DATA NOTE

Whole genome sequence and genome-wide distributed single nucleotide polymorphisms (SNPs) of the Black Bengal goat

[version 1; peer review: 2 approved with reservations]

Md. Bazlur Rahman Mollah ¹, Md. Shamsul Alam Bhuiyan²,
M.A.M. Yahia Khandoker², Md. Abdul Jalil³, Gautam Kumar Deb⁴,
Md. Panir Choudhury³, Nure Hasni Desha ³

¹Poultry Biotechnology and Genomics Laboratory, Department of Poultry Science, Bangladesh Agricultural University, Mymensingh, 2202, Bangladesh

²Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh, 2202, Bangladesh

³Goat and Sheep Production Research Division, Bangladesh Livestock Research Institute, Savar, 1341, Bangladesh

⁴Biotechnology Division, Bangladesh Livestock Research Institute, Savar, 1341, Bangladesh

v1 First published: 21 Mar 2019, 8:318
<https://doi.org/10.12688/f1000research.18316.1>
Latest published: 21 Mar 2019, 8:318
<https://doi.org/10.12688/f1000research.18316.1>

Abstract

The Black Bengal goat (BBG) is a dwarf sized heritage goat (*Capra hircus*) breed from Bangladesh, and is well known for its high fertility, excellent meat and skin quality. Here we present the first whole genome sequence and genome-wide distributed single nucleotide polymorphisms (SNPs) of the BBG. A total of 833,469,900 raw reads consisting of 125,020,485,000 bases were obtained by sequencing one male BBG sample. The reads were aligned to the San Clemente and the Yunnan black goat genome which resulted in 98.65% (properly paired, 94.81%) and 98.50% (properly paired, 97.10%) of the reads aligning, respectively. Notably, the estimated sequencing coverages were 48.22X and 44.28X compared to published San Clemente and the Yunnan black goat genomes respectively. On the other hand, a total of 9,497,875 high quality SNPs ($Q \geq 20$) along with 1,023,359 indels, and 8,746,849 high quality SNPs along with 842,706 indels were identified in BBG against the San Clemente and Yunnan black goat genomes respectively. The dataset is publicly available from NCBI BioSample (SAMN10391846), Sequence Read Archive (SRR8182317, SRR8549413 and SRR8549904), with BioProject ID PRJNA504436. These data might be useful genomic resources in conducting genome wide association studies, identification of quantitative trait loci (QTLs) and functional genomic analysis of the Black Bengal goat.

Keywords

Black Bengal goat, whole genome sequence, short reads, SNP

Open Peer Review

Approval Status ? ?

	1	2
version 1		
21 Mar 2019	view	view

1. **Jason Miller** , Shepherd University,
Shepherdstown, USA

2. **Almas A. Gheyas**, University of Edinburgh,
Midlothian, UK

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Genomics and Genetics** gateway.

Corresponding author: Md. Bazlur Rahman Mollah (mbrmollah.ps@bau.edu.bd)

Author roles: **Mollah MBR:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bhuiyan MSA:** Conceptualization, Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Khandoker MAMY:** Conceptualization, Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Jalil MA:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Deb GK:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Choudhury MP:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Desha NH:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported partly by Bangladesh Agricultural University Research System (BAURES) [2018/671]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2019 Mollah MBR *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mollah MBR, Bhuiyan MSA, Khandoker MAMY *et al.* **Whole genome sequence and genome-wide distributed single nucleotide polymorphisms (SNPs) of the Black Bengal goat [version 1; peer review: 2 approved with reservations]** F1000Research 2019, 8:318 <https://doi.org/10.12688/f1000research.18316.1>

First published: 21 Mar 2019, 8:318 <https://doi.org/10.12688/f1000research.18316.1>

