In Silico Analysis of Genome Wide Non-Synonymous Single Nucleotide Polymorphisms in Indigenous Cattle Breeds of Pakistan

Umer Farooq 1,2, Nimra Murtaza 2, Abubakar Siddique 1, Bilal Saleem 1, Obaid Ur Rehman 1, Nageen Zahra 1, Muhammad Uzair 1, Muhammad Naem Riaz 1,3* and Muhammad Ramzan Khan 1*

1 National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Centre, Park Road, Islamabad, Pakistan
2 Department of Biosciences, COMSATS University Islamabad, Park Road, Islamabad, Pakistan
3 Animal Biotechnology Program, Animal Sciences Institute, National Agricultural Research Centre, Park Road, Islamabad, Pakistan

ABSTRACT

Genomic selection programs for yield enhancement and disease resistance have become a reality with the availability of highly detailed genomic information. This information is critical to highlight genetic polymorphisms related with economically important traits including milk and meat yield, development, and resistance against diseases. In this present study, our main objective was to identify the deleterious SNPs and their associated genes which possibly disrupt protein’s structure and function, as well as lead to genetic disease. We performed genome wide reference-based sequence alignment and functional annotation to identify deleterious non-synonymous SNPs (nsSNPs) in cattle breeds of Pakistan. For this purpose, genomic data of four different purpose cattle breeds including Bhagnari, Cholistani, Sahiwal and Red Sindhi was analyzed. Comparison with taurine reference genome ARS-UCD.1.2.99 discovered 29,032,662 genomic variations of which 25,469,157 were single nucleotide polymorphisms (SNPs) and 3,563,505 were Insertion/Deletions (InDels). Functional annotation identifies 122,943 missense SNPs that may possibly affect economically important traits. Using sequence and structure based computational tools SIFT, we identified 154 deleterious variants in 134 genes. Gene enrichment highlighted the presence of these genes in different biological processes including developmental, signaling, transport, metabolic and homeostasis. These findings are useful resource for further exploration into the molecular processes associated with these variances.

INTRODUCTION

Livestock have traditionally played important part in development of human societies for a long time. According to a rough estimate, livestock supports the livelihoods of approximately 0.6 billion farmers in underdeveloped nations (Thornton et al., 2006). Domestication of cattle has a major role in development of a country’s agriculture economy. They contribute more than milk, meat, fiber skin and fuel productions. Based on their physical appearances nearly 800 cattle breeds are divided into two categories, Bos taurus (taurine) and Bos indicus (indicine) (Felius et al., 2011). Beside many other reported differences, the appearance of a hump over the shoulders is one of the major phenotypic characteristics used to distinguish indicine breeds form taurine breeds (Zeder et al., 2006). Indicine cattle grow in tropical and subtropical climates such as Africa, Southeast Asia, Brazil, northern Australia, southern China, and sections of the United States thanks to its physiological characteristics. Taurine, on the other hand, is typically found in more developed European countries, North America, and Australia, where it has a greater metabolic rate and dietary requirements. There are collectively 35 recognized cattle breeds in Pakistan and India (Felius et al., 2014) which make considerable contribution to the agriculture
economy. After goats, cows are the second most common animal species in Pakistan with an estimated 51.5 million animals (Finance Division, 2021), accounting for 3.2% of world cattle population (Singh et al., 2014). According to environmental conditions, most cattle breeds are created as drought breeds like Bhagnari, while Red Sindhi and Sahiwal are milk breeds and Tharparkar, Achai, Gabrali, and Cholistan are dual-purpose cattle breeds (Felius et al., 2014).

The process of natural and human driven selection has changed the cattle genotypes leading to diverse genetic and phenotypic profiles, and adaptation to temperature and tropical environments (MacHugh et al., 1997; Porto-Neto et al., 2013). Since the completion of the taurine cattle genome (Elsik et al., 2009), thousand bull genome project (Hayes and Daetwyler, 2019), the worldwide bovine genome sequencing, and the HapMap project (Gibbs et al., 2009), a considerable number of genetic variants including single nucleotide polymorphisms (SNPs) and InDels (Insertions/ Deletions) that have been recorded in a publicly available database (Sherry et al., 2001; Eck et al., 2009; Elsik et al., 2009; Gibbs et al., 2009; Iqbal et al., 2019).

