DATA NOTE

Open Access

The genome of the Black Bengal goat (*Capra hircus*)

Amam Zonaed Siddiki[1*] , Abdul Baten[2,7], Masum Billah[1], Mohammad Atique Ul Alam[1], Kazi Shefaul Mulk Shawrob[1], Sourav Saha[1], Muntaha Chowdhury[1,6], Atif Hasan Rahman[3], Michael Stear[4], Gous Miah[5], Mahadia Kumkum[1], Mohammad Sirazul Islam[1], Mohammad Alamgir Hossain[1], A. K. M. Moniruzzaman Mollah[6] and Md. Kabirul Islam Khan[5]

Abstract

Objectives: Black Bengal goat (*Capra hircus*), a member of the Bovidae family with the unique traits of high prolificacy, skin quality and low demand for food is the most socioeconomically significant goat breed in Bangladesh. Furthermore, the aptitude of adaptation and disease resistance capacity of it is highly notable which makes its whole genome information an area of research interest.

Data description: The genomic DNA of a local (Chattogram, Bangladesh) healthy male Black Bengal goat (*Capra hircus*) was extracted and then sequenced. Sequencing was completed using the Illumina HiSeq 2500 sequencing platform and the draft assembly was generated using the "ARS1" genome as the reference. MAKER gene annotation pipeline was utilized to annotate 26,458 gene models. Genome completeness was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) which showed 82.5% completeness of the assembled genome.

Keywords: Black Bengal goat, *Capra hircus*, Whole genome sequencing, Assembly, Annotation

Objective

Black Bengal goat (BBG) belongs to the Bovidae family and found throughout Bangladesh, West Bengal, Bihar, and Orissa regions of northeastern India. It is estimated that more than 90% of the goat population in Bangladesh comprised the Black Bengal, the remainder being Jamunapari and their crosses [1]. Higher prolificacy, fertility, resistance against common diseases, adaptability to the adverse environmental condition, early maturity, seasonality and superiority in the litter size are some of the outstanding features of BBG. Besides, it produces excellent quality flavored, tender and delicious meat with low intramuscular fat and fine skin of extraordinary quality for which there is tremendous demand all over the world [1, 2]. Moreover, it plays a vital role in the economy of

Bangladesh by contributing 1.66% of the GDP (Gross Domestic Product) (DLS 2017).

Fortunately, the market demand of Black Bengal goat is emerging. This gives breeders of original/rare breeds an opportunity to expand the stock and preserve its genetic diversity. One of the primary goals in managing goat populations is to maintain high-level genetic diversity and low-level inbreeding. To estimate the future breeding potential of a goat breed, it is necessary to characterize the genetic structure and evaluate the level of genetic diversity within the breed. Moreover, a long term genetic approach can be used to improve the spectacular economic characteristics of BBG [3].

Therefore, the genetic characterization of the entire BBG genome is essential in characterizing its economic traits as well as adaptive capability. With the availability of whole genome sequence, the targeted areas for genetic improvements are now: goat prolificacy, growth rate, meat quality, skin quality, disease resistance, and survivability. A complete and accurate reference to the goat

*Correspondence: zsiddiki@cvasu.ac.bd
[1] Genomics Research Group, Department of Pathology and Parasitology, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University (CVASU), Chattogram 4225, Bangladesh
Full list of author information is available at the end of the article

Siddiki *et al. BMC Res Notes*   (2019) 12:362

Page 2 of 3

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	DNA isolation and library preparation methodology	.docx	https://figshare.com/s/67c8971f5a69277d55d0
Data file 2	Whole genome assembly data	FASTA	NCBI GeneBank (Accession numbers:GCA_004361675.1)(https://www.ncbi.nlm.nih.gov/assembly/GCA_004361675.1)
Data set 3	Whole genome sequence	FASTA	NCBI GeneBank (Accession numbers SMSF01000001- SMSF01003972) (https://www.ncbi.nlm.nih.gov/nuccore/SMSF00000000.1)
Data set 4	Annotation files	.tsv	https://figshare.com/s/a7873490048898ac1d56

genome is an essential component of advanced genomic selection of product characteristics.