Introduction

The Black Bengal goat (BBG) is a small-sized breed of goat (*Capra hircus*) distributed throughout Bangladesh, West Bengal, Bihar, and Orissa regions of northeastern India. The predominant coat color of this breed is black but it is also found in brown, white and gray (Jalil *et al.*, 2018). It is a heritage goat breed of Bangladesh, and well known for its high fertility, excellent meat and skin quality. This animal is a source of high quality meat, milk, and leather, and contributes substantially to the economy of Bangladesh (Amin *et al.*, 2000; Faruque *et al.*, 2017). The BBG is reported to have originated from wild goat, also known as the bezoar or Pasang (*Capra aegagrus*) (Herre & Röhrs, 1990), having introgressed genes from the markhor (*Capra falconeri*). Inheritance of genetic materials from the goats from the Southern region of China to the BBG has been hypothesized, given the historical cultural and geographical connection between South China and the Bengal across the South-Eastern offshoot of the Tibetan plateau (Nozawa, 1991). Despite its economic importance, no large-scale genomic resource is available to date for this goat breed. Here we used the Illumina HiSeq sequencing platform to sequence the whole genome of the BBG, generated short reads and identified high quality genome wide distributed single nucleotide polymorphisms (SNPs). These data might provide useful insight for conducting genome wide association studies, the identification of quantitative trait loci (QTLs) and functional genomic analysis of the Black Bengal goat.

Methods

Experimental animal

The experimental goat was reared at Bangladesh Livestock Research Institute (BLRI) goat farm under semi-intensive management system including slatted floor, well ventilated open sided house attached to pasture. All efforts were made to ameliorate harm to the animal. A small piece of ear tissue from an adult (30 months old) pedigreed goat (BioSample SAMN10391846) was collected by ear punching using a sterilized tissue puncher following local anesthesia (Lidocaine hydrochloride, 2%) on the right ear and immediately frozen into liquid nitrogen. Prior to ear punching, the goat was handled calmly with great care by a trained animal operator to prevent distress and injury to the animal and the handler. The tissue punching site was finally treated with antiseptic cream (Cetrimide, 0.5% and chlorhexidine digluconate 0.1%). All the animal procedure conformed the guidelines of the AWEC (Animal Welfare and Ethical Committee) of Bangladesh Agricultural University.

Sample processing

The tissue was finely ground by Micro Pestle (Sigma-Aldrich cat # SIAL501ZZ0), and high molecular weight DNA was extracted from the fresh frozen tissue using the Phenol:chloroform:isoamyl alcohol method (Sambrook & Russell, 2001). DNA purity was evaluated by NanoDrop 1000 Spectrophotometer (Life Technologies, CA, USA) and 0.8% agarose gel electrophoresis. DNA quantity was quantified using Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Life Technologies, CA, USA cat # Q32851). DNA was fragmented by acoustic disruption using Covaris S220 ultrasonicator and then underwent end repair,

detailing, adapter ligation and purification (NEBNext UltraII DNA library Prep Kit cat # E7645S) following manufacturer instructions. The purified DNA was further selected for the right size before PCR amplification for library construction. The preliminary quantification and dilution of the library was performed using Qubit 2.0 Fluorometer, and, then Agilent 2100 Bioanalyzer was used to determine the insert size and nucleic acid concentration of the resulting library. The effective concentration of each sample in the library mixture was determined by qPCR (ABI 7500, Applied Biosystems, CA, USA) using the KAPA Library Quantification Kit (Cat. # KK4824) following the manufacturer's standard protocol with the primer pair Primer 1: 5'-AAT GAT ACG GCG ACC ACC GA-3' Primer 2: 5'-CAA GCA GAA GAC GGC ATA CGA-3'. The PCR conditions were as follow: an initial denaturation at 95°C for 5 min followed by 35 cycles (denaturation at 95°C for 30 sec, annealing/extension/data acquisition at 60°C for 45 sec) and melt curve analysis at 65 – 95°C before sequencing to ensure the accuracy of the sample concentration.

