

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted <i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data generation and sample collection is described in Methods. A detailed description of data used within this study is provided within Supplementary Data 2 and Supplementary Data 7.

Data analysis

Code used for analyses are publicly available on Github: <https://github.com/xiqtcacf/IndonesianCattle-Scripts>. Software used and their version numbers are described in Methods and provided here: FastQC v0.11.9; MultiQC v1.13; PALEOMIX BAM pipeline (branch 'pub/2022/africa'); AdapterRemoval v2.3.2; BWA-mem v0.7.17-r1188; samtools v1.11; Angsd v0.925; GENMAP v1.2.0; RepeatMasker v4.1.1; PCAngsd v0.985; SATC; bcftools v1.14; BEAGLE v3.3.2; vcfpR v0.8.0; HaploNet v0.5; PLINK v1.90b6.24; PLINK2.0; TreeMix v1.12; OptM v0.1.9; ADMIXTOOLS2 v2.0.0; MAFFT v7; Jalview v2.11.4.0; POPART 1.7.1b; TreeViewer v2.2.0; BEAST v1.10.4; TRACER v1.7.2; TreeAnnotator v1.10; Figtree v1.4.4; MSMC2 v2.1.3; LOTER v1.0; Hmmer v0.6.9; EMU v1.01; NgsRelate v2; Ancestry_HMM v1.0.2; g:GOST v110; ComplexUpset package v1.3.3; R v4.3.3; Haplostrips v1.2.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data used in this study are described within the Article and Supplementary Information. Sample locations and their sources are described within Supplementary Data 2. Raw fastaq files generated within this study and their associated metadata are publicly available on the NCBI database under BioProject accession PRJNA1108075 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1108075/]. The details of other downloaded data used in this study are described in Supplementary Data 2 and Supplementary Data 7.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender n.a.

Reporting on race, ethnicity, or other socially relevant groupings n.a.

Population characteristics n.a.

Recruitment n.a.

Ethics oversight n.a.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-------------------|--|
| Study description | In order to elucidate the genomic resources in Indonesian cattle and how they have been shaped by their complex and poorly resolved history, we generated 233 whole-genomes from Indonesian cattle and banteng, 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. Depending on analyses, samples were treated as individual-level units or population-level units. |
| Research sample | Samples are from the genus Bos, including zebu (Bos indicus), banteng (B. javanicus), taurine (Bos taurus), and gaur (B. gaurus), chosen because the study aim was 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, and are therefore the taxa and species of interest. Samples sequenced in this study include 179 samples from Indonesian cattle breeds Aceh (31), Pesisir (37), Pasundan (26), Jabres (33), Madura (34) and Sumba Ongole (18), 51 Bali cattles from Bali (19), Kupang (17) and Australia (15); and three captive Javan banteng from Texas, USA. We combined them with 81 publicly available genomes from 2 Javan banteng, 8 Bali cattle, 42 zebu, 27 taurine, and 2 gaur, resulting in 314 samples in total for mapping and downstream analyses. Exact location coordinates for all samples and reference for downloaded samples can be found in Supplementary Data 2. In addition, two chromosome-level assembly, taurine cattle (BosTau9, GenBank: GCA_002263795.2-ARS-UCD1.2) and water buffalo (WaterBuffalo, GenBank: GCA_003121395.1-UOA_WB_1) were used as reference in this study and described in the article. The Y chromosome of taurine cattle (GenBank: CM001061.2) were used and described in the article. We additionally downloaded 19 samples from mainland Asia to hope providing additional evidence for the origins of cattle introduction into Indonesia during revisions. As local ancestry inference can potentially be affected by the choice of reference genome, we also downloaded a recently available banteng reference genome (RefSeq: GCF_032452875.1-ARS-OSU_banteng_1.0) to assesse whether mapping to a banteng reference would impact the LOTER analyses during revisions. |
| Sampling strategy | No sample size calculations were performed, but were chosen based on availability and breadth across different locations in Indonesia to answer research questions described in the Introduction. |
| Data collection | Samples were gathered from scientific collections at different institutions and contributed by different co-authors. Sample |

contributions and metadata were recorded by members in labs within and/or were collaborators with researchers 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.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

As described in Results, 81 samples were removed due to suspected duplication (54 samples) and relatedness (27 samples of first and second degree), (Supplementary Data 1 and Supplementary Data 2).

Reproducibility

For Haplonet, we set the number of ancestry (K) from 3–12, with 50 independent runs for each K, and 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 3–7. For TreeMix, we assume 0–10 migration events. For each number of migration events, we ran 100 iterations using bootstrap. We built the phylogenetic tree for the Y chromosome (Ychr) by running the Markov chain-Monte Carlo (MCMC) chain for 10e7 steps, sampling trees and parameters every 1000 steps. For divergence time inference using MSMC2, we used the phased callable regions of two individuals per population, randomly sampling 10 million SNPs from the genome. For estimation of admixture time using Ancestry_HMM, we quantified uncertainties by doing 100 bootstrap replicates for each population using a block size of 5000 SNPs. LOTER was performed by random sampling based on a randomly chosen seed. For other analyses, results were primarily dependent on the underlying data. Assuming that the data are treated and analysed appropriately, results for these analyses are reproducible. Details about sample storage are provided in the Methods section in the manuscript.

Randomization

Samples were initially grouped based on location, and were then classified into populations after investigating population structure between samples.

Blinding

Blinding was not used within this study, but analyses for all samples were treated the same after sample exclusion, described in Methods. Blinding was not performed because this study does not measure or evaluate an exposure.

Did the study involve field work?

☐ Yes

☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

n.a.

Wild animals

The tissue and DNA samples are sourced from pre-existing scientific collections, and therefore no field collection was carried out specifically for the present study.

Reporting on sex

n.a.

Field-collected samples

Details about sample storage are provided in the Methods section in the manuscript.

Ethics oversight

All samples were provided directly by scientists (co-authors). Samples were collected complying with local and international legislation. All sample collection predates the Nagoya protocol. The research was carried out in compliance with the Code of Conduct for Responsible Research of the University of Copenhagen.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

| | |
|-----------------------|--|
| Seed stocks | <i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i> |
| Novel plant genotypes | <i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i> |
| Authentication | <i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i> |