

Whole genome analysis of Black Bengal goat from Savar Goat Farm, Bangladesh

Ziqrul Haq Chowdhury

Bangladesh Agricultural Research Council

Nazmul Hussain

Bangladesh Agricultural University

Saam Hasan

North South University

Ajran Kabir

Bangladesh Agricultural University

Muket Mahmud

Bangladesh Agricultural University

Mahdi Robbani

North South University

Tahmina Tabassum

North South University

Tamanna Afroze

North South University

Aura Rahman

North South University

Rafiqul Islam

Bangladesh Agricultural Research Council

Maqsud Hossain (✉ muhammad.maqsud@northsouth.edu)

North South University <https://orcid.org/0000-0001-9794-411X>

Research note

Keywords: Capra hircus, Bangladesh, Black Bengal goat

Posted Date: July 31st, 2019

DOI: <https://doi.org/10.21203/rs.2.12258/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on October 24th, 2019. See the published version at <https://doi.org/10.1186/s13104-019-4700-7>.

Abstract

Objectives Single Nucleotide Polymorphisms (SNPs) play critical roles in genetic diversity and disease. Many traits and diseases are linked with exonic SNPs that are significant for gene function, regulation or translation. This study focuses on SNPs that potentially act as the genetic basis for desirable traits in the Black Bengal Goat. This variety of goat native to South Asia, and is identified as one of the most commercially important meat producing animals in the world. The aim of this study was to sequence the genome of Black Bengal Goats and identify SNPs that might play a significant role in determining meat quality in the organism. The study focuses on exonic SNPs for their greater likelihood of affecting the final translated protein product. The genes were also filtered according to their functional relevance to meat quality. The study is based on a single observation. Results Approximately 76,000 exonic variants were identified in the study. After filtration using a Wilcoxon test based score, the number came down to 49,965 which were found to be distributed in 11,568 genes. The functional pathways affected by these variations included fatty acid metabolism and degradation, which are important processes that influence meat quality.

Introduction

The Black Bengal goat (*Capra hircus*) is a regional breed found in Bangladesh and eastern India. It is characterized by its smaller stature and is preferred in the livestock industry for its higher reproductive rate and quality of meat, both being key indicators of economic value (1). Meat quality in goats is measured by several indicators, the most important of which is the degree of tenderness. Previous studies have investigated the correlation between tenderization of meat and the expression of various genes (2). The goal of this study was to identify specific genetic markers that may correlate to improved meat quality observed in the Black Bengal goat using a whole genome sequencing (WGS) approach. Variant calling and SNP profiling was carried out on the sequenced data through comparison with previously shortlisted key genes so as identify those variants that may affect said genes and possibly confer changes that lead to desirable characteristics in the organism. The variants were also annotated to determine their downstream effect on transcriptional and translational processes. The rationale of this study was to identify SNPs, either unique changes or differences in the number of such variants, in various genes that may help in better understanding and characterizing the factors that determine meat quality in goats.

Methods

Blood sample from a female Black Bengal goat aged 4 years was collected from Savar Goat Farm where the goats are reared by maintaining their history. The blood samples were transported to the Department of Microbiology and Hygiene, Bangladesh Agricultural University (BAU) for DNA extraction and sample processing. All goats were handled following standard guidelines of AWEEC (Animal Welfare and Experimental Ethical Committee) of Bangladesh Agricultural University [approval number AWEEC/BAU/2019(13)].

DNA from the blood sample was extracted using QIAamp® DNA Blood Mini kit (QIAGEN, Germany) following manufacturer's instructions. DNA purity was evaluated by NanoDrop 1000 Spectrophotometer (Life Technologies, CA, USA) and 0.8% agarose gel electrophoresis. Sequencing was carried out at the Genewiz Sequencing Centre in Suzhou, China using an Illumina HiSeq platform that produced 1,151,332 trimmed reads and a 109-fold mean coverage.

For genome mapping BWA's mem algorithm (3) was selected with the minimum mapping quality set to 20 and unmapped reads filtered out using SAMtools (4) before indexing. The resulting bam file was sorted by coordinate from the left most position. Following this, read groups were added and PCR duplicates were marked using Picard (<http://broadinstitute.github.io/picard/>). For single nucleotide polymorphism (SNP) profiling and variant calling, a modified version of the GATK best practices pipeline was followed (DePristo et al., 2011), Base recalibration was performed with BaseRecalibrator, and finally variants were called with HaplotypeCaller in GVCF mode. Variants were further analysed on HaplotypeCaller based on several statistical parameters including QD, Wilcoxon test scores and fisher test score for strand bias.

