

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

Corresponding Author: Dr Rasmus Heller

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Wang et al report over 200 genomes from East/Southeast Asian Bos species: Indonesian cattle and domesticated banteng (*Bos javanicus*). They find widespread banteng ancestry in certain Indonesian cattle breeds such as Madura and Pesisir, with distinct breed-specific patterns. They disentangle this admixture using local ancestry inferences, revealing distinct admixture patterns in Sumatran compared to Javan cattle; timing of admixture using local ancestry also indicates that these were distinct processes within the last ~2000 years. The authors also estimate heterozygosity of the zebu & banteng sources of admixture, inferring distinct population histories. The authors estimate diversity parameters of various East/Southeast Asian breeds, inferring high heterozygosity in Indonesian cattle, low heterozygosity in Bali cattle/banteng, and notable runs of homozygosity in certain populations of banteng. The authors finally assess the potential selection acting on introgressed banteng variants in Indonesian cattle, assessing functional enrichment within breeds and regions with an excess of banteng ancestry across breeds, finding evidence in selection milk-associated, pigmentation-associated, and immune-associated genes.

This is a well-written, informative paper that coincidentally produces a valuable resource of Bos genomes. The paper advances the current understanding of Bos diversity, ancestry, and population history, using current population genetic techniques and some creative analyses (e.g. leveraging the local ancestry traits to infer the genetic makeup of the unsampled source populations). The inference of human-directed admixture with wild Bos (due to the mtDNA patterns and large introgression despite biological barriers) is also a very interesting finding. To our knowledge these data described here are considerably advancing the genetic resources available for these species and breeds. The paper is building off a literature of (cited) work of Indonesian cattle and Bali cattle; the scale (in terms of genome wide data and individuals) and quality of the analyses. In general the data & analyses supports the conclusions of the paper - this has the very real potential to guide efforts of preserving or utilizing genetic variation existing in Indonesian cattle breeds; as will the signals of low diversity in Bali cattle guide breed maintenance. The general findings of negative selection against introgressed variation, and potential convergent adaptive introgression, will also contribute to ongoing discussions within evolutionary biology of the role of introgression.

The authors provide the raw data and the scripts used in this study on their github. The methods are detailed and in combination with the available scripts the work can be reproduced. However, it would be helpful if the authors could specify the number of SNPs used for each analysis to enhance transparency and provide a clearer understanding of the scope of the dataset. (I cannot seem to find the Haplonet script in the github repository).

The methodology employed in this study is well-structured and robust; the use of multiple methods for local ancestry inference is particularly commendable. However, some points of the analyses are criticised below and should be resolved - these represent our major comments and are constructive. But overall, the methodology is well-justified and meets the expected standards in our field.

Major comments

Line 745-751 - BEAGLE imputation.

How well did the genotype imputation with BEAGLE 3 perform, has genotype imputation been tested on Bali cattle/Banteng and diverse cattle breeds (such as the Indonesian breeds), how well would these potential diverse haplotypes impute? In

line 750-751 a visualization of distribution of genotype discordance between original and imputed VCF files was mentioned, but we cannot find these in the manuscript or SI. This section of the analyses - underlying many downstream parts of the manuscript - does require a little more scrutiny: the validation of genotype imputation of the Bali cattle/Banteng and similar diverse cattle breeds or a more in-depth discussion of the genotype discordance and imputation performance. The highest coverage Bentang individual has an average genome coverage of 34.86X, it would be interesting to see how well these high-coverage genotypes are imputed. Were imputed genotypes filtered after imputation, if yes please add this information.

Line 772-779 - Regions of Homozygosity.

Could you clarify why the threshold of 5 heterozygous SNPs per sliding window was chosen for the ROH analysis? Was this parameter based on previous research? If so, please provide the relevant citation. If not, please justify the selection of this threshold. One approach to do so would be to estimate the potential heterozygosity error rate, which could result from sequencing or alignment errors, and demonstrate how this aligns with the chosen parameter. If potential false-positive heterozygous rate due to imputation might be a problem, please specify and show accordingly. Also, please specify the number of SNPs used for this analysis.

Line 849-872 Local ancestry inference

It is welcoming to see that three different methods were used to infer local ancestry—this approach is very comprehensive. However, there are some concerns regarding the use of the inbred Australian Bali cattle population as one of the ancestry source references. Given the high amount of ROH and low diversity in this population, could this have affected the local ancestry inference? Also considering the variation observed in Figure 4E and S10-11. Would it be possible to rerun the analyses on a subset of Bali cattle, excluding the Australian Bali cattle, to assess the impact? If not, could you clarify whether including this population would not significantly affect the results?

Minor comments

The figures in general should be made colour-blind friendly, even if just removing red or green.

Line 61: For the sample numbers in the abstract - can they confirm this is with the "suspected duplicated samples" (line 156) removed? It would be best to provide the number of unique genomes generated (and at line 134).

Line 109: adaptation "and to the smallholder husbandry practiced throughout the archipelago" - not sure exactly what's being suggested here, that they have "adapted to husbandry practices"? Perhaps the authors mean that their gene pool has been shaped by husbandry practices - "adapted" to us means something specific with regard to selection pressures. Can the authors clarify?

Lines 197-198: can you provide F_{st} ranges with between cattle and banteng, to contextualize the lower F_{st} with banteng seen in some Indonesian breeds (line 198-199)

Figure 1a - please increase the size of the symbols, they don't pop out against the background enough.

Figure 1b: can you confirm the outlier Kupang point is N_31B, the putative F1 hybrid?

Line 210: Can you clarify which reference alignment was used for heterozygosity estimation? BosTau9 or Waterbuffalo. In general, it is not clear which dataset was used for which analysis e.g. this applies ROH as well/Figure 2 caption.

Line 251: please specify "some indicine cattle introgression"

Line 767: cattle instead of cattles

Line 777-778: The sentence "We merged distinct ROHs within a 100 kb distance" raised some concerns. PLINK offers built-in parameters, such as --homozyg-gap (which defaults to a 1000 kb gap between SNPs) and --homozyg-density (<https://www.cog-genomics.org/plink/1.9/ibd#homozyg>), to control the distance between SNPs and SNP density. We wondered why these parameters weren't explored further, and why the decision was made to merge ROHs within a 100 kb distance instead. Are there other studies of Bos species which also utilize this parameter choice? It would be helpful to validate the merging of ROHs that PLINK would typically split to ensure the approach is sound.

Line 990-1012: Segment Overlapping introgressed segments among LOTER, Hmmix, and UX . We were curious about why the top 5% quantile regions from the different methods don't overlap. Is this a result of the different approaches or sensitivities of the methods? Additionally, how were the values from Hmmix and Ux (mentioned in line 1012) used for further validation. Was this done using a specific threshold?

Line 1001: This is the first mention of Uabc80, should this be Uabc20? (Line 923)

Line 1005-1007: "Considering the tendency of Hmmix to predict more false positives on regions of banteng ancestry than LOTER, we only kept windows that passed the top 5% quantile of LOTER values in each cattle group." Should be accompanied by either a citation or a justification.

Supplementary Figure 25: Would it be possible to give the points a black outline in this figure. It is difficult to determine individuals.