Due to the advancements and cheap cost of next-generation sequencing (NGS), it is now feasible to simultaneously generate sequence data and estimate SNP allele frequencies in a range of reference populations (Kumar et al., 2012). Whole genome SNP genotyping analysis detected genomic regions targeted by selection, which, for example, contain immune and environmental adaptation related genes. This has opened new doors for studying the genetic variation and processes which help in adaptation to different biogeographic regions (Iso-Touru et al., 2016; Weldenegodguad et al., 2019). SNPs are the most prevalent type of genetic variation that are stable, have a low mutation rate, are inherited in a mendelian manner, and are relatively inexpensive to genotype (Snelling et al., 2005). As a result, they are routinely utilized as biomarkers in genetic research. Advances in bioinformatics and statistical tools have also improved our understanding of demographic evolution, the possible role of genomic structural variations, adaptation during domestication and selection, and the biological functions of these genomic variations in livestock breed (Gutenkunst et al., 2009; Li and Durbin, 2011; Jacob et al., 2020).

Regardless of these recent breakthroughs, there is a severe gap in whole-genome investigations of Asian cow breeds and European bovine breeds that are commonly employed for meat and dairy production (Kawahara-Miki et al., 2011; Choi et al., 2014). In this study, we analyzed the whole genome sequence data for four different purpose cattle breeds of Pakistan. Our main goal was to identify and analyze the missense variants resulting from nsSNP and predict their effect on resulting amino acid sequence, their structure, functionality, and stability. Missense variations not only modify the tertiary structure of proteins, but also lead to formation of deleterious phenotypes. Results of this study will be useful in furthering research into the genetic pathways underpinning features of interest in Indian cattle.

MATERIALS AND METHODS

Data collection

The whole genome sequences of Pakistani Bos indicus cattle samples included in this study were recently made available in public domain (Iqbal et al., 2019). Sequence data is available in China National GenBank Sequence Archive under project accession number CNP0000189. Total genomic DNA was extracted form whole blood (10 mL) samples and were sequenced using BGISEQ-500. For this study, we choose four different purpose breeds including Sahiwal (CNS0014619), Bhagnari (CNS0014606), Cholistan (CNS0014604), and Red Sindhi (CNS0014620).

Reads alignment and mapping

Quality check of sequence reads was performed using fast QC (Andrews, 2010) tool to identify low quality reads and adapter sequences. Raw sequences were first subjected to filtering process to identify and remove low quality reads and adapters using trimmomatic software (Bolger et al., 2014). Burrows-Wheeler Alignment (BWA) (Li and Durbin, 2009) algorithm was used with default parameters to create index files for reference genome ARS-UCD1.2 Btau5.0.1Y accessed from the 1000 Bull Genome project resources (Hayes and Daetwyler, 2019). Short reads of each selected cattle sample were mapped to reference genome using BWA with option “bwa-mem”. SAM files generated as result of alignment were converted to their binary equivalent (BAM) files and sorted using SAM tools (Li et al., 2009). PCR duplicates were removed using Picard tools “Mark Duplicate” command line utility from aligned reads (Picard tools by broad institute https://broadinstitute.github.io/picard/). Base Quality Score Recalibration (BQSR) was performed on mapped reads to resolve high or low base quality scores estimation. Final BAM files were used for downstream variant calling.

Variant calling

Single nucleotide polymorphisms (SNPs) along with insertion and deletions mutations (InDels) were identified using “Haplotype Caller” tool of genome analysis toolkit (GATK) which performs local de-novo assembly of haplotypes in genomic variation regions (McKenna et al.,
GATK was used according to guidelines of 1000 Bull Genome Project available at http://www.1000bullgenomes.com/doco/1000bullsGATK3.8pipelineSpecifications_Run8_Revision 20191101.docx (Fernandes Júnior et al., 2020). GATK “Select Variants” mode was used to separate SNPs and InDels into separate files. All variants including SNPs and InDels were discovered as difference from the reference genome ARS-UCD1.2 sequence. To remove the false positive calls, variants were filtered with GATK “Variant Filtration” argument using hard-filters with the following exclusion criteria: (1) quality by depth - QD<2.0; (2) Fisher Strand test - FS>60.0; (3) root mean square of the mapping quality score - MQ < 40.0; (4) and SOR > 9.0. SNPs and INDELS of autosomal chromosomes + X were left after filtration process.

