

RESEARCH

Open Access



Genomic evidence refutes the hypothesis that the Bornean banteng is a distinct species

Xin Sun¹ , Marta Maria Ciucani² , Jacob Agerbo Rasmussen² , M. Thomas P. Gilbert^{1,2} and Mikkel-Holger S. Sinding^{3*}

Abstract

The banteng (*Bos javanicus*) is an endangered species within the wild Asian *Bos* complex, that has traditionally been subdivided into three geographically isolated subspecies based on (i) mainland Southeast Asia (*B. j. birmanicus*), (ii) Java (*B. j. javanicus*), and (iii) Borneo (*B. j. lowi*). However, analysis of a single Bornean banteng mitochondrial genome generated through a genome skimming approach was used to suggest that it may actually represent a distinct species (Ishige et al. in *Mitochondrial DNA A DNA Mapp Seq Anal* 27(4):2453–4. <http://doi.org/10.3109/19401736.2015.1033694>, 2016). To explore this hypothesis further, we leveraged on the GenBank (NCBI) raw read sequencing data originally used to construct the mitochondrial genome and reconstructed its nuclear genome at low (0.2×) coverage. When analysed in the context of nuclear genomic data representing a broad reference panel of Asian *Bos* species, we find the Bornean banteng affiliates strongly with the Javan banteng, in contradiction to the expectation if the separate species hypothesis was correct. Thus, despite the Bornean banteng's unusual mitochondrial lineage, we argue there is no genomic evidence that the Bornean banteng is a distinct species.

Keywords: Bornean banteng, Genomics, Phylogeny, Taxonomy

Introduction

The wild Asian *Bos* complex consists of three recognised species in Southeast Asia, the banteng (*B. javanicus*), the gaur (*Bos gaurus*), and the kouprey (*Bos sauveli*) that likely went extinct in the twentieth century [1]. The subspecies level structure and diversity of this group has previously only been characterised using mitochondrial DNA, revealing clear paraphyletic structure, with the Bornean banteng mitochondrial lineage nesting inside the diversity of the gaur [2]. Based on this observation, it has been suggested that the Bornean banteng may not be a banteng at all, but rather a distinct species in the wild Asian *Bos* complex [2]. Mitochondrial DNA based phylogenetic signals are, however, not always concordant with

those generated from autosomal nuclear DNA [3]. Therefore, we elected to leverage on the fact that the mitochondrial genome in question was generated through genome skimming of shotgun sequence generated from the tooth of a banteng bull collected in 2010 in Malua, Sabah, Borneo, thus un-analysed nuclear DNA data also exists for the specimen in the public record. We therefore mapped this unused fraction of the original genome skim data to the water buffalo reference genome, enabling us to recover the autosomal genome of the Bornean banteng at an average depth of 0.2×. In this communication we put this partial genome into a phylogenetic context.

Materials and methods

We downloaded the Borneo banteng shotgun genome skim dataset (NCBI: SRA: DRA00172) generated by [2], computationally processed it through the pipeline recently used to generate and investigate the first kouprey genomes [4] along with wild Asian *Bos* genome

*Correspondence: mhssinding@gmail.com

³ Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 1350 Copenhagen, Denmark
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article 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 licence, and indicate if changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

wide variation. Specifically, reads were mapped using the PALEOMIX pipeline v1.2.13.2 [5]. Low quality and missing bases were trimmed from the reads with default settings, and adaptors, dimers and sequences shorter than 25 bp were removed using AdapterRemoval 2.2.4 [6]. To avoid biases associated with mapping to an ingroup [7] in downstream ancestry analyses, retained reads were mapped against scaffolds above 50,000 bp of the outgroup—de novo Water buffalo (*Bubalus bubalis*) UMD_CASPUR_WB_2.0 [8]. The reads were then aligned to the reference using bwa-aln v0.7.16a algorithm [9] including minimum base mapping quality to optimise the initial coverage. Filters targeting mapping and base quality were added at a later stage as appropriate in the specific analysis. Next, Picard MarkDuplicates v2.18.0 (<http://broadinstitute.github.io/picard/>) was used to filter for PCR and optical duplicates. Finally, GATK v4.1.0.0 [10, 11] was used to perform the indel realignment step with no external indel database. All analyses were performed using a set of reference alignments published in [4]. The specific analyses performed was a mitochondrial neighbour-joining phylogeny made using MEGA 10 [12] with 500 bootstraps. ANGSD v0.921 [13] was used to estimate the genotype likelihood at the varying sites for scaffolds with a length above 10 kbp. The resulting output in beagle format was subsequently used as input in PCAngsd v0.973 to perform a genotype likelihood based principal component analysis (PCA). The covariance matrix was used as input in Rstudio [14] to calculate eigenstrat and eigenvalues and plot the components using ggplot2 [15]. We randomly selected 1000 sequences of 5000 bp to construct gene trees using RAxML [16], which were then used to build the species tree. A nuclear genome phylogeny (species tree) was made in ASTRAL-III [17], with DiscoVista being implemented to evaluate the support of alternative topologies [18]. All analyses were performed using the same parameters as detailed in [4].