Signed: Kevin G Daly & Jolijn A. M. Erven

(Remarks on code availability)

I looked over ~half the code - it's great to see this being available. In particular the genotype calling is very detailed. This will certainly contribute to reproducibility of the study analyses.

Reviewer #2

(Remarks to the Author)

This is an important paper that addresses, for the first time, the diversity and origins of Indonesian cattle at the full genome level 233 whole genome sequences, including 179 new Indonesian cattle genomes. Why is it important? Indonesian cattle are a unique genetic mix-up of different cattle lineages, including indicine and indigenous banteng, possibly on top of some ghost bovid species. The Indonesian Peninsula has been at the crossroads of trading routes, which may have included domesticated Bovids. The source of the Banteng genome in Southeast Asia might be the Indonesian Peninsula, and as such, understanding the diversity of the source population and its influence on today's modern cattle in the region provides important baselines for further study.

Main finding include that The Madura breed is one of Indonesian cattle's most genetically diverse breeds. This is due to the unprecedented genetic introgression from banteng and other Bos species, comprising up to 36.6% Madura genome. The paper claims the Zebu ancestors for the Indonesian cattle are from three sources, two of which are from mainland southeast and Eastern Asia, challenging the previous studies that pinpointed the Zebu ancestors to India. The paper identified breed-specific adaptive introgression and convergent adaptive introgression, such as for the coat color. This study highlighted insights into the origin, history, and introgression of Indonesian admixed cattle using population structure, genetic diversity, and admixture process in Indonesian cattle.

The paper is very well written. The findings have not been overinterpreted with conclusion based on sound analysis,

Strengths of the Study

Comprehensive genomic analysis of 233 whole genomes, supplemented with global cattle diversity data. Robust methodologies, including admixture modelling and ancestry inference, yield high-resolution insights into population history and genetic diversity. Identifying novel single-nucleotide polymorphisms (SNPs) linked to adaptive traits, expanding the genetic toolkit for livestock improvement.

Limitations and Future Directions

While the study offers several new and scientific insights, several areas need further exploration:

Underrepresentation of Wild Bos Species: Limited genomic data from wild Bos species restricts the ability to characterize introgression sources fully.

Reference bias: All conclusions are based on the SNPs generated using female taurine from Hereford and female water buffalo from the Mediterranean breed as a reference. Harnessing the pan-genomic approach would give more insights into ancestral history and introgression. In other words, while they are today taurine and indicine de novo genome of reference, it is not the case for Bos banteng. As one of the main points of the paper is to highlight the important genetic contribution of the latter, to which extent the results presented here may be different if a Bos banteng de novo genome had been available and used as a reference, should be mentioned in the discussion section

Broader Contextualization: The paper is a little bit weak in integrating historical, ecological, and cultural data. This should be expanded in the discussion to enrich further interpretations of cattle dispersal and admixture in Southeast Asia.

Indicine references from different geographic areas: Besides Indonesia (geographic gap now addressed in this study), identifying the origins of zebu introgression requires references indicine breeds from the geographic areas where the source population might have originated. But for some, they are lacking and/or are too few (for example, the Southern part of the Indian subcontinent, including Sri Lanka, South East Asia, today Bangladesh etc.). To what extent should the inclusion of such samples further highlight the results and perhaps even change some of the conclusions that should be discussed?

The title of the paper is informative but also unfortunate. In my opinion, my suggestion will be to replace 'megadiverse' with 'unique diversity' (the word 'megadiverse' might be more appropriate for a 'new and view).

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

We thank the authors for their thoughtful and constructive response to our comments. The revised manuscript carefully incorporates the feedback provided, and the changes have, in our view, strengthened the work. Our initial concerns have been addressed, and we have no further comments. However, the use of the word "pure" in Line 195 could be replaced with a value-neutral term such as "unadmixed".

Imputation

Thank you for including a downsampling approach to test imputation accuracy — we are pleased to see that the imputation accuracy is high for the downsampled Banteng individuals. This is very promising! The addition of this test, along with the accompanying visuals, has strengthened the imputation approach and its ability to accurately impute Banteng ancestry (particularly in low-coverage individuals).

ROH

Thank you for expanding on the approach to ROH calling. We acknowledge the sensitivity of ROH inference to technical artifacts and understand the challenge of defining "true" ROH. With the additional validation, visualisations and explanation provided, we find the current approach appropriate and well-supported.

Local ancestry inference

We are grateful for the further exploration of reference panel configurations. We now agree with the authors that including the full population as a reference is the most robust choice.

Color Blindness Consideration

Thank you for exploring color blindness – We recognize the challenges involved in selecting accessible color palettes when visualizing numerous populations or subsets.

LAI overlapping genes

We appreciate the addition of visualizations showing overlapping genes across different local ancestry inference methods. These observations also add valuable support to the statement in Lines 1005–1007.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

This is a revised version of a previously submitted manuscript. The authors have conducted a comprehensive response to the reviewers, including some new analyses. It is greatly appreciated. The work will catalyse further studies on the genetic diversity of Southeast Asian cattle and beyond. I have no further comments on the manuscript. Well done

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

The code is well structured and annotated. However, the README file can be enhanced more to give detained instruction and guidance on how to reproduce the results.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <https://creativecommons.org/licenses/by/4.0/>

RESPONSE TO REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

Wang et al report over 200 genomes from East/Southeast Asian Bos species: Indonesian cattle and domesticated banteng (*Bos javanicus*). They find widespread banteng ancestry in certain Indonesian cattle breeds such as Madura and Pesisir, with distinct breed-specific patterns. They disentangle this admixture using local ancestry inferences, revealing distinct admixture patterns in Sumatran compared to Javan cattle; timing of admixture using local ancestry also indicates that these were distinct processes within the last ~2000 years. The authors also estimate heterozygosity of the zebu & banteng sources of admixture, inferring distinct population histories. The authors estimate diversity parameters of various East/Southeast Asian breeds, inferring high heterozygosity in Indonesian cattle, low heterozygosity in Bali cattle/banteng, and notable runs of homozygosity in certain populations of banteng. The authors finally assess the potential selection acting on introgressed banteng variants in Indonesian cattle, assessing functional enrichment within breeds and regions with an excess of banteng ancestry across breeds, finding evidence in selection milk-associated, pigmentation-associated, and immune-associated genes.

This is a well-written, informative paper that coincidentally produces a valuable resource of Bos genomes. The paper advances the current understanding of Bos diversity, ancestry, and population history, using current population genetic techniques and some creative analyses (e.g. leveraging the local ancestry traits to infer the genetic makeup of the unsampled source populations). The inference of human-directed admixture with wild Bos (due to the mtDNA patterns and large introgression despite biological barriers) is also a very interesting finding. To our knowledge these data described here are considerably advancing the genetic resources available for these species and breeds. The paper is building off a literature of (cited) work of Indonesian cattle and Bali cattle; the scale (in terms of genome wide data and individuals) and quality of the analyses. In general the data & analyses supports the conclusions of the paper - this has the very real potential to guide efforts of preserving or utilizing genetic variation existing in Indonesian cattle breeds; as will the signals of low diversity in Bali cattle guide breed maintenance. The general findings of negative selection against introgressed variation, and potential convergent adaptive introgression, will also contribute to ongoing discussions within evolutionary biology of the role of introgression.

