DATA NOTE

Open Access

Genome sequencing of captive white tigers from Bangladesh

Ashutosh Das[1*], Md Shahadat Hossain Suvo[2], Mishuk Shaha[1] and Mukta Das Gupta[3]

Abstract

Objectives The Bengal tiger *Panthera tigris tigris*, is an emblematic animal for Bangladesh. Despite being the apex predator in the wild, their number is decreasing due to anthropogenic activities such as hunting, urbanization, expansion of agriculture and deforestation. By contrast, captive tigers are flourishing due to practical conservation efforts. Breeding within the small captive population can produce inbreeding depression and genetic bottlenecks, which may limit the success of conservation efforts. Despite past decades of research, a comprehensive database on genetic variation in the captive and wild Bengal tigers in Bangladesh still needs to be included. Therefore, this research aimed to investigate the White Bengal tiger genome to create a resource for future studies to understand variation underlying important functional traits.

Data description Blood samples from Chattogram Zoo were collected for three white Bengal tigers. Genomic DNA for all collected samples were extracted using a commercial DNA extraction kit. Whole genome sequencing was performed using a DNBseq platform. We generated 77 Gb of whole-genome sequencing (WGS) data for three white Bengal tigers (Average 11X coverage/sample). The data we generated will establish a paradigm for tiger research in Bangladesh by providing a genomic resource for future functional studies on the Bengal white tiger.

Keywords White Bengal tigers, Whole genome sequencing

Objective

The critically endangered Bengal tiger, *Panthera tigris tigris*, is a native subspecies of the Indian subcontinent. The Bengal tiger population in India started to drop over a century ago. By 1970, less than 2,000 tigers remained in the wild, similar to the global tiger population reduction. Approximately 2,900 wild tigers remain in Indian reserves, making up over 60% of the total number of wild tigers worldwide [1]. According to estimates from the Bengal Tiger Conservation Activity (BAGH) project, there were 114 tigers in the Bangladesh Sundarbans [2]. Despite several conservation efforts, numerous factors, including habitat loss, deforestation, altered land cover, human disturbance of the forests, poaching, hunting, illegal wildlife trade, climate change, natural disasters and inadequate legal frameworks [3], are contributing to the extinction of the tiger population in the Sundarbans.

By contrast, captive tigers are flourishing. Appropriately maintained captive populations of wild animals have been shown to represent a "genetic reservoir" of their natural counterparts, providing insurance against extinction in the wild and aiding in public education, research, and fundraising. Small, isolated populations that experience inbreeding have minimal genetic variety among their individuals and are very vulnerable to extinction. According to estimates, the Bengal tiger population

*Correspondence:
Ashutosh Das
ashutosh.das@cvasu.ac.bd
[1] Department of Genetics and Animal Breeding, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University, Khulshi Chattogram-4225, Bangladesh
[2] Chattogram Zoo, Akbar Shah Chattogram-4202, Bangladesh
[3] Department of Microbiology and Veterinary Public Health, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University, Khulshi Chattogram-4225, Bangladesh

Das *et al. BMC Genomic Data*   (2024) 25:52

Page 2 of 4

possesses the most genetic diversity, making it the ideal gene pool reservoir for conservation efforts [4, 5]. Using available genomic resources, a high-quality reference genome for Bengal tigers [6] and other tiger genomic data [5, 7–10], we can conduct a comparative genomic analysis and determine the genetic diversity of Bengal tigers. Therefore, we generated this data to compile more

comprehensive genomic information, which will be helpful for future research into the variants causing significant colour phenotypes.

Data description.