Sequencing

Sequencing was performed on the Illumina system (HiseqX) according to manufacturer's instructions. The samples were sequenced using a 2 × 150 paired-end (PE) configuration (GENEWIZ, Suzhou, China) using Illumina Truseq SBS Kit v4 (cat # FC-401-4003) in high output mode. Base calling was achieved with the sequencer built-in software **Real-Time Analysis (RTA)** (v1.5.15.1), which performs real-time conversion of the four fluorescent signals obtained from CCD (charge-coupled device) to binary base call (BCL) data. BCL data were then converted to fastq files using **bcl2fastq** (v2.17, Illumina). Data demultiplexing was then performed simultaneously based on index information.

Primary analysis was performed using the sequencer's built-in software **HCS** (v3.4.0) to determine whether the read can pass the chastity filter based on the signal quality of the first 25 cycles. If the read had no more than 2 out of the 25 cycles with chastity values below 0.6, the read was called PF (Pass Filter). PF clusters converted by **bcl2fastq** were called PF data and stored in FASTQ format. The raw data were filtered to remove adapter sequences, PE reads having Q scores of < 20 and N composition of >10%. After primary cleaning of reads, mitochondrial genomes were removed. Then the remaining high quality, contamination-free reads were aligned to both the San Clemente (**GCA_001704415.1**) and the Yunnan black goat genome (**GCA_000317765.2**) separately using **Bowtie2** (v2.3.4.3) (Langmead & Salzberg, 2012). **Samtools** (v1.9) (Li *et al.*, 2009) was used to convert the resulting SAM sequence alignment files to BAM format, followed by sorting, indexing and quality filtering. **BCFtools** (v1.9) (Narasimhan *et al.*, 2016) was used to call and filter the variants.

Validation

A total of 833,469,900 raw reads consisting of 125,020,485,000 bases were obtained by sequencing of one male BBG sample (BioSample SAMN10391846). After the QC, a total of 812,209,030 reads containing 118,911,538,136 bases were kept

which was 97.45% of the total raw reads. After quality filtration and removal of the mitochondrial genome, the reads were aligned to the San Clemente and the Yunnan black goat genome which resulted in 98.65% (properly paired, 94.81%) and 98.50% (properly paired, 97.10%) of the reads aligning, respectively. Additionally, a total of 9,497,875 high quality SNPs ($Q \geq 20$) along with 1,023,359 indels were identified in BBG versus the San Clemente genome (See underlying data (Mollah *et al.*, 2019a)). Similarly 8,746,849 high quality ($Q \geq 20$) SNPs along with 842,706 indels were identified BBG versus the Yunnan black goat genome (See underlying data (Mollah *et al.*, 2019b)). The transition and transversion ratio was 2.27 and 2.29 in BBG against the San Clemente and the Yunnan black goat respectively.

Data availability

Underlying data

Capra hircus isolate:Black Bengal Goat breed:Black Bengal Goat (goat). BioProject PRJNA504436:

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA504436>

This project contains the following underlying data:

- [SRR8549904](#) (Alignment of BBG whole genome sequences with Yunnan Black goat)
- [SRR8549413](#) (Alignment of BBG whole genome sequence with San Clemente goat)
- [SRR8182317](#) (WGS of Black Bengal Goat: Adult male ear tissue)

Figshare: Genome-wide distributed SNPs identified in the Black Bengal Goat versus the San Clemente genome. <https://doi.org/10.6084/m9.figshare.7708010.v1> (Mollah *et al.*, 2019a)

This project contains the following underlying data:

- [BBG_aln_SanClemente.bam.calls.q20.bcf](#) (SNP data identified in the Black Bengal Goat versus the San Clemente genome)
- [BBG_aln_SanClemente.bam.calls.q20.bcf.stats](#) (output file from analysis of .bcf file with bcftools)

Figshare: Genome-wide distributed SNPs identified in the Black Bengal goat versus the Yunnan Black goat genome. <https://doi.org/10.6084/m9.figshare.7707929.v1> (Mollah *et al.*, 2019b)