Data description
At first, A 3 years old male healthy Black Bengal goat (BBG) without known genetic diseases was selected for blood collection. Genomic DNA from each animal was isolated from the EDTA-blood, using the Addprep genomic DNA extraction kit (South Korea) (detailed methodology in Data file 1—Table 1). The quality and quantity of the DNA were assessed by the Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) and Infinite F200 microplate reader (TECAN), according to the manufacturer's instruction. The status of the DNA was visually inspected by 0.8% agarose gel electrophoresis. Purified genomic DNA was sent for library preparation (detailed methodology in Data file 1—Table 1) and whole genome sequencing (WGS) at BGI Group (Shenzhen, Guangdong, China). A total of 40 Gb (Gigabase pair) (14-fold) of subread bases with a read length of 150 bp were generated using next-generation sequencing (NGS) technology on an Illumina HiSeq 2500 platform (detailed methodology in Data file 1—Table 1).

After sequencing, quality of the raw sequencing reads were inspected using FastQC version 0.11.8 [4]. Reads were quality controlled including removing adaptor sequences, contamination and low-quality reads from raw reads using Trimmomatic V0.32 [5]. A total of 247,325,362 clean reads were included in the assembly. Subsequently, for de novo assembly we used ABySS v. 2.1.5 assembler [6], which generated 32,94,295 contigs (minimum contig size 200 bp). Next, ABACAS v.1.3.1 pipeline was used with the reference genome ARS1 (GCA_001704415.1) [7] to arranging, ordering, and orientation of the assembled genome [8]. The genome assembly data has been deposited in the NCBI GenBank under the Accession number GCA_001704415.1 (Data file 2—Table 1). The final assembled genome size of BBG is 3.04 Gb with 724.80 Mb (Megabase pair) gaps and GC content of 41.77%. Completeness of the genome was assessed with benchmarking universal single-copy orthologs (BUSCO) version 3.0.2 [9] which showed 82.5% completeness.

Genes were annotated using Maker version 3.0 pipeline [10] which identified 26,458 gene models. RepeatMasker V 4.0.9 [11] using the latest version of the repbase database [12] identified 31.85% repeat elements in the genome. Finally, InterProScan V 5.33–72.0 [13] was used to identify the gene ontology (GO) terms, which identified a total of 12,589 GO terms and 8173 genes have at least 1 associated GO term. The whole genome sequence data has been submitted in the NCBI GenBank under the Accession numbers SMSF01000001–SMSF01003972 (Data file 3—Table 1).

Limitations
The number of unassembled regions in the genome is 3943 and the total number of bases placed in this gap is 724,808,570 bp.

Siddiki *et al. BMC Res Notes*   (2019) 12:362

Page 3 of 3

Availability of data materials

The genome sequence information has been accessible at DDBJ/ENA/Gen-Bank under the Accession Numbers SMSF01000001–SMSF01003972 and the assembled genome at GCA_001704415.1. The version reported in this paper is the first version, SMSF00000000.1.

Ethics approval and consent to participate

The experiments discussed in this investigation were approved by the Institute Review Committee of Chattogram Veterinary and Animal Sciences University.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Genomics Research Group, Department of Pathology and Parasitology, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University (CVASU), Chattogram 4225, Bangladesh. [2] AgResearch, Private Bag 11008, Palmerston North 4410, New Zealand. [3] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh. [4] Department of Animal, Plant and Soil Sciences, School of Life Sciences, AgriBio, La Trobe University, Bundoora, VIC 3083, Australia. [5] Department of Genetics and Animal Breeding, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University (CVASU), Chattogram 4225, Bangladesh. [6] Department of Biological Sciences, Asian University for Women (AUW), Chattogram 4000, Bangladesh. [7] Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia.

References

1. Husain SS. A study on the productive performance and genetic potentials of Black Bengal goats. A Ph.D. Thesis, Bangladesh Agricultural University, Mymensingh. 1993.
2. Islam M, Nahar TN, Haq S. Prospect of goat production in Bangladesh. Asian Livestock (FAO). 1991.
3. Faruque S, Chowdhury SA, Siddiquee NU, Afroz MA. Performance and genetic parameters of economically important traits of Black Bengal goat. J Bangladesh Agric Univ. 2010;8(1):67–78.
4. FastQC program. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 12 Jan 2017.
5. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
6. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19(6):1117–23.
7. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017;49(4):643.
8. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics. 2009;25(15):1968–9.
9. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
10. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188–96.
11. Smit A, Hubley R, Green P. RepeatMasker open-4.0. 2013–2015. Seattle, WA, USA: Institute for Systems Biology; 2015. https://www.repeatmasker.org/faq.htmlb .
12. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6(1):11.
13. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33(suppl_2):W116–W12020.

Publisher's Note

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