Following the ethical rules and procedures of Chattogram Zoo Bangladesh, blood samples were taken from three white tiger cubs of three months' age, including one
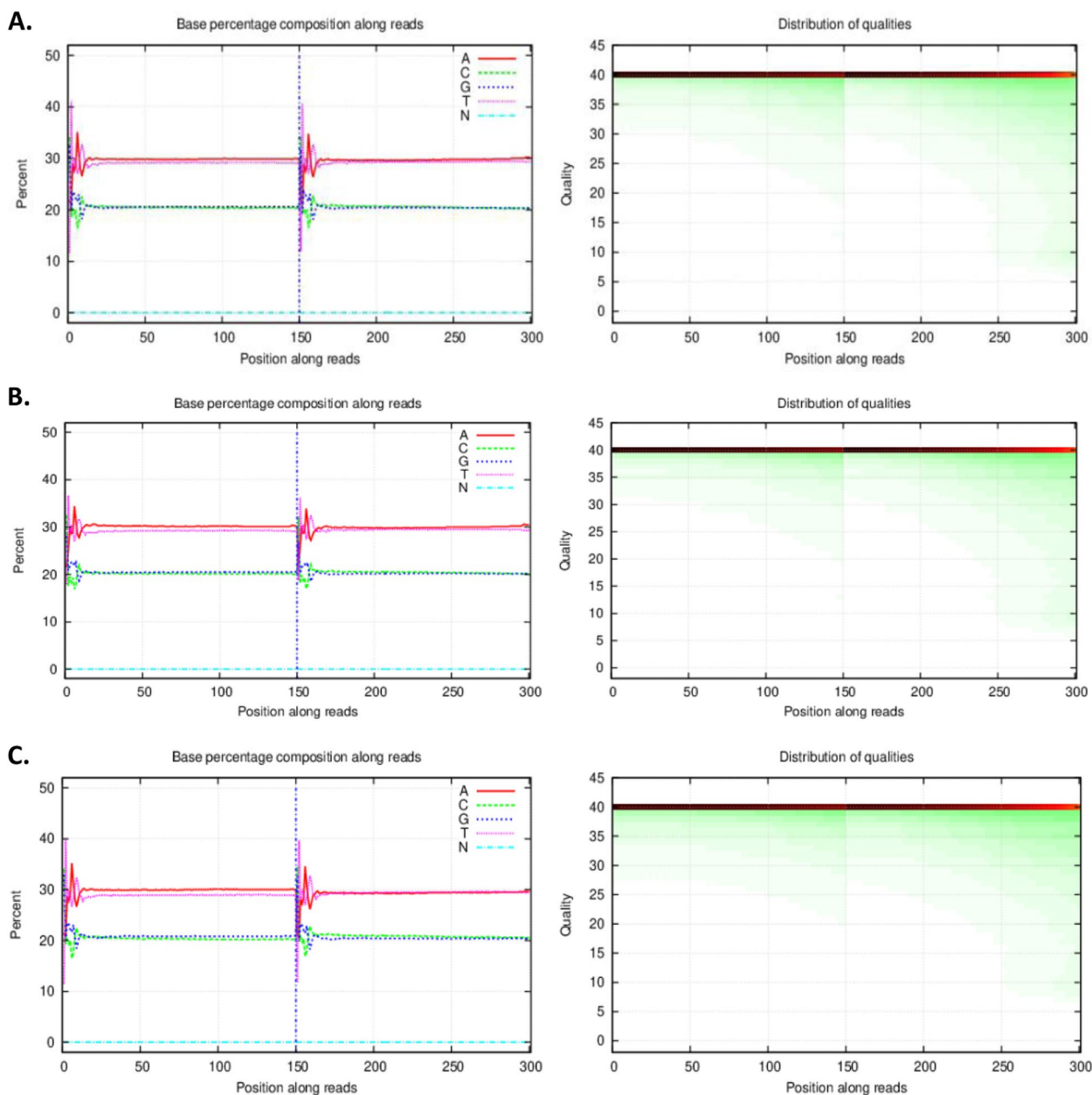

Fig. 1 The distribution of base percentage and qualities along reads. In the left figures, x-axis represents base position along reads, y-axis represents base percentage at the position; each color represents a type of nucleotide. Under normal conditions, the sample does not have AT/GC separation. It is normal to see fluctuations in the first several bp positions, which is caused by random primer and the instability of enzyme–substrate binding at the beginning of the sequencing reaction. In the right figures, x-axis represents base position along reads, y-axis represents base quality; each dot represents the base quality of the corresponding position along reads, color intensity reflects the number of nucleotides, a more intense color along a quality value indicates a higher proportion of this quality in the sequencing data. **A**, **B** and **C** represent for sample no. 1, 2 and 3 respectively