The variants were annotated with ANNOVAR (5), that was provided with an annotation database built using gene annotation data from UCSC (6). In the next step, gene variants that have been previously implicated in meat tenderization were analysed. Only those variants were considered which passed the BaseQRankSumTest filter, a Mann-Whitney-Wilcoxon U-test. Further filtration was carried out by the ReadPosRankSum values which are calculated based on the result of the BaseQRankSumTest for each variant. The Manhattan plot was generated using qqman package available on R (7). Further functional annotation and pathway analysis of the final RNA editing sites was carried using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (8).

Results

After initial variant calling, variants not belonging to exonic regions as per Annovar's algorithm (5), were filtered out in order to focus on exonic variants. This is since exonic variants are more likely to affect the final protein product, possibly through alterations of gene function, expression, translation or even post-translational modifications. Following this filtration, the study was left with around 76,000 variants, a number that was further brought down to 49,965 after a second filtration based on Wilcoxon test scores as calculated by the BaseQSum calculator functionality of GATK's variant annotator (9).

These 49,965 exonic variant sites were found to be distributed in a total of 11,568 genes. These variants were categorized according to their effect on translation, i.e. synonymous, non-synonymous, stop-gain, frame-shift or unknown. To statistically analyze the association study, the u-based z approximation derived from the rank-sum test was used instead of the more conventional p-values.

Once more, GATK's variant annotation functionalities were used to calculate these values which are outputted as the ReadPosRankSum values. Negative values were omitted owing to those representing presence of the alternate alleles near the ends of reads which are often erroneous in these analyses. The remaining positive values were used to plot the frequency of variants per chromosome and used to speculate on possible chromosome-wise association between the variants and possible desired traits (Figure 1).

Genes that prior studies had implicated in the meat tenderization process, encode proteins including calpain-1 (Capn1), calpain-2 (Capn2), calpastatin (Cast), caspase 3 (Casp3), caspase 9 (Casp9), α B-crystallin (Cryab), heat shock protein 27 (Hsp27), heat shock protein 40 (Hsp40) and heat shock protein 70 (Hsp70) (Saccà et al., 2019). Among these, this study found variants in HSP70, CASP3, CAPN1, CAPN2 and CAST encoding genes. Table 1 shows the positions and types of variants found in these genes.

Table 1. Variants found in genes previously implicated in meat tenderization process.

Gene	Chromosome	Position	Base change	Amino acid change
CAST	NC_030814	14434086:exon3:c	G83A:p	R28>K
CAST	NC_030814	14434086:exon18:c	C1245G:p	V415>V
CAST	NC_030814	14434086:exon28:c	G2052A:p	E684>E
CAST	NC_030814	14434086:exon25:c	C1863T:p	P621>P
CAST	NC_030814	14434086:exon14:c	G972A:p	P324>P
HSP70.1	NC_030830	22440674:exon1:c	G309A:p	V103>V
HSP70.1	NC_030830	22440674:exon1:c	C993T:p	I331>I
HSP70.1	NC_030830	22440674:exon1:c	C1653T:p	S551>S
HSP70.1	NC_030830	22440674:exon1:c	C1528T:p	L510>L
HSP70.1	NC_030830	22440674:exon1:c	C24A:p	G8>G
CAPN1	NC_030836	43822608:exon17:c	C1803A:p	R601>R
CAPN2	NC_030823	25579185:exon12:c	G1522A:p	D508>N
CAPN2	NC_030823	25579185:exon13:c	T1560A:p	L520>L
CAPN2	NC_030823	25579185:exon17:c	G1764A:p	G588>G
CAPN2	NC_030823	25579185:exon18:c	C1833T:p	Y611>Y
CAPN2	NC_030823	25579185:exon3:c	C327T:p	A109>A
CASP3	NC_030834	30502447:exon3:c	C267T:p	N89>N

Further analysis looking into the number and nature of non-synonymous variants that occurred in this study's specimen and the changes they resulted in for the encoded protein revealed the final list to contain a total of 20,879 non-synonymous variants that were distributed in 6850 genes. The results of the gene ontology search also provided some promising findings. In particular, the functional clustering analysis showed that the processes most likely impacted by the identified genes were fatty acid metabolism and fatty acid degradation. Both of these processes may hold significant contributions to the leanness and eventual quality of meat produced from the concerned specimen (10). The third pathway with the most significant scores from the list of genes was the PPAR pathway (peroxisome proliferative-activated receptor) which too is concerned with fatty acid metabolism. Table 2 shows these results alongside the associated statistical scores as calculated by DAVID's in-built algorithm.