The authors provide the raw data and the scripts used in this study on their github. The methods are detailed and in combination with the available scripts the work can be reproduced. However, it would be helpful if the authors could specify the number of

SNPs used for each analysis to enhance transparency and provide a clearer understanding of the scope of the dataset. (I cannot seem to find the Haplonet script in the github repository).

***Answer: Thanks for the comment. We have added the number of SNPs used for each analysis in the SI as new Supplementary Data 4 and added the statement to L156-L158 in the revised manuscript with marked changes. We also have supplied HaploNet scripts to the study page on github.**

The methodology employed in this study is well-structured and robust; the use of multiple methods for local ancestry inference is particularly commendable. However, some points of the analyses are criticised below and should be resolved - these represent our major comments and are constructive. But overall, the methodology is well-justified and meets the expected standards in our field.

***Answer: We thank the reviewers for their positive assessment and constructive comments. We will address their concerns individually below.**

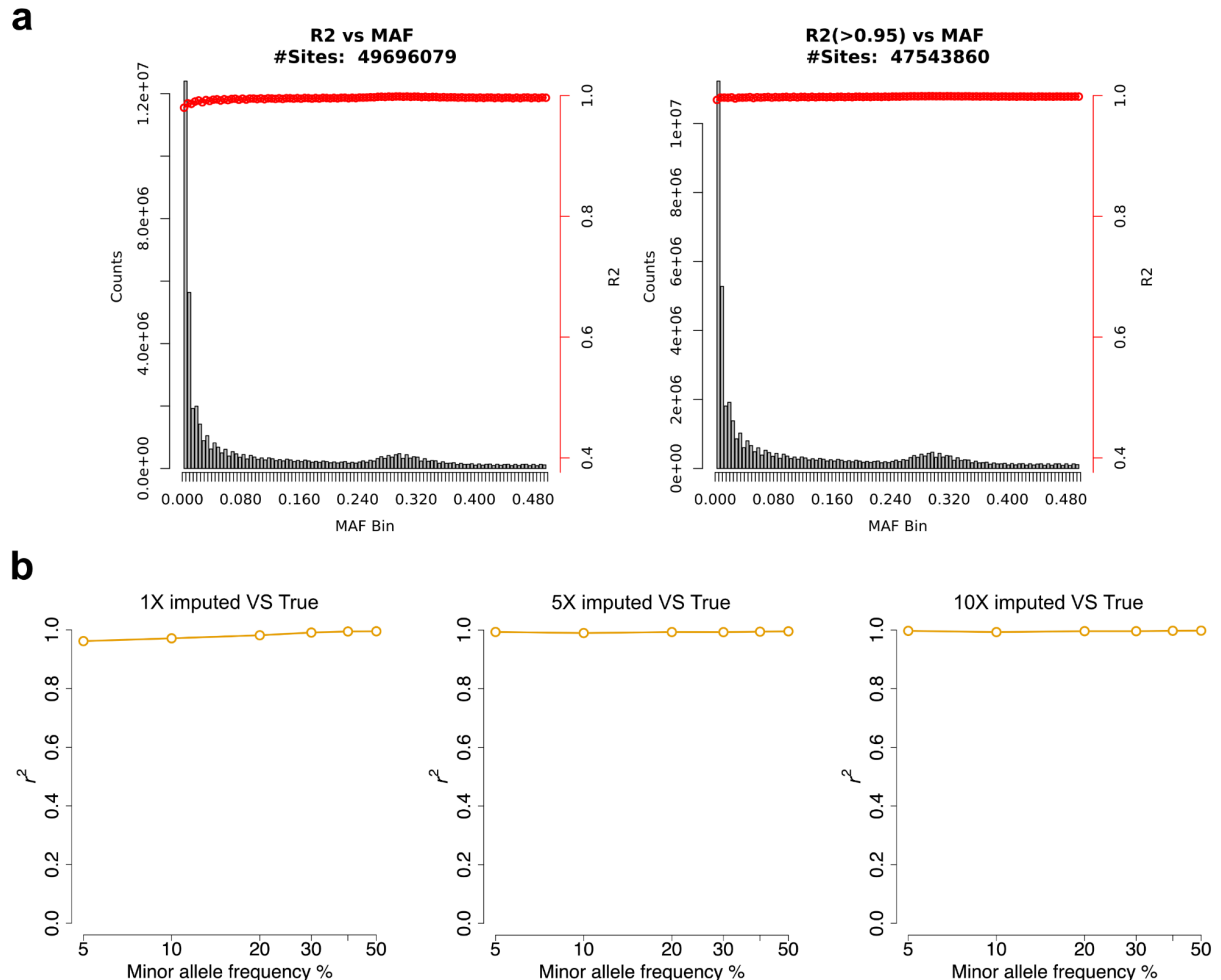
Major comments

Line 745-751 - BEAGLE imputation.

How well did the genotype imputation with BEAGLE 3 perform, has genotype imputation been tested on Bali cattle/Banteng and diverse cattle breeds (such as the Indonesian breeds), how well would these potential diverse haplotypes impute? In line 750-751 a visualization of distribution of genotype discordance between original and imputed VCF files was mentioned, but we cannot find these in the manuscript or SI. This section of the analyses - underlying many downstream parts of the manuscript - does require a little more scrutiny: the validation of genotype imputation of the Bali cattle/Banteng and similar diverse cattle breeds or a more in-depth discussion of the genotype discordance and imputation performance. The highest coverage Bentang individual has an average genome coverage of 34.86X, it would be interesting to see how well these high-coverage genotypes are imputed. Were imputed genotypes filtered after imputation, if yes please add this information.

***Answer: Thanks for pointing out this omission. The statement in the manuscript was referring to L756-758, but the corresponding plot was accidentally not included in the SI. We have now added it as Supplementary Fig. 30a. In addition, we conducted the suggested analysis, downsampling the highest-depth banteng individual (34.86X, LIB112407_Banteng_85B_Texas) to 1X, 5X, 10X depth, then imputing the downsampled data, and comparing these imputed genotypes with the high-quality genotype calls for**

the full data from this sample. As can be seen in the plots below (restricted to sites on Chromosome 1), this analysis showed a very high imputation accuracy. We have added these results to L758-764 in the revised manuscript with marked changes and to the SI as new Supplementary Fig. 30b.



Supplementary Figure 30. Evaluation of accuracy of imputation. a, Visualization of minor allele frequency (MAF) distribution of genotype discordance between the original vcf and imputed vcf genotype files. R^2 is the estimated squared correlation ($0 \leq R^2 \leq 1$) between the allele dosage with highest posterior probability in the genotype probabilities file and the true allele dosage for the marker. b, Visualisation of MAF distribution (chromosome 1) of individual LIB112407_Banteng_85B_Texas (depth of 34.86 X) between imputed genotypes (downsampled into depths of 1X, 5X and 10X) and the high-quality genotype calls for the full data from this sample. MAFs were calculated using all of the individuals in this study.

Line 772-779 - Regions of Homozygosity.

Could you clarify why the threshold of 5 heterozygous SNPs per sliding window was chosen for the ROH analysis? Was this parameter based on previous research? If so, please provide the relevant citation. If not, please justify the selection of this threshold. One approach to do so would be to estimate the potential heterozygosity error rate,

which could result from sequencing or alignment errors, and demonstrate how this aligns with the chosen parameter. If potential false-positive heterozygous rate due to imputation might be a problem, please specify and show accordingly. Also, please specify the number of SNPs used for this analysis.