This project contains the following underlying data:

- [BBG_aln_Yunnan.bam.calls.q20.bcf](#) (SNP data identified in the Black Bengal Goat versus the Yunnan Black goat genome)
- [BBG_aln_Yunnan.bam.calls.q20.bcf.stats](#) (output file from analysis of .bcf file with bcftools)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Grant information

This work was supported partly by Bangladesh Agricultural University Research System (BAURES) [2018/671].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Bangladesh Livestock Research Institute (BLRI) for providing pedigreed Black Bengal goat for sample collection. We also acknowledge the BdREN (Bangladesh Research and Education Network, UGC) for sharing high performance computing infrastructure.

References

- Amin MR, Husain SS, Islam AB: **Evaluation of Black Bengal goats and their cross with the Jamunapari breed for carcass characteristics.** *Small Rumin Res.* 2000; **38**(3): 211–215.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Faruque M, Choudhury M, Ritchie C, *et al.*: **Assessment of performance and livelihood generated through community based goat production in Bangladesh.** *SAARC J Agric.* 2017; **14**(2): 12–19.
[Publisher Full Text](#)
- Herre W, Röhrs M: **Haustiere - zoologisch gesehen.** 2nd ed. Springer Spektrum, Heidelberg. 1990.
[Publisher Full Text](#)
- Jalil M, Choudhury M, Kabir M, *et al.*: **Morphometric characterization of Black Bengal Goat under farming condition in Bangladesh.** *Asian J Med Biol Res.* 2018; **4**(1): 95–104.
[Publisher Full Text](#)
- Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods.* 2012; **9**(4): 357–359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Handsaker B, Wysoker A, *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics.* 2009; **25**(16): 2078–2079.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mollah MBR, Bhuiyan MSA, Khandoker MAMY, *et al.*: **Genome-wide distributed SNPs identified in the Black Bengal Goat versus the San Clemente genome.** *figshare.* Fileset. 2019a.
<http://www.doi.org/10.6084/m9.figshare.7708010.v1>
- Mollah MBR, Bhuiyan MSA, Khandoker MAMY, *et al.*: **Genome-wide distributed SNPs identified in the Black Bengal goat versus the Yunnan Black goat genome.** *figshare.* Fileset. 2019b.
<http://www.doi.org/10.6084/m9.figshare.7707929.v1>
- Narasimhan V, Danecek P, Scally A, *et al.*: **BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data.** *Bioinformatics.* 2016; **32**(11): 1749–1751.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nozawa K: **Domestication and history of goats.** In: Majjala K. (Ed.), *World Animal Science: Disciplinary Approach.* Elsevier, Amsterdam. 1991; 391–404.
- Sambrook JF, Russell DW: **Molecular Cloning: A Laboratory Manual.** 3rd ed. Cold Spring Harbor Laboratory Press, New York. 2001.
[Reference Source](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 24 April 2019

<https://doi.org/10.5256/f1000research.20035.r46138>

© 2019 Gheyas A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Almas A. Gheyas

The Roslin Institute, University of Edinburgh, Midlothian, UK

The article is mostly well written and provides sufficient background about the importance of the Black Bengal goat as a rationale for generating the genomic data. There are, however, several places in the Methods section, which would require further information and/or clarification. I highlight the points below:

1. *"DNA was fragmented by acoustic disruption using Covaris S220 ultrasonicator and then underwent end repair, detailing..."*:

By "detailing", do you mean dA-tailing?

2. *"The purified DNA was further selected for the right size before PCR amplification for library construction."*

What was considered the "right size" and how were the fragments of desired size range selected?

3. *"The effective concentration of each sample in the library mixture..."*.

Since the sequencing was performed on a single BBG sample, the authors need to clarify that it was sequenced in a multiplex with other unrelated samples.

4. *"If the read had no more than 2 out of the 25 cycles with chastity values below 0.6..."*.