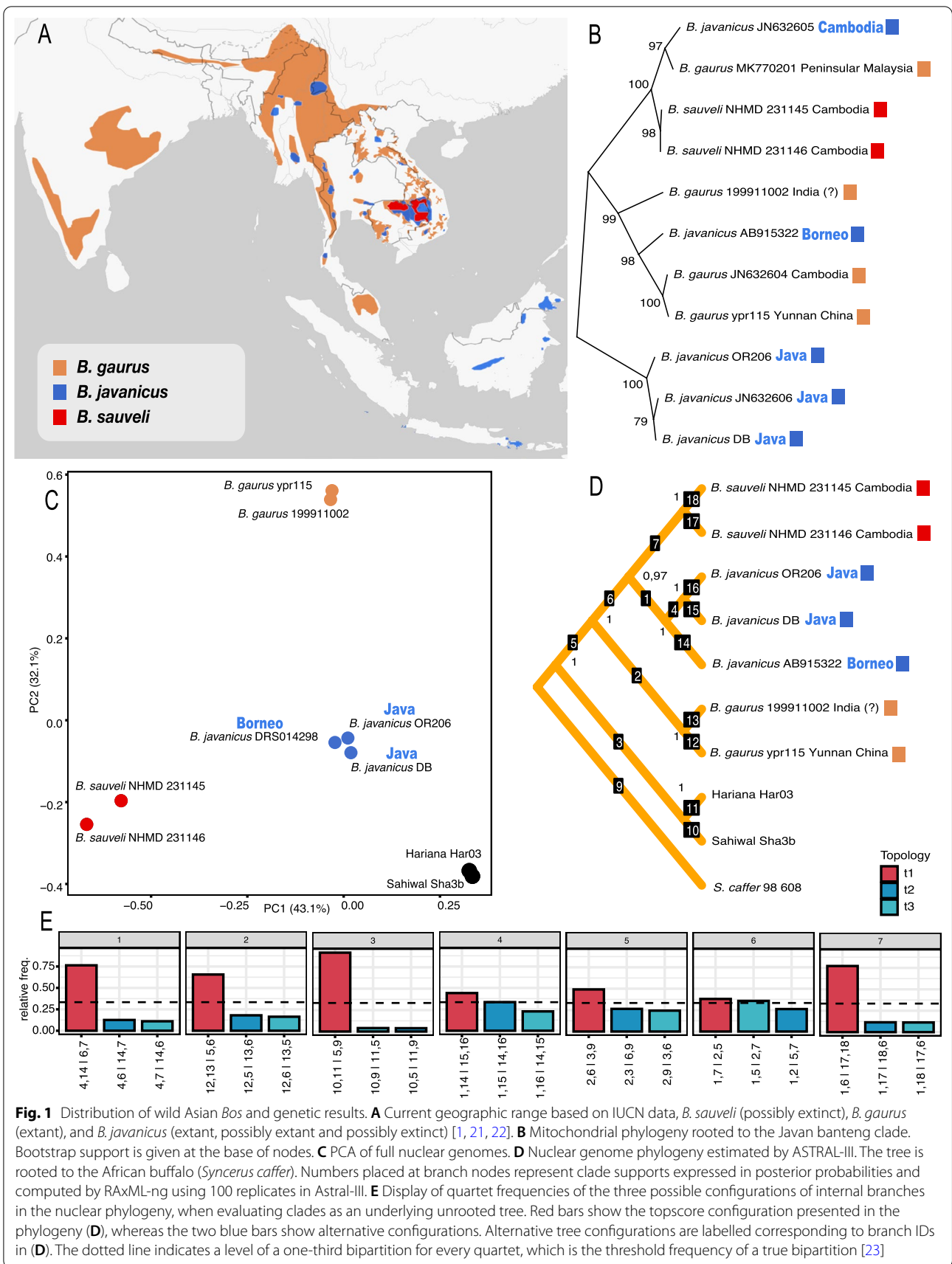
Results and discussion

Today the species level distribution of banteng (Fig. 1A) exists as three geographically isolated groups (one on the Southeast Asian mainland, a second on Java and a third on Borneo), which correspond with the three described banteng subspecies. Consistent with the prior mitochondrial-based analysis of the Bornean subspecies, we replicate the paraphyletic mitochondrial structure of banteng (Fig. 1B), finding that the Bornean banteng lineage is nested within that of the gaur. To explore the nuclear genomic affiliation of the Bornean banteng, we generated a PCA based on two gaur, two kouprey, two zebu cattle (*Bos taurus indicus*) and two captive banteng of Javan origin (based on both zoo records, also supported by the mitochondrial data). The results show clustering of all

the distinct species, including notably, the Bornean banteng with the Javan bantengs (Fig. 1C). To further evaluate a phylogenetic affiliation at the autosomal level, we used the same genomic dataset to create an ASTRAL-III nuclear phylogeny, allowing scaffold-based evaluation of the support for alternative topologies using DiscoVista. Again all species form individual clusters, supported by the highest posterior probability of 1, with the Bornean banteng robustly placed on a branch as sister to the two Javan banteng (Fig. 1D). Additionally, despite the limited amount of data available for performing the analyses (constrained by the, 0.2× coverage of the Bornean banteng), the DiscoVista results (Fig. 1E) replicate previous findings of extensive incomplete lineage sorting (ILS) at the diversification of wild Asian *Bos* [4]. This is compatible with a polytomic origin of the three species [4], we hypothesise may explain the inconsistency of the mitochondrial versus the nuclear phylogenetic affiliation. A comparable scenario is seen in ILS in bison, where European wisent (*Bison bonasus*) and American bison (*Bison bison*) carry highly divergent mitochondrial lineages [19]. As a result American bison mitochondria form a sister lineage to yak (*Bos mutus*), while wisent mitochondria form a sister lineage to cattle (*Bos taurus*). However at autosomal level European wisent and American bison are unambiguously related and monophyletic.