***Answer: Thanks for the comment. ROH calling is notoriously sensitive to data artefacts, and each data set needs a careful assessment of the performance of the ROH calling. Our approach for determining settings for the ROH calling was similar to that taken in (Liu et al. 2024), and we have now added a citation of this paper in the Method section (L787-793). Our final settings for the ROH calling was based on exploratory analyses that used visualization of ROH calling along with heterozygosity in bins and SNP density across chromosomes to compare ROH calling accuracy under different settings. Here, we determined that the settings reported in L795-797 yielded the best ROH calling of those considered. We include plots used to reach this decision by adding an example to the Supplementary as the new Supplementary Fig. 6.**

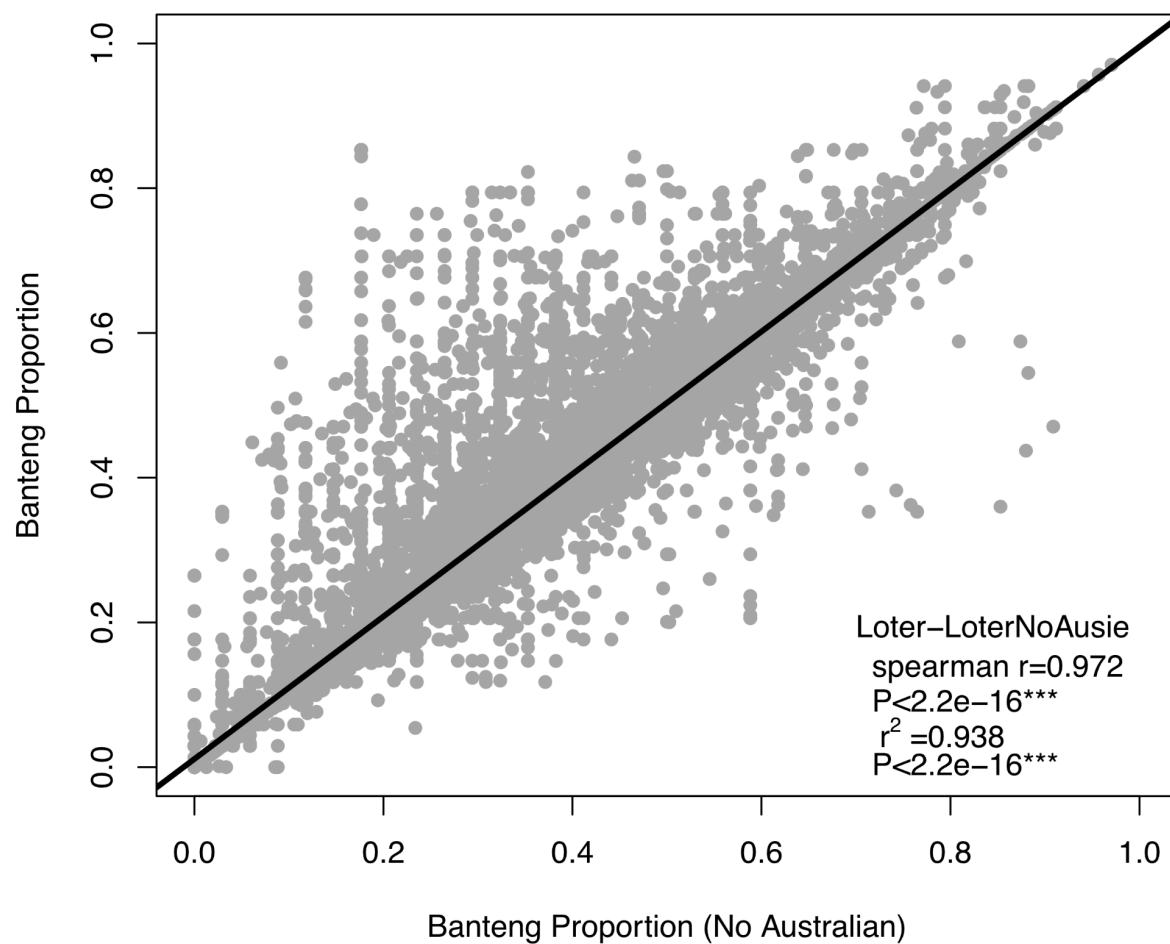
We acknowledge that ROH calling (and especially their exact length distribution) is challenging, and we can see from our plots that our ROH calling is still not perfect, as it still splits up some stretches that appear like longer contiguous ROHs, making us potentially overlook some very long ROHs (but probably not undercall the total amount of FROH in each sample). We have added some reservations to this effect in the revised manuscript (L219-224). Also, we emphasize that the main result here is the comparison between samples of similar quality and depth, which we think remains informative regardless of whether we have achieved “perfect” ROH calling.

Line 849-872 Local ancestry inference

It is welcoming to see that three different methods were used to infer local ancestry—this approach is very comprehensive. However, there are some concerns regarding the use of the inbred Australian Bali cattle population as one of the ancestry source references. Given the high amount of ROH and low diversity in this population, could this have affected the local ancestry inference? Also considering the variation observed in Figure 4E and S10-11. Would it be possible to rerun the analyses on a subset of Bali cattle, excluding the Australian Bali cattle, to assess the impact? If not, could you clarify whether including this population would not significantly affect the results?

***Answer: Thanks for the comment. We have now rerun the LOTER analyses removing Bali cattle from Australia from the reference panel for Madura population. What we found was a very high correlation between this and the original LOTER results (see plot below), however with inferred banteng ancestry being systematically lower when removing the Australian Bali cattle from the reference population. This is consistent with our expectation that it should not be problematic to include a low-diversity population in one**

of the reference populations, as long as it is not admixed with another population potentially closer to the other possible reference population(s). In this case, including Australian Bali cattle simply adds a few extra banteng haplotypes not found in the rest of the banteng individuals of the reference panel, without which some regions of true banteng ancestry are overlooked by LOTER. For this reason, we decided to retain the use of the original reference panel that includes Australian Bali cattle. We have added these results to L379-382 in the revised manuscript with marked changes and to the SI as new Supplementary Fig. 17.