Table 2. Pathways affected by the identified SNPs and their associated statistical scores.

Annotation Cluster 1	Enrichment Score: 3.16	Count	p-Value	Benjamini
KEGG_PATHWAY	Fatty acid degradation	17	3.40E-06	1.80E-04
KEGG_PATHWAY	Fatty acid metabolism	15	1.00E-03	9.70E-03
KEGG_PATHWAY	PPAR signalling pathway	13	9.20E-02	3.40E-01
Annotation Cluster 2	Enrichment Score: 2.8	Count	p-Value	Benjamini
KEGG_PATHWAY	Protein digestion and absorption	36	2.90E-08	3.80E-06
KEGG_PATHWAY	ECM-receptor interaction	22	5.20E-04	5.50E-03
KEGG_PATHWAY	Focal adhesion	34	2.30E-02	1.20E-01
KEGG_PATHWAY	Amoebiasis	20	3.10E-02	1.50E-01
KEGG_PATHWAY	PI3K-Akt signaling pathway	32	9.40E-01	9.90E-01

Discussion

A total of 11,564 genes were identified from a list of 47,241 exonic variants. This represents a large number of genes, which could theoretically represent close to 50% of all human protein coding genes. This does bring the results of this study into doubt as it increases the likelihood of discovering variants by chance and sequencing errors. However, this study still provides some evidence regarding the significance of certain gene variants in influencing beneficial characteristics of the Black Bengal goat. Among these variants, two contained non-synonymous changes that led to alterations in the encoded amino acid. CAST contained a variant that would lead to an arginine being substituted for lysine at position 28. CAPN2 contained a change that would result in an aspartic acid being substituted by asparagine at position 508. In addition several variants in HSP70 protein were also found, along with considerable numbers of more in other HSP proteins. The HSP family has previously been heavily linked with the process of meat tenderization (2). The lipid metabolism pathways being highlighted by the ontology analysis is another indication of specific variants contributing to the observable characteristics of these organisms (10).

In conclusion, we would like to report this study as a preliminary insight into the possible genetic basis of improved meat quality in Black Bengal goats, with findings that deserve further research and investigation. The goat remains an important organism in the livestock industry and hence its understanding and improvement hold significant economic value. Further genomic characterization and analysis of this breed of goat with a far greater number of samples would be able to verify our initial

findings and add new evidence that would allow a better discerning of the unique variants that may confer this species its desirable traits.

Limitations

The primary limitation of this work is that it has a sample size of one. A larger dataset comprising multiple samples, including different breeds, would allow for more robust statistical analysis as well as establishing confidence against false positives. We hope to build on this study and develop a more verified genetic basis of the concerned traits in the Black Bengal goat in future.

Declarations

Ethics approval and consent to participate

All goats were handled by following the standard guidelines of the AWEEC (Animal Welfare and Experimental Ethical Committee) of Bangladesh Agricultural University [approval number AWEEC/BAU/2019(13)]

Consent for Publication

Not Applicable

Availability of data and materials

The sequenced genome is available and raw data was submitted to GenBank with the accession no: PRJNA533512.

Competing Interests

The authors declare there is no competing interest.

Funding

This work was supported by Bangladesh Agricultural Research Council (BARC), Farmgate, Dhaka, Bangladesh.

Author's Contributions

1. Shah Md. Ziqrul Haq Chowdhury- Conceptualization, funding acquisition, investigation, supervision, writing, review and editing
2. KHM Nazmul Hussain Nazir- Conceptualization, DNA extraction, investigation, methodology, project administration, resources, supervision, writing – original draft preparation, writing – review and editing
3. Saam Hasan- Computational and statistical analysis, writing – original draft preparation
4. Ajan Kabir Samin- Conceptualization, sample collection, experimentation, writing – original draft preparation
5. Muket Mahmud- Conceptualization, sample collection, experimentation, writing – original draft preparation
6. Mahdi Robbani- Computational and statistical analysis
7. Tahmina Tabassum- Computational and statistical analysis
8. Tammana Afroze- Computational and statistical analysis
9. Aura Rahman- Writing – Writing, final formatting, review and editing
10. Rafiqul Islam- Conceptualization, sample collection, experimentation, writing – original draft preparation
11. Maqsd Hossain- Conceptualization, resources, computational and statistical analysis supervision, writing – review and editing

Acknowledgements

The authors would like to thank Savar Goat Farm, Savar, Dhaka for providing Black Bengal goat for sample collection.