Should ILS not be the explanation, one alternative explanation could be that low levels of hybridisation might have led to fixation of an exotic mitochondrial lineage in either (or both) of the insular banteng groups. Potentially low levels of gaur admixture might have occurred during the formation of the Bornean banteng subspecies either prior to, or during, the isolation of Borneo after the last glacial maximum [20]. Alternatively we speculate that the Javan banteng could potentially have inherited the unique mitochondrial lineage from an unknown, now extinct population. These are all interesting scenarios, that research including ancient genomes may clarify in the future.

However, regardless of the origin of the banteng's paraphyletic mitochondrial structure, the 0.2× nuclear genome clearly supports affiliation of the Bornean banteng with the banteng species. Despite this, we highlight that these results by no means decrease the conservation value of the Bornean banteng, but rather allow focus and coherence by clarifying species affiliation. According to IUCN, the current banteng species population is decreasing in size, and consists of between 4000 and 8000 mature individuals that are isolated, dispersed and as a whole endangered [21]. What fraction of these are Bornean is largely unknown—while probably no more than 350 individuals live in Sabah, only very small fragmented populations exist in Kalimantan, and likely very



few—if any—persist in Sarawak [21]. Ultimately therefore we hope that our genomic characterization of the Bornean banteng population might contribute to both future attempts to clarify its evolutionary history, as well as guide future conservation efforts.

Acknowledgements

The authors would like to thank the authors of [2] for making the raw sequencing reads behind their mitochondrial de novo genome available, hence allowing this autosomal investigation.

Author contributions

The authors XS, MTPG and MHSS contributed to the study conception and design. Material preparation, data collection and analysis were performed by XS, MMC, JAR and MHSS. The first draft of the manuscript was written by MHSS and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work and MHSS was supported by the Carlsberg Foundation (Grant number: CF20-0355). XS was supported by the Norwegian Environment Agency under the project number 2018/2326 granted to MTPG.