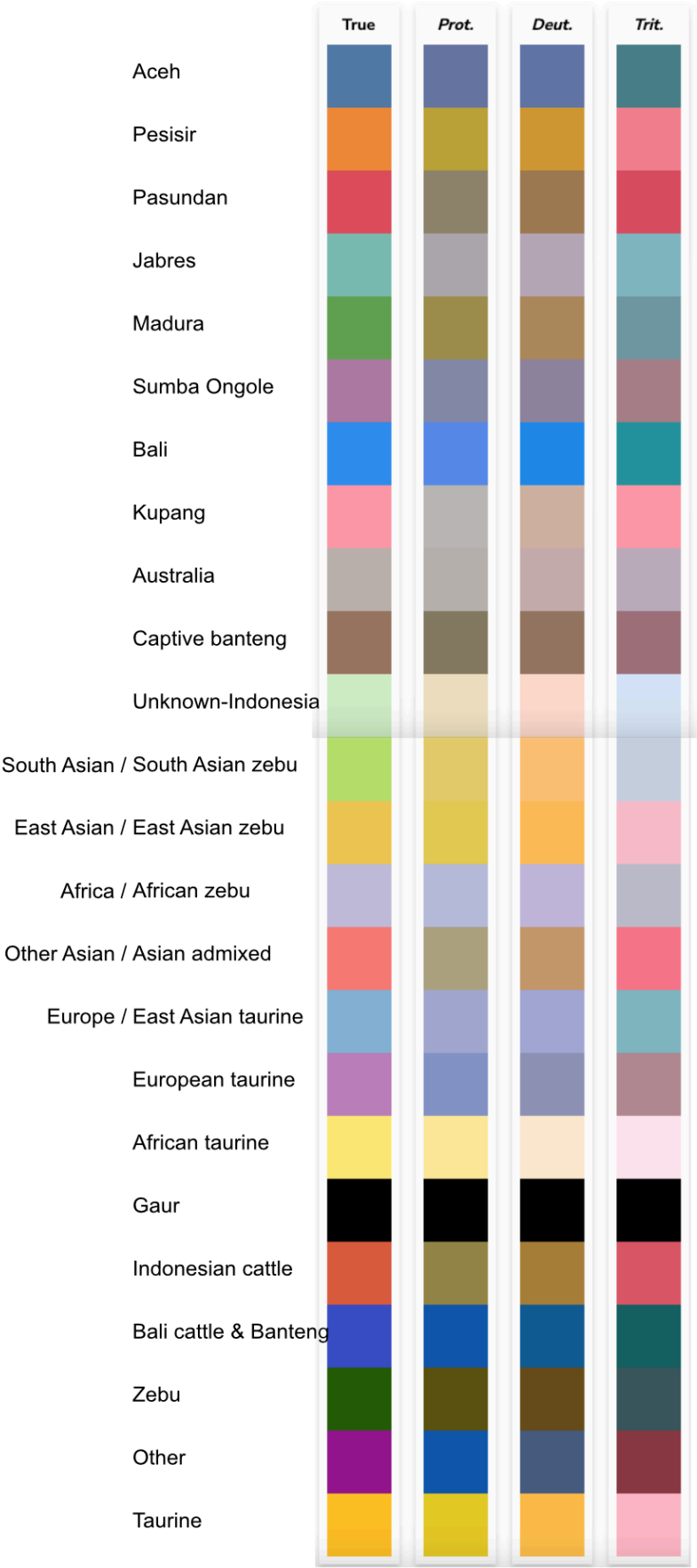
Supplementary Figure 17. Correlation between banteng ancestry proportion in Madura population when including and excluding Australia population from the ancestry source reference panel.

Minor comments

The figures in general should be made colour-blind friendly, even if just removing red or green.

***Answer:** We have done our best to choose a visually appealing and clear color scheme. The challenge is that we have a total of 24 different groupings that we use in the manuscript, which is a very high number to avoid all possible colour-blind issues while keeping an appealing color palette. Following the reviewer comment, we used a website (<https://davidmathlogic.com/colorblind/#%234E79A7-%23F28E2B-%23E15759-%2376B7B2-%2359A14F-%23B07AA1-%231C86EE-%23FF9DA7-%23BAB0AC-%239C755F-%23CCEBC5-%23B3DE69-%23EDC948-%23BEBADA-%23FB8072-%2380B1D3-%23BC80BD-%23FFED6F-%23000000-%23DD6336-%233743C8-%231E5C05-%23971C8F-%23FFC300>) to assess the color blind friendliness of our color scheme and found it to be high for a 24 color palette (see below). We also want to highlight that in plots where different cattle groups are being shown, we either use different symbols along with different colors (e.g. Fig. 1b, Fig. 2b) or demarcate the different groups on one of the axes (e.g. Fig. 2a, Fig. 3b, Fig. 3c, Fig. 4a). We therefore chose to retain the current color palette for the sake of clarity and aesthetics. We would be happy to change it if the reviewer can provide concrete examples of where it causes confusion.

Color Palette



Line 61: For the sample numbers in the abstract - can they confirm this is with the "suspected duplicated samples" (line 156) removed? It would be best to provide the number of unique genomes generated (and at line 134).

***Answer: Thanks for the comment. Following this recommendation, we have changed the number in the Abstract to be the actual unique individuals, after removing duplicates. We also clarified the sample sizes of newly generated and downloaded data in L150-154 in the revised manuscript with marked changes.**

Line 109: adaptation "and to the smallholder husbandry practiced throughout the archipelago" - not sure exactly what's being suggested here, that they have "adapted to husbandry practices"? Perhaps the authors mean that their gene pool has been shaped by husbandry practices - "adapted" to us means something specific with regard to selection pressures. Can the authors clarify?

***Answer: Thanks for the comment. The full statement in the manuscript reads: "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", and that is the meaning we wanted to convey - i.e. that selection imposed by both the physical environment (climate, diseases) and the husbandry practices (e.g. feeding with particular items, requiring cattle to perform a specific type of work) has acted on the gene pool of indigenous cattle, and that it is therefore meaningful to say that they "adapted" to these conditions. We do not see any problem in stating that cattle can adapt to local husbandry practices (if the reality is that their gene pool has been shaped by such practices in a manner consistent with selection), but maybe this is a subjective opinion, and we would be happy to consider alternative wording suggestions.**

Lines 197-198: can you provide F_{ST} ranges with between cattle and banteng, to contextualize the lower F_{ST} with banteng seen in some Indonesian breeds (line 198-199)

***Answer: We have now added F_{ST} ranges between unadmixed cattle and banteng, and between Indonesian cattle and banteng in L195 in revised manuscript with masked changes.**

Figure 1a - please increase the size of the symbols, they don't pop out against the background enough.

***Answer: Thanks for the comment. We have now increased the size of each sampling location marking in a revised Fig. 1 in the manuscript.**

Figure 1b: can you confirm the outlier Kupang point is N_31B, the putative F1 hybrid?

***Answer: Yes, this is correct. We have now highlighted this in the figure legend in L204-205 in the revised manuscript and highlighted the changes.**

Line 210: Can you clarify which reference alignment was used for heterozygosity estimation? BosTau9 or Waterbuffalo. In general, it is not clear which dataset was used for which analysis e.g. this applies ROH as well/Figure 2 caption.

***Answer: We used BosTau9 for the heterozygosity estimation. In response to this comment we have added a new table with an overview of which data is being used in which analyses as new Supplementary Data 4.**

Line 251: please specify "some indicine cattle introgression"

***Answer: We rephrase this sentence in L254-255 in the revised manuscript and highlighted the changes.**

Line 767: cattle instead of cattles

***Answer: Corrected. See L780 in the revised manuscript.**

Line 777-778: The sentence "We merged distinct ROHs within a 100 kb distance" raised some concerns. PLINK offers built-in parameters, such as `--homozyg-gap` (which defaults to a 1000 kb gap between SNPs) and `--homozyg-density` (<https://www.cog-genomics.org/plink/1.9/ibd#homozyg>), to control the distance between SNPs and SNP density. We wondered why these parameters weren't explored further, and why the decision was made to merge ROHs within a 100 kb distance instead. Are there other studies of Bos species which also utilize this parameter choice? It would be helpful to validate the merging of ROHs that PLINK would typically split to ensure the approach is sound.