References

1. Rout PK, Joshi MB, Mandal A, Laloe D, Singh L, Thangaraj K. Microsatellite-based phylogeny of Indian domestic goats. *BMC Genet.* 2008;9(11).
2. Saccà E, Corazzin M, Bovolenta S, Piasentier E. Meat quality traits and the expression of tenderness-related genes in the loins of young goats at different ages. *Animal.* 2019;1–10.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
4. 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.

5. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
6. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, SUGnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004;32:D493–6.
7. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw.* 2018;3(25).
8. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
9. Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491–501.
10. Goetsch AL, Merkel RC, Gipson TA. Factors affecting goat meat production and quality. *Small Rumin Res.* 2011;101:173–81.

Figures

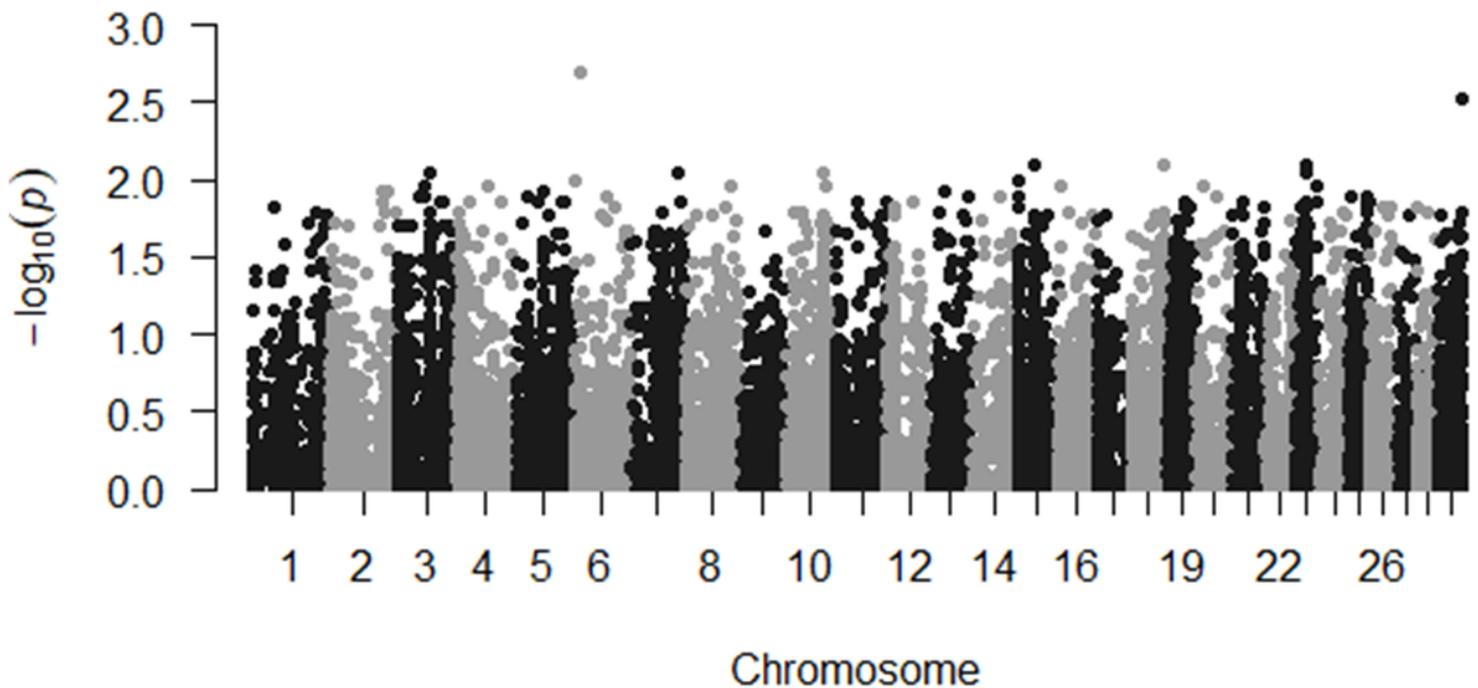


Figure 1

Manhattan Plot of U-based Z Approximation Scores of SNPs per chromosome. The graph shows SNPs and their associated ReadPosRankSum values for each chromosome.