Das *et al. BMC Genomic Data*   (2024) 25:52

Page 3 of 4

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)
Figure 1	The distribution of base percentage and qualities along reads	Document file (.pdf)	Figshare, https://doi.org/10.6084/m9.figshare.24996869 [13]
Data file 1	WGS data of white Bengal tiger sample 1	SRA file (.fastq.gz)	NCBI Sequence Read Archive https://identifiers.org/ncbi/insdc.sra:SRR24305815 [14]
Data file 2	WGS data of white Bengal tiger sample 2	SRA file (.fastq.gz)	NCBI Sequence Read Archive https://identifiers.org/ncbi/insdc.sra:SRR24459545 [15]
Data file 3	WGS data of white Bengal tiger sample 3	SRA file (.fastq.gz)	NCBI Sequence Read Archive https://identifiers.org/ncbi/insdc.sra:SRR24632529 [16]
Table 2	Basic statistics of whole genome sequence data for captive white tiger from Bangladesh	Document file (.docx)	Figshare, https://doi.org/10.6084/m9.figshare.25902517.v1 [17]

Table 2 Basic statistics of whole genome sequence data for captive white tiger from Bangladesh

Sample Name	Clean Reads	Clean Base	Read Length	Q20(%)	Q30(%)	GC(%)
WBT_01	147,848,227	44,354,468,100	PE150	96.92	92.55	41.15
WBT_02	145,372,272	43,611,681,600	PE150	97.07	93.01	40.72
WBT_03	88,031,025	26,409,307,500	PE150	97.29	92.55	41.35

WBT White Bengal tiger

female and two male cubs. Blood samples were collected aseptically from the cephalic vein using sterile butterfly needles. Blood sample were placed in Vacutainer tubes containing ethylene diamine tetraacetic acid (EDTA) as the anticoagulant. Total genomic DNA was extracted from blood samples using Monarch Genomic DNA Purification Kit (New England Biolabs, UK) according to the manufacturer's guidelines. Thermo Scientific, USA's NanoDropTM One Microvolume UV–Vis Spectrophotometer was used to evaluate the extracted DNA's quality and purity. All samples shown a decent purity with a 160/280 values ranged from 2.06–2.38. For sequencing and library construction (Short Insert library), purified genomic DNA was transferred to Beijing Genomics Institute (BGI, Hong Kong). The DNBSEQ Short-read library preparation instructions provided by the manufacturer were followed for the development of the sequencing libraries. We used a DNBseq platform to do whole genome sequencing(WGS).

High-performance computing resources were used for WGS bioinformatics. Low-quality raw paired readings were removed using SOAPnuke [11] after the raw reads were assessed for quality. In a nutshell, low-quality or adapter sequences in the raw data were filtered first. Many data processing steps were taken to get rid of contaminants and provide reliable data. The filter parameters for the SOAPnuke program were "-n 0.001 -l 10 –adaMR 0.25". The filtering steps were 1) Filter adapter: delete the whole read if the sequencing read matches 25.0% or more of the adapter sequence (a maximum of two base mismatches is permitted); 2) Filter low-quality data: remove the whole sequencing read if bases with a quality value of less than 10 make up at least 50.0% of the read; 3) Eliminate N: Delete the whole read and discard any N information that makes up 0.1% or more of the sequencing read; 4) To obtain clean readings, Phred + 33 was set as the output read quality value for the system. The quality of data was examined after filtering. Base percentage compositions showed all sequenced samples had high-quality data after filtering (Fig. 1) [13]. Burrows-Wheeler Aligner (BWA) software [12] was used to align high-quality reads to the reference *Panthera tigris tigris* genomes, the PanTigT.SI.v4 [6], using the default BWA mem settings.

For three white Bengal tigers that were sequenced, we produced 77 Gb of data (Table 1, Data file 1, 2 and 3) [14–16]. The average genome coverage was 11X. A description of the clean data is shown in Table 2 [17]. Mapping reads encompass 2363074012 base pairs of

Das *et al. BMC Genomic Data*   (2024) 25:52

Page 4 of 4

WGS data which covered 98.46% of the reference tiger genome in the current investigation.

Limitations.