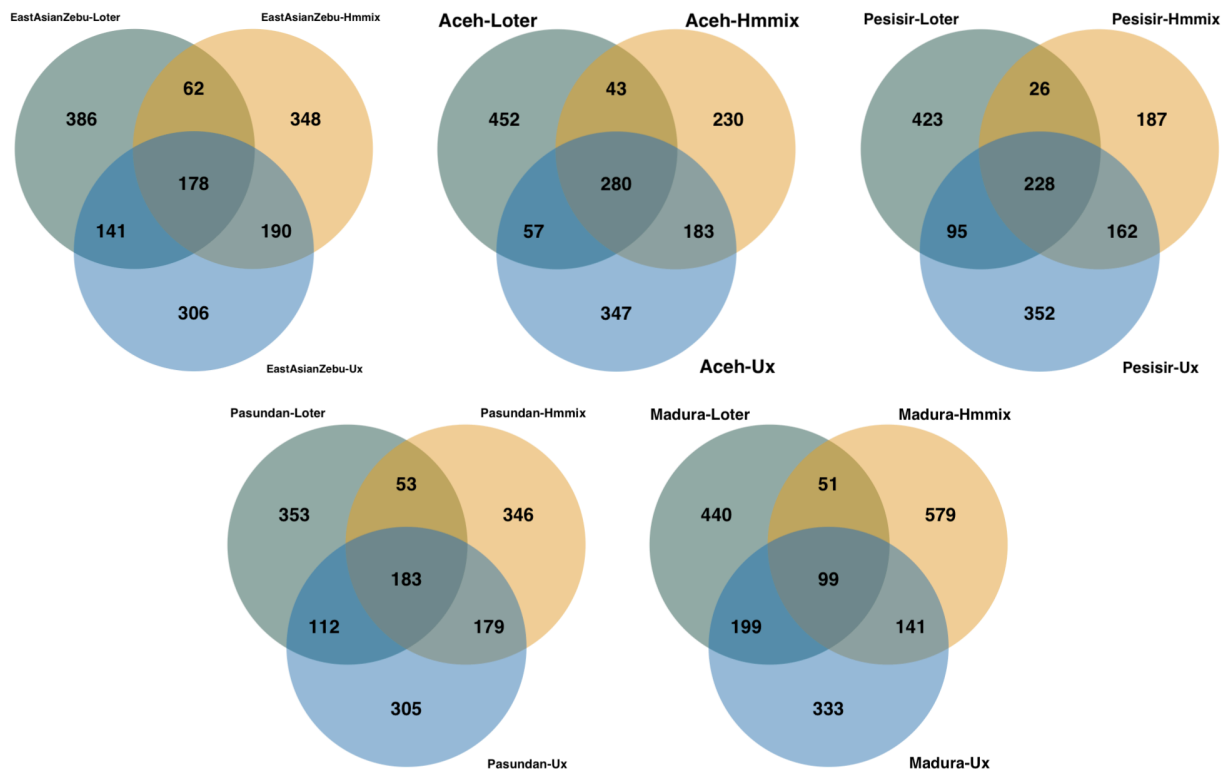
***Answer: Thanks for the comment. As stated in a response above, we explored and visually assessed different settings for ROH calling. We acknowledge that ROH detection is difficult to generalize, as different data sets will need different filtering approaches to get optimal ROH calling. Our approach is visual, and based on the logic that long ROHs that are caused by recent inbreeding are visible if one plots heterozygosity in this manner along chromosomes, notably if heterozygosity is relatively high outside of ROHs. Therefore, a visual performance check of different settings is our chosen approach. We have added reservations about the inherent difficulty in identifying ROHs in L219-224, and elaborated on our chosen approach in L787-804 of the revised manuscript.**

Line 990-1012: Segment Overlapping introgressed segments among LOTER, Hmmix, and U_x . We were curious about why the top 5% quantile regions from the different

methods don't overlap. Is this a result of the different approaches or sensitivities of the methods? Additionally, how were the values from Hmmix and Ux (mentioned in line 1012) used for further validation. Was this done using a specific threshold?

***Answer:** The top 5% of each method does overlap (see Supplementary Fig. 23), although not perfectly. We believe this is due to the fundamental differences in how these methods estimate introgression. For example, Hmmix consistently infers a higher proportion of introgressed ancestry than LOTER and Uabc, and this 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 used the observation that these adaptive introgression statistics with potentially different strengths and weaknesses are strongly, although not perfectly correlated, mainly as an overall sanity check, or robustness analysis. We have clarified this in L1052-1057, and L1063 in the revised manuscript with masked changes.

We also reported the gene list targeted by top 5% outliers that were overlapping between any two out of three methods as new Supplementary Data 9-11, and the Venn diagram plot showing the number of overlapping genes for each Indonesian breed as new Supplementary Fig. 25 as shown below. We have added these new results in L438-441 in revised manuscript and highlighted changes.



Supplementary Figure 25. Venn diagram plot showing the number of overlapping genes for each Indonesian breed between any two out of three methods (LOTER, Hmmix, and U_x). Gene ID lists are shown in Supplementary Data 9-11.

Line 1001: This is the first mention of Uabc80, should this be Uabc20? (Line 923)

***Answer: Yes, thanks for the comment. We corrected this sentence in L1048 in the revised manuscript.**

Line 1005-1007: "Considering the tendency of Hmmix to predict more false positives on regions of banteng ancestry than LOTER, we only kept windows that passed the top 5% quantile of LOTER values in each cattle group." Should be accompanied by either a citation or a justification.

***Answer: See response above, this is based on the empirical observation that Hmmix systematically infers a higher proportion of introgressed ancestry in each population relative to the two other methods. This has now been clarified in L1052-1057.**

Supplementary Figure 25: Would it be possible to give the points a black outline in this figure. It is difficult to determine individuals.

***Answer: Please see updated Supplementary Fig. 31.**

Signed: Kevin G Daly & Jolijn A. M. Erven

Reviewer #1 (Remarks on code availability):

I looked over ~half the code - it's great to see this being available. In particular the genotype calling is very detailed. This will certainly contribute to reproducibility of the study analyses.

***Answer: Thanks, we agree that method transparency greatly aids in reproducibility.**

Reviewer #2 (Remarks to the Author):

This is an important paper that addresses, for the first time, the diversity and origins of Indonesian cattle at the full genome level 233 whole genome sequences, including 179 new Indonesian cattle genomes. Why is it important? Indonesian cattle are a unique genetic mix-up of different cattle lineages, including indicine and indigenous banteng, possibly on top of some ghost bovid species. The Indonesian Peninsula has been at the

crossroads of trading routes, which may have included domesticated Bovids. The source of the Banteng genome in Southeast Asia might be the Indonesian Peninsula, and as such, understanding the diversity of the source population and its influence on today's modern cattle in the region provides important baselines for further study.

Main findings include that The Madura breed is one of Indonesian cattle's most genetically diverse breeds. This is due to the unprecedented genetic introgression from banteng and other Bos species, comprising up to 36.6% Madura genome. The paper claims the Zebu ancestors for the Indonesian cattle are from three sources, two of which are from mainland southeast and Eastern Asia, challenging the previous studies that pinpointed the Zebu ancestors to India. The paper identified breed-specific adaptive introgression and convergent adaptive introgression, such as for the coat color. This study highlighted insights into the origin, history, and introgression of Indonesian admixed cattle using population structure, genetic diversity, and admixture process in Indonesian cattle.