Do you mean 2 bases out of 25 cycles?

5. *"The raw data were filtered to remove adapter sequences, of < 20..."*

Does this Q score represent average value per read?

6. *"After primary cleaning of reads, mitochondrial genomes were removed."*

Please clarify why and how the mitochondrial genome was removed?

7. The study mentions using Bowtie2 for mapping sequence reads against reference genomes but does not provide any detail about parameters used for mapping. If the default setting was used, that needs to be mentioned. I am also wondering if any post-alignment processing (e.g. marking duplicate reads, indel realignment etc.) was applied. These processing steps are important as these will affect the variant calling and false discovery rate.

8. The authors mention about variant calling with BCFtools but do not provide any details about parameters used for calling/filtering variants. The only filtration criteria mentioned is $Q > 20$ and the filtered set has been called "high quality". I am not convinced if calling variants from a single individual with $Q > 20$ can be called "high quality" without performing further validation or without providing any information about false discovery rate. Besides I am wondering if the same criteria were used for calling/filtering both SNPs and indels? Please clarify this. I would also suggest the authors to remove the words "high quality" as in my opinion these are only a preliminary set of variables, which would require further filtration and validation before being used in further studies.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatic analysis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 April 2019

<https://doi.org/10.5256/f1000research.20035.r46140>

© 2019 Miller J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jason Miller

College of Natural Sciences and Mathematics, Shepherd University, Shepherdstown, WV, USA

With reference genome assemblies already available for the goat species *Capra hircus*, this manuscript announces a set of whole-genome shotgun reads from one individual of a breed that had yet to be sequenced. The data size is small, constituting less than one single run of an Illumina HiSeq X. The data were enhanced by mapping and SNP calling and those data are also made available. The manuscript is presented as a Data Note for which the content and format seem appropriate.

- The abstract and introduction provide possibly excessive background on the goat breed. The text includes “dwarf sized”, “small-sized”, “high fertility”, “excellent meat and skin quality”, “high quality meat”, “contributes substantially to the economy”, and “Despite its economic importance.” Even with citations, these descriptions are imprecise for a scientific journal. The introduction could reference one respectable study of this goat’s importance. It should address the need for a genomic explanation for certain traits, with references if possible.
- Two genomic references were used for SNP discovery and at least one mitochondrial genome sequence was used for filtering reads. The manuscript provides accessions for both genomic sequences but none for the mitochondria. The manuscript should provide literature citations for all of these reference sequences.
- The introduction should briefly review previous goat sequencing projects, e.g. the Yunnan goat (Dong *et al.*, 2013¹), the domestic goat (Bickhart *et al.*, 2017²), and some appropriate goat mitochondria study (e.g. Colli *et al.*, 2015³).
- The text should explain that one library was prepared from one tissue sample and then multiplexed for sequencing with several (unrelated?) libraries. The text only hints at this by using the words “mixture” and “demultiplexing”. It is critical to understand that the reads were multiplexed since otherwise the reported number of reads seems low for a HiSeq run.
- The English grammar could be polished in a few places: “transition and transversion ratio” should say “transition to transversion”; “were as follow:” should say “as follows”; “sequencer built-in software” should say “sequencer’s built-in software”. The Illumina instrument name seems to be spelled HiSeq X, not HiseqX.

References

1. Dong Y, Xie M, Jiang Y, Xiao N, et al.: Sequencing and automated whole-genome optical

mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol.* 2013; **31** (2): 135-41

[PubMed Abstract](#) | [Publisher Full Text](#)

2. Bickhart DM, Rosen BD, Koren S, Sayre BL, et al.: Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet* . 2017; **49** (4): 643-650 [PubMed Abstract](#) | [Publisher Full Text](#)

3. Colli L, Lancioni H, Cardinali I, Olivieri A, et al.: Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC Genomics*. 2015; **16**: 1115 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for creating the dataset(s) clearly described?

Partly

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research