Since all the samples come from individuals from the same parent, performing a genome-wide association study to identify genomic regions associated with a particular phenotype were not possible.

Abbreviations
BWA   Burrows-Wheeler Aligner
DNA   Deoxyribonucleic acid
Gb   Gigabase
WGS   Whole genome sequencing

Availability of data and materials
Figure 1 and Table 2 described in this Data Note can be freely and openly accessed on FigShare (https://figshare.com/) [13, 17]. The data described in this Data note can be freely and openly accessed from the NCBI Bioproject PRJNA961947. Raw data have been deposited separately in the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) with open accession ID SRP434520 [14–16].

Declarations

Ethics approval
All animal procedures including collecting blood samples from tigers were approved by the research ethics committee of Chattogram Veterinary and Animal Sciences University (Memo no. – CVASU/DIR(R&E)EC/2024/688/1/13).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Jhala YV, Qureshi Q, Nayak AK. Status of Tigers, Copredators and Prey in India, 2018. New Delhi: National Tiger Conservation Authority, Government of India; 2020.
2. Aziz MA, Kabir MJ, Shamsuddoha M, Ahsan MM, Sharma S, Chakma S, Second phase status of tigers in Bangladesh Sundarban, et al. Department of Zoology. Bangladesh: Janhangirnagar University; WildTeam Bangladesh; Forest Department; 2018.
3. Islam MZ. A Reality Check of the Global TX2 Goals of Doubling the Bengal Tiger (Panthera tigris tigris) Population by 2022 in the Sundarbans Mangrove Forest. Contemp Probl Ecol. 2023;16(6):868–85.
4. Armstrong EE, Khan A, Taylor RW, Gouy A, Greenbaum G, Thiéry A, et al. Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. Mol Biol Evol. 2021;38(6):2366–79.
5. Luo SJ, Liu YC, Xu X. Tigers of the world: Genomics and conservation. Annual review of animal biosciences. 2019;15(7):521–48.
6. Shukla H, Suryamohan K, Khan A, Mohan K, Perumal RC, Mathew OK, Menon R, Dixon MD, Muraleedharan M, Kuriakose B, Michael S. Near-chromosomal de novo assembly of Bengal tiger genome reveals genetic hallmarks of apex predation. GigaScience. 2023;12:giac112.
7. Zhang L, Lan T, Lin C, Fu W, Yuan Y, Lin K, et al. Chromosome-scale genomes reveal genomic consequences of inbreeding in the South China tiger: A comparative study with the Amur tiger. Mol Ecol Resour. 2022;23:330–47. https://doi.org/10.1111/1755-0998.13669.
8. Khan A, Patel K, Shukla H, et al. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proc Natl Acad Sci. 2021;118(49): e2023018118.
9. Liu YC, Sun X, Driscoll C, et al. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. Curr Biol. 2018;28(23):3840–9.
10. Cho YS, Hu L, Hou H, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun. 2013;4(1):2433.
11. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7(1):gix120.
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. bioinform. 2009;25(14):1754–60.
13. Das A, Suvo MSH, Shaha M, Gupta MD. Genome sequencing of captive white tiger from Bangladesh. 2024. Figshare. https://doi.org/10.6084/m9.figshare.24996869.
14. Das A, Suvo MSH, Shaha M, Gupta MD. Genome sequencing of captive white tiger from Bangladesh. NCBI Sequence Read Archive. 2024. https://identifiers.org/ncbi/insdc.sra:SRR24305815
15. Das A, Suvo MSH, Shaha M, Gupta MD. Genome sequencing of captive white tiger from Bangladesh. NCBI Sequence Read Archive. 2024. https://identifiers.org/ncbi/insdc.sra:SRR24459545
16. Das A, Suvo MSH, Shaha M, Gupta MD. Genome sequencing of captive white tiger from Bangladesh. NCBI Sequence Read Archive. 2024. https://identifiers.org/ncbi/insdc.sra:SRR24632529
17. Das A, Suvo MSH, Shaha M, Gupta MD. Genome sequencing of captive white tiger from Bangladesh. 2024. Figshare. https://doi.org/10.6084/m9.figshare.25902517.v1.

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