The paper is very well written. The findings have not been overinterpreted with conclusion based on sound analysis,

Strengths of the Study

Comprehensive genomic analysis of 233 whole genomes, supplemented with global cattle diversity data.

Robust methodologies, including admixture modelling and ancestry inference, yield high-resolution insights into population history and genetic diversity. Identifying novel single-nucleotide polymorphisms (SNPs) linked to adaptive traits, expanding the genetic toolkit for livestock improvement.

***Answer: We thank the reviewer for this positive assessment.**

Limitations and Future Directions

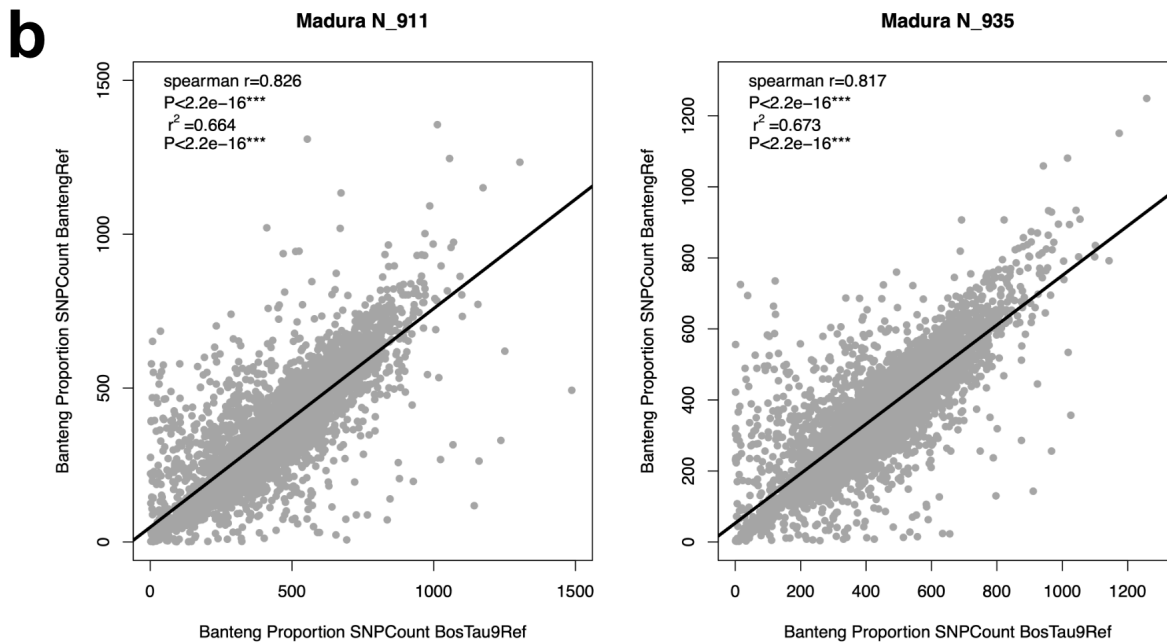
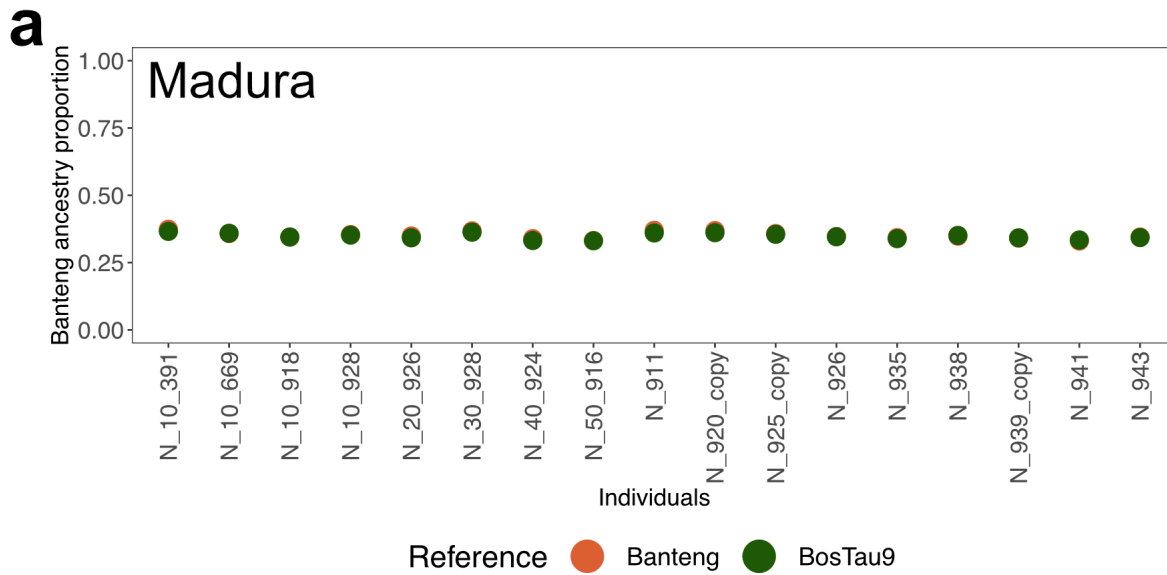
While the study offers several new and scientific insights, several areas need further exploration:

Underrepresentation of Wild Bos Species: Limited genomic data from wild Bos species restricts the ability to characterize introgression sources fully.

***Answer: We agree with the reviewer about this limitation. Unfortunately, the limited availability and accessibility of such data/samples place some restrictions on what is achievable.**

Reference bias: All conclusions are based on the SNPs generated using female taurine from Hereford and female water buffalo from the Mediterranean breed as a reference. Harnessing the pan-genomic approach would give more insights into ancestral history and introgression. In other words, while they are today taurine and indicine de novo genome of reference, it is not the case for Bos banteng. As one of the main points of the paper is to highlight the important genetic contribution of the latter, to which extent the results presented here may be different if a Bos banteng de novo genome had been available and used as a reference, should be mentioned in the discussion section

***Answer: Thanks for the comment and the constructive suggestion. Following this suggestion, we mapped our data to a newly generated banteng reference genome (*Bos javanicus*, RefSeq: GCF_032452875.1-ARS-OSU_banteng_1.0) and redid a number of the analyses, notably the LOTER analysis, to investigate the sensitivity of the LAI conclusions to the choice of reference genome. Results were very similar to the results based on the bosTau9 mapping, and we added this analysis to the manuscript with masked changes (L379-382; L891-907; Supplementary Fig. 18).**



Supplementary Figure 18. Evaluation of how reference genome (Banteng and BosTau9) influence local ancestry inference by LOTER. a, Individual admixture proportion from LOTER when using Banteng and BosTau9 as reference genome. b, Correlation between SNP counts inferred to be of banteng ancestry in each 50 kb genomic window for two individuals (N_911 and N_935) from Madura population.