Availability of data and materials

All data analysed in this study is publicly available and previously published, the NCBI affiliation of each specimen is given in Fig. 1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹University Museum, NTNU, 7491 Trondheim, Norway. ²Center for Evolutionary Hologenomics, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1353 Copenhagen, Denmark. ³Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 1350 Copenhagen, Denmark.

Received: 18 July 2022 Accepted: 9 September 2022

Published online: 20 September 2022

References

- Timmins RJ, Burton J, Hedges S. *Bos sauveli*. The IUCN Red List of Threatened Species 2016. 2016;e.T2890A46363360. <https://doi.org/10.2305/IUCN.UK.2016-2.RLTS.T2890A46363360.en>.
- Ishige T, Gakuhari T, Hanzawa K, Kono T, Sunjoto I, Sukor JRA, et al. Complete mitochondrial genomes of the tooth of a poached Bornean banteng (*Bos javanicus lowii*; Cetartiodactyla, Bovidae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016;27(4):2453–4. <https://doi.org/10.3109/19401736.2015.1033694>.
- Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*. 2019;364(6446):eaav6202. <https://doi.org/10.1126/science.aav6202>.
- Sinding MHS, Ciucani MM, Ramos-Madriral J, Carmagnini A, Rasmussen JA, Feng S, et al. Kouprey (*Bos sauveli*) genomes unveil polytomic origin of wild Asian Bos. *iScience*. 2021;24(11):103226. <https://doi.org/10.1016/j.isci.2021.103226>.
- Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 2014;9(5):1056–82. <https://doi.org/10.1038/nprot.2014.063>.
- Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016;9:88. <https://doi.org/10.1186/s13104-016-1900-2>.
- Gopalakrishnan S, Castruita JS, Sinding MHS, Kuderna L, Rääkkönen J, Petersen B, et al. The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp, population genomics. *BMC Genomics*. 2017. <https://doi.org/10.1186/s12864-017-3883-3>.
- Williams JL, Iamartino D, Pruitt KD, Sonstegard T, Smith TPL, Low WY, et al. Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2n = 50). *GigaScience*. 2017. <https://doi.org/10.1093/gigascience/gjx088>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–8. <https://doi.org/10.1038/ng.806>.
- Kumar S, Stecher G, Li M, Nnyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547–9. <https://doi.org/10.1093/molbev/msy096>.
- Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform*. 2014;15:356. <https://doi.org/10.1186/s12859-014-0356-4>.
- Team R. RStudio: integrated development for R. RStudio, Boston: PBC. 2020.
- Wickham H. Data analysis. In: Wickham H, editor. *ggplot2: elegant graphics for data analysis*. Cham: Springer International Publishing; 2016. p. 189–201. https://doi.org/10.1007/978-3-319-24277-4_9.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform*. 2018;19(Suppl 6):153. <https://doi.org/10.1186/s12859-018-2129-y>.
- Sayyari E, Whitfield JB, Mirarab S. DiscoVista: interpretable visualizations of gene tree discordance. *Mol Phylogenet Evol*. 2018;122:110–5. <https://doi.org/10.1016/j.ympev.2018.01.019>.
- Wang K, Lenstra JA, Liu L, Hu Q, Ma T, Qiu Q, et al. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun Biol*. 2018;1:169. <https://doi.org/10.1038/s42003-018-0176-6>.
- Hanebuth T, Stattegger K, Grootes PM. Rapid flooding of the sunda shelf: a late-glacial sea-level record. *Science*. 2000;288(5468):1033–5. <https://doi.org/10.1126/science.288.5468.1033>.
- Gardner P, Hedges S, Pudyatmoko S, Gray TNE, Timmins RJ. *Bos javanicus*. The IUCN Red List of Threatened Species 2016. 2016;e.T2888A46362970. <https://doi.org/10.2305/IUCN.UK.2016-2.RLTS.T2888A46362970.en>.
- Duckworth JW, Sankar K, Williams AC, Samba Kumar N, Timmins RJ. *Bos gaurus*. The IUCN Red List of Threatened Species 2016. 2016;e.T2891A46363646. <https://doi.org/10.2305/IUCN.UK.2016-2.RLTS.T2891A46363646.en>.
- Allman ES, Degnan JH, Rhodes JA. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. *J Math Biol*. 2011;62(6):833–62. <https://doi.org/10.1007/s00285-010-0355-7>.

Publisher's Note

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