BMC Genomic Data

DATA NOTE

Open Access

Genome assembly of the *Bos gaurus* in Vietnam using nanopore sequencing

Thi Dieu Thuy Nguyen[1*], Thi Binh Nguyen Tran[2] and Loan T. Nguyen[3*]

Abstract

Objectives The gaur (*Bos gaurus*) is one of the largest wild cattle species in the Bovidae family and is among the most threatened mammals in Vietnam. Listed as vulnerable on the IUCN Red List since 1986, gaurs are found in various provinces across Vietnam, including Lai Chau, Son La, Thanh Hoa, Nghe An, and several others. Their population has drastically declined, with current estimates showing about 30-50 individuals in the Northwest and fewer than 300 in the Central Highlands. Conservation efforts are crucial, and genome assembly is vital, providing valuable information for effective conservation strategies.

Data description Using Oxford Nanopore sequencing technologies, this study generated over 49 Giga-bases of trimmed whole genome sequencing reads from the muscle tissue of a deceased gaur collected in 2014 in Quang Nam, Vietnam. These reads were assembled into a 2.61 Gb draft genome, comprising 32,004 contigs, 54 scaffolds, with an N50 of 324.5 kilobases and a GC content of 41.86%. The BUSCO completeness value, using the mammalia_odb10 dataset, was 83.4%. This Vietnamese gaur genome assembly will serve as a valuable resource for conservation genetic research in this species.

Keywords *Bos gaurus*, Nanopore sequencing, Vietnam, Genome assembly, Long reads

Objective

The gaur (*Bos gaurus*), one of the largest wild cattle species in the Bovidae family, is among the most threatened mammals. This species has been listed as vulnerable on the IUCN Red List since 1986 [1, 2]. Wild gaurs inhabit southern and southeastern Asia, including India, Nepal, Bhutan, Bangladesh, Myanmar, Thailand, China, Laos, Cambodia, Vietnam, and Malaysia [2–5]. In Vietnam, gaurs are found in various provinces such as Lai Chau, Son La, Thanh Hoa, and Nghe An [2, 6]. Their population has drastically declined, with current estimates showing about 30–50 individuals in the Northwest and fewer than 300 in the Central Highlands [6].

Studies on the Vietnamese gaur genome have primarily focused on mitochondrial DNA. A 2023 study examined the genetic relationship of Vietnamese gaurs from the Central Highlands with other gaurs from Southeast Asia and India, indicating the closest genetic proximity of Vietnamese gaurs to *Bos frontalis* (0.0086 ± 0.0022) and *Bos gaurus* (0.0182 ± 0.0054) [6].

The gaur has unique genetic features not found in domesticated cattle, such as enhanced resistance to biting arthropods, attributed to bovidic acid secretion [7, 8]. This hydroxyfuranoid fatty acid could be a safer

*Correspondence:
Thi Dieu Thuy Nguyen
ntdthuy@ibt.ac.vn
Loan T. Nguyen
t.nguyen3@uq.edu.au
[1]Institute of Biology, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[2]Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi, Vietnam
[3]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	ONT sequence reads of *B. gaurus* genomic DNA	fastq file (.fastq)	NCBI SRA Database Accession number SRR29504129 https://identifiers.org/ncbi/insdc.sra:SRR29504129 [10]
Data file 2	Genome assembly of *B. gaurus*	fasta file (.fasta)	EMBL-EBI ENA Accession number GCA_965225615 https://identifiers.org/insdc.gca:GCA_965225615 [11]

alternative to DEET-based repellents, but it has not been commercialized.

To date, only one genome assembly for *Bos gaurus* has been published on NCBI, from the USDA. A genome assembly from Vietnamese gaurs could serve as a reference for non-domesticated cattle. With a divergence of approximately 4.5 million years from taurine cattle, this assembly could provide valuable information for effective conservation strategies, insights into natural and artificial selection, and identifying genes related to traits of interest.

Data description

Samples were collected from a deceased wild gaur (*Bos gaurus*) in Quang Nam, Vietnam, in 2014. The carcass was transported to the Vietnam Museum of Nature in a refrigerated vehicle for conservation purposes, with the intention that the samples could be used for scientific studies. Muscle tissue was sampled approximately 48 h post-mortem. After being stored in ethanol at -20 °C for over five years, the samples were processed for extraction and sequencing.

Total DNA was extracted from the muscle sample using the Gentra Puregene Tissue kit (QIAGEN, Germany), following the manufacturer's instructions. DNA quantity and quality were assessed using a Qubit™ 4 Fluorometer (Thermo Scientific, USA) and a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE).

Sequencing libraries were prepared using the Ligation Sequencing Kit v10 (SQK-LSK110) from Oxford Nanopore Technologies (ONT, UK) with modifications as described in Hayes et al. [9], and sequenced on a PromethION device with R9 flow cells (ONT, UK). Raw signal files (FAST5) were used for base calling using Guppy Basecaller version 6.4.6 in super accuracy mode, resulting in over 55 million reads (Qscore ≥ 10) with an average read length of 2.1 kb (Table 1, Data file 1, [10]).