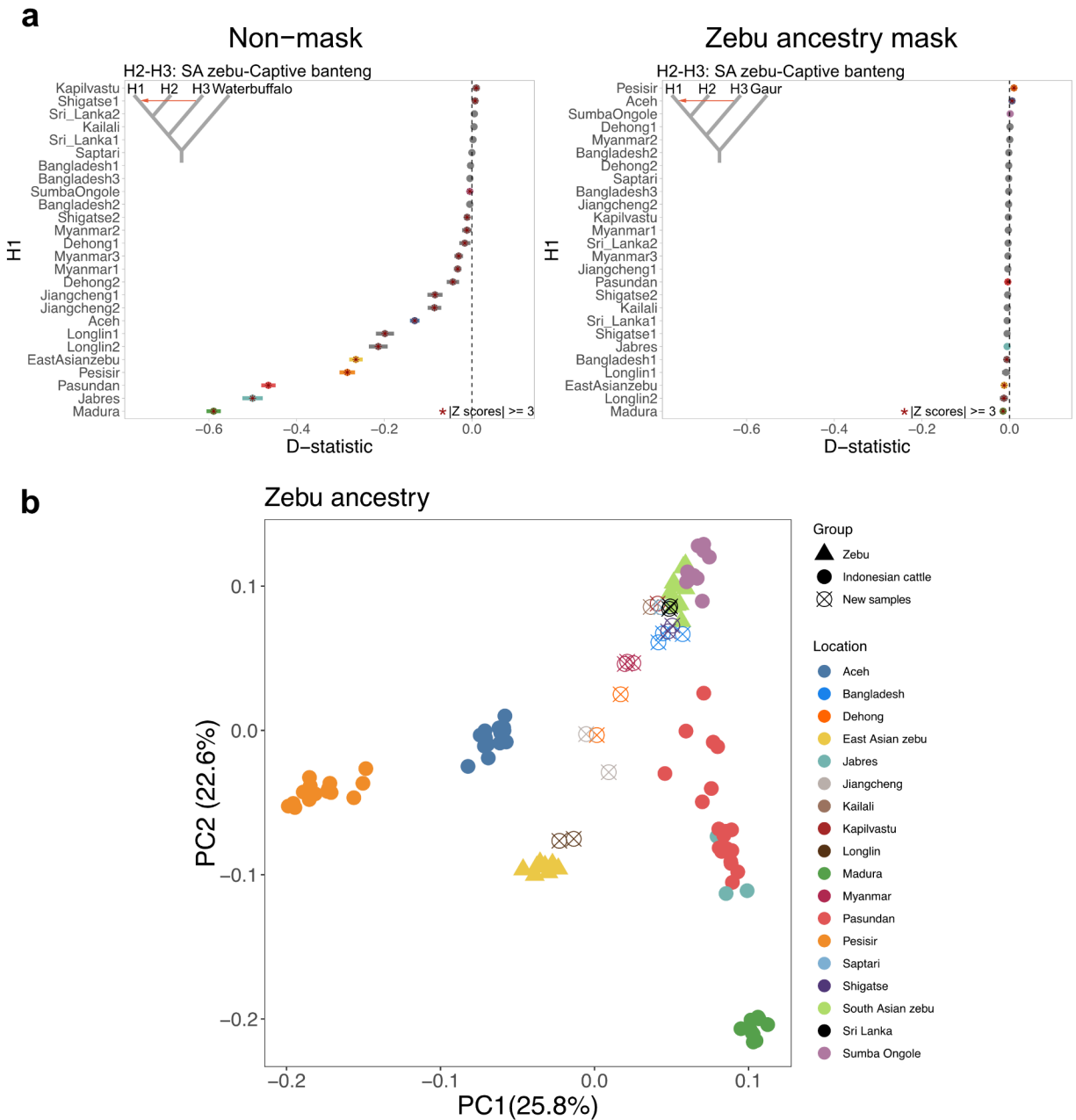
Broader Contextualization: The paper is a little bit weak in integrating historical, ecological, and cultural data. This should be expanded in the discussion to enrich further interpretations of cattle dispersal and admixture in Southeast Asia.

***Answer: Thanks for this suggestion. We strived to keep the manuscript concise, which to some extent conflicts with the integration of the sources of information mentioned here. Also, we highlight that historical and cultural sources of relevance to the main conclusions are not necessarily consistent and provide limited tangible evidence. However, following the reviewer comment we investigated further and found additional literature based on a recent study considering historical information and linguistics, and this information has now been integrated into the Discussion (L547-553).**

Indicine references from different geographic areas: Besides Indonesia (geographic gap now addressed in this study), identifying the origins of zebu introgression requires references indicine breeds from the geographic areas where the source population might have originated. But for some, they are lacking and/or are too few (for example, the Southern part of the Indian subcontinent, including Sri Lanka, South East Asia, today Bangladesh etc..). To what extent should the inclusion of such samples further highlight the results and perhaps even change some of the conclusions that should be discussed?

***Answer: Thanks for the comment. A recent study published during the preparation of our manuscript did add a few new genomes from mainland Asia that could be relevant for our analyses (although none from further south than Myanmar in northernmost Southeast Asia). Therefore, in response to this comment we added a selected subset of these new samples (Supplementary Data 7) to assess whether they provide additional evidence for the origins of cattle introduction into Indonesia. Whereas these samples do provide insights into the dispersal of zebu cattle from South Asia to East Asia, they did not resolve key questions about the origins of Indonesian cattle. We expect that only additional genome data from further south in the Southeast Asian mainland has the potential to clarify these questions, assuming that extant cattle in these regions retain some of the ancestry from the original dispersals.**

To explore the information content of these newly sequenced SEA cattle samples, we added the relevant samples to the EMU PCA analysis in Supplementary Fig. 11b and modified the Results, Discussion and Methods accordingly in the revised manuscript with masked changes (L316-320; L328-333; L530-553; L1010-1025).



Supplementary Figure 11. D-statistics calculated by ADMIXTOOLS2 a, when using all sites (left panel), and masked genome segments of zebu ancestry origin inferred by LOTER (right panel). Data are presented as the estimated D-statistic \pm 3 standard errors. Star represents significant allele sharing for each combination. b, PCA analysis on genome segments of zebu ancestry origin when adding newly downloaded additional samples from northernmost Southeast Asia and East Asia (e.g. Myanmar and southern China), as inferred by LOTER.

The title of the paper is informative but also unfortunate. In my opinion, my suggestion will be to replace 'megadiverse' with 'unique diversity' (the word 'megadiverse' might be more appropriate for a 'new and view'.

***Answer: Thanks, we acknowledge that this term might be considered inappropriate in this context. We have replaced the previous title with “Multiple origins and adaptive introgression shaped the unique diversity of Indonesian cattle”.**

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

***Answer: Thanks, we warmly support this reviewing option.**

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

***Answer: Thanks, we warmly support this reviewing option.**

Reference

Liu X., Lin L., Sinding M.-H.S., Bertola L.D., Hanghøj K., Quinn L., Garcia-Erill G., Rasmussen M.S., Schubert M., Pečnerová P., Balboa R.F., Li Z., Heaton M.P., Smith T.P.L., Pinto R.R., Wang X., Kuja J., Brüniche-Olsen A., Meisner J., Santander C.G., Ogotu J.O., Masembe C., da Fonseca R.R., Muwanika V., Siegismund H.R., Albrechtsen A., Moltke I., Heller R. 2024. Introgression and disruption of migration routes have shaped the genetic integrity of wildebeest populations. *Nat. Commun.* 15:2921.