 (2021).
111. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
112. Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A. & Korneliussen, T. S. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8**, giz034 (2019).
113. Chen, N. et al. Ancient genomes reveal tropical bovid species in the Tibetan Plateau contributed to the prevalence of hunting game until the late Neolithic. *Proc. Natl Acad. Sci. USA* **117**, 28150–28159 (2020).
114. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
115. Rosen, B. D. et al. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, gaaa021 (2020).
116. Hu, Z.-L., Park, C. A. & Reecy, J. M. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res* **50**, D956–D961 (2022).
117. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207–W212 (2023).
118. Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
119. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (*No Title*) (2024).
120. Marnetto, D. & Huerta-Sánchez, E. *Haplostrips*: revealing population structure through haplotype visualization. *Methods Ecol. Evol.* **8**, 1389–1392 (2017).

Acknowledgements

We are grateful to all sample providers from BPTU Indrapuri, Aceh Besar, Aceh; BPPIBTSP Ciamis, Dinas Ketahanan Pangan dan Peternakan Jawa Barat; BPTU-HPT Padang Mangatas, Sumatera Barat; Dinas Peternakan Kabupaten Kupang NTT, East Nusa Tenggara. RH and CN were supported by a Danmarks Frie Forskningsfond Sapere Aude research grant (DFF8049-00098B), and RH, LB and XW were supported by an European Research Council Starting Grant (No. 853442). MHSS was supported by The Carlsberg Foundation (CF20-0355).

Author contributions

Methodology: X.W., C.N., S.G.A., G.G.E., L.D.B., A.K., J.K., K.H., J.M. and T.B. Samples collection and lab work: A.A.C., A.P.P., D.S.P., Y.A.T., P.S., C.D.G., R.C.S., B.D.R., D.E.H., M.P.H., T.P.L.S., M.H.S.S., D.D.S., M.A. and B.P. Writing - original draft: X.W., C.N. and R.H. Writing - review & editing: X.W., C.N., S.G.A., G.G.E., L.D.B., A.K., J.K., K.H., J.M., T.B., C.J.A.B., A.A.C., A.P.P., D.S.P., Y.A.T., P.S., C.D.G., Y.J., J.A.L., R.C.S., B.D.R., D.E.H., M.P.H., T.P.L.S., L.F., G.L., M.H.S.S., D.D.S., M.A., B.P. and R.H. Supervision: R.H. All authors contributed to the writing of the manuscript and approved the paper for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62692-z>.

Correspondence and requests for materials should be addressed to Rasmus Heller.

Peer review information *Nature Communications* thanks Kevin Daly, who co-reviewed with Jolijn Erven; and Olivier Hanotte, who co-reviewed with Nigatu Ayele, for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Biology, University of Copenhagen, Copenhagen N, Denmark. ²Bioinformatics Research Centre, Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. ³National Centre for Biological Sciences (NCBS), Bangalore, India. ⁴Centre for Ecological Sciences, Indian Institute of Science, Bangalore, Karnataka, India. ⁵Biological and Precision Psychiatry, Mental Health Centre Copenhagen, Copenhagen University Hospital, København, Denmark. ⁶Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, København, Denmark. ⁷Global Ecology | Partuyarta Ngadluku Wardli Kuu, College of Science and Engineering, Flinders University, Adelaide, SA, Australia. ⁸Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage, Wollongong, NSW, Australia. ⁹Genetic Conservation and Genome Laboratory, IPB University, Bogor, Indonesia. ¹⁰Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia. ¹¹Department of Animal Science, Faculty of Agriculture, Universitas Tanjungpura, Pontianak, Indonesia. ¹²Research Center for Animal Husbandry, National Research and Innovation Agency, Jakarta, Indonesia. ¹³Faculty of Medicine and Veterinary Medicine, Universitas Nusa Cendana, Kupang, Indonesia. ¹⁴Department of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, Northwest Agriculture and Forestry University, Yangling, China. ¹⁵Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands. ¹⁶Ten Triple X Ranch, Glen Rose, TX, USA. ¹⁷U.S. Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD, USA. ¹⁸Department of Animal and Food Sciences, Oklahoma State University, Stillwater, OK, USA. ¹⁹U.S. Department of Agriculture, Agricultural Research Service, U.S. Meat Animal Research Center, Clay Center, NE, USA. ²⁰Paleogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich, Germany. ²¹School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. ²²Palaeogenomics and Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford, UK. ²³Department of Biology, Faculty of Science, IPB University, Bogor, Indonesia. ²⁴Division of Reproduction and Obstetrics, School of Veterinary Medicine and Biomedical Sciences, IPB University, Bogor, Indonesia. ²⁵These authors contributed equally: Xi Wang, Casia Nursyifa. ✉ e-mail: rheller@bio.ku.dk