The sequence reads were processed using Filtlong v0.2.1 to remove any reads under 500 bp, resulting in over 49 gigabases of high-quality reads with an N50 of 2.5 kb. The pre-processed reads were then assembled de novo using Flye (version 2.9) with the parameters --nano-hq for the ONT super accuracy base-call file, --genome-size = 2.7G, and -i 3 for the number of polishing iterations. This resulted in a draft genome assembly of 2.61 Gb, consisting of 32,004 contigs with a contig N50 of 324.5 kb (Table 1, Data file 2, [11]). The draft genome assembly was evaluated with BUSCO, receiving a completeness value of 83.4%.

So far, this will be the second published genome assembly for *Bos gaurus*, with the first being from a healthy female published by the USDA.

Limitations

The current genome assembly is fragmented and has low coverage. This is because the DNA was collected 48 h post-mortem and stored at -20 °C for over 5 years before extraction. Additionally, the extracted DNA remained in the freezer for more than 3 years before sequencing. Higher coverage can be achieved using the same sequencing technology. For scaffolding, the Pore-C method published by Nguyen et al. in 2023 could be utilized [12].

Abbreviations

IUCN International Union for Conservation of Nature
NCBI National Centre for Biotechnology Information
ONT Oxford Nanopore Technologies
USDA United States Department of Agriculture

Acknowledgements

The authors acknowledge the People's Committee of Quang Nam Province for granting permission to the Vietnam Museum of Nature (under the Vietnam Academy of Science and Technology) to use the samples for scientific studies.

Authors' contributions

TDTN and LTN designed the study. TDTN collected the sample. TTBN extracted DNA. LTN performed the sequencing and genome assembly. All authors wrote the manuscript. All authors declare no competing interest.

Data availability

The data described in this Data note can be freely and openly accessed on NCBI SRA under Bioproject ID PRJNA 1127286 [10], and via the EMBL-EBI European Nucleotide Archive (ENA) under accession number GCA_965225615 [11]. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 July 2024 / Accepted: 20 May 2025
Published online: 01 July 2025

References

1. IUCN Red List of Threatened Species. Available from: https://www.iucnredlist.org/species/2891/46363646. Accessed 27 May 2025.
2. Nguyen MH. The status of vulnerable Gaur Bos Gaurus and endangered Banteng Bos javanicus in Ea so nature reserve and Yok don and Cat Tien National parks, Vietnam. Oryx. 2009;43(1):129–35.
3. Ellerman JR, Morrison-Scott TCS. Checklist of Palaearctic and Indian mammals 1758 to 1946. BM(NH). 1951. pp. 1–810. https://www.biodiversitylibrary.org/part/92546.
4. Corbet GB, Hill JE. The mammals of the Indomalayan region: a systematic review. Oxford: Oxford university press; 1992.
5. Mohanarangan A, Distribution. Ecology and Conservation of the Gaur (Bos gaurus, H. Smith, 1824). In Ambawat RS, Rahmani AR, editors. Animal diversity, natural history and conservation. 2011. pp. 77–94.
6. Nguyen Thi Phuong Mai, Diem NTH, Tung HT, Long LT. The genetic relationship of Vietnamese Gaurs assessed by MtDNA sequences. J Entomol Zool Stud. 2023;11(1):7–12.
7. Ishii H, Krane S, Itagaki Y, Berova N, Nakanishi K, Weldon PJ. Absolute configuration of a hydroxyfuranoid acid from the pelage of the genus Bos, 18-(6 S, 9 R, 10 R)-bovidic acid. J Nat Prod. 2004;67(8):1426–30.
8. Tran K, Chauhan K. Structural activity of Bovidic acid and related compounds as feeding deterrents against Aedes aegypti. Biopesticides Int. 2007;3(1):53–7.
9. Hayes BJ, Nguyen LT, Forutan M, Engle BN, Lamb HJ, Copley JP, et al. An epigenetic aging clock for cattle using portable sequencing technology. Front Genet. 2021;12:760450.
10. NCBI Sequence Read Archive. https://identifiers.org/ncbi/insdc.sra:SRR29504129; 2024.
11. Nguyen TDT, Tran TBN, Nguyen LT. Genome assembly of the Bos gaurus in Vietnam using nanopore sequencing. EMBL-EBI European Nucleotide Archive, Accession:GCA_965225615; 2025. Available at https://identifiers.org/insdc.gca:GCA_965225615.
12. Nguyen LT, Jung H, Ma J, Andersen S, Ross E. Long-read Pore-C shows the 3D structure of the cattle genome. Anim Prod Sci. 2023;63(11):972–82.

Publisher's Note