Annotation of variations

Genetic variants were annotated using SnpEff program with annotation database ARS-UCD1.2.99 (Cingolani et al., 2012). SnpEff software assigns each SNPs and InDels to a functional class and provide several fields of information describing the affected transcripts and proteins, if applicable. SNPs and InDels were assigned to different functional classes including exonic, intronic, intergenic, splice site acceptor, splice site donor, splice site region, downstream, upstream, UTR 3 prime and UTR 5 prime region. InDels were also allocated to the in-frame deletion and insertion, disruptive in-frame deletion, and insertion functional classes whereas functional classes stop lost, stop retained, initiator codon were assigned to SNPs only. Variants were categorized into missense, nonsense, and silent variants. Missense or non-synonymous single base pair substitution result in amino acid substitution which affects the proteins phenotypically and functionally.

Detection of deleterious SNPs and go enrichment

SIFT tool (Vaser et al., 2016) was used to predicts the genes with tolerated and deleterious SNPs from the missense variants. This program identifies the tolerated and deleterious SNPs to predict the impact of amino acid substitution on phenotypic and functional alterations in protein molecules. SIFT program to identifies tolerance score (TI) for each in range from 0.0 to 1.0. The SIFT score ≤ 0.05 labels the non-synonymous variant as deleterious to protein function (Ng and Henikoff, 2003; Sim et al., 2012). Gene enrichment analysis was performed to identify over-represented biological processes using genes with deleterious SNPs. This gene lists comprising was submitted to the ClueGO (Bindea et al., 2009), a Cytoscape (Shannon et al., 2003) plug-in that combines Gene Ontology and KEGG pathways to produce a well-organized GO/ pathway annotation network, allowing researchers to see if these genes are linked to key pathways.

RESULTS AND DISCUSSION

Sequence data and alignment

Genomic sequences data of 4 different cattle breeds of Pakistan was analyzed which included dairy breeds Sahiwal and Red Sindhi, dual purpose breed Cholistani and drought tolerant breed Bhagnari. Individual sample sequence reads ranged from 754,986,078 to 1,605,207,524. Mapping of each sample to cattle reference genome ARS-UCD1.2 using BWA software yielded 83% to 99% alignment or sample reads (Table I). SAM alignment files were sorted and then converted to their binary format (BAM) files. Flow diagram depicts the bioinformatics analysis performed on the new generation sequence data (Fig. 1).

Identification and annotation of variants

Alignment BAM files (including Bhagnari, Sahiwal, Cholistani and Red Sindhi) were pooled together to perform variant calling. Total ~25.46 M SNPs and ~3.56 M InDels were identified in the 29 autosomes and X chromosome against bovine reference genome ARS-UCD1.2 (Fig. 2 and Supplementary Table S1). Among the 3,563,505 InDels, 1,87,986 were deletions. Number of SNPs present per individual samples were 13.19, 18.96, 15.63, and 15.21 million in Bhagnari, Cholistani, Sahiwal, and Red Sindhi, respectively. The transition to transversion (Ts/Tv) ratio was pretty much similar across the samples, 2.27 in Bhagnari, 2.3 in Cholistani, 2.28 in Sahiwal and 2.3 in Red-Sindhi (Table II). This is indicative of high quality of our SNPs identified. Number of SNPs and Indels depend on the length of chromosomes. Out of total SNPs and InDels, only 104152 (0.41%) SNPs and 409 (0.01%) InDels matched to previously reported variations in dbSNP build 150.
Table I. Number of raw reads and aligned reads in each sample.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Sample ID</th>
<th>Read length (bp)</th>
<th>Total reads (bp)</th>
<th>Mapped reads</th>
<th>Total reads aligned %</th>
<th>Unmapped reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholistani</td>
<td>CNS0014604</td>
<td>50</td>
<td>1,262,644,830</td>
<td>1,238,899,685</td>
<td>98.11</td>
<td>23,745,145</td>
</tr>
<tr>
<td>Bhagnari</td>
<td>CNS0014606</td>
<td>50</td>
<td>754,986,078</td>
<td>633,661,223</td>
<td>83.93</td>
<td>121,324,855</td>
</tr>
<tr>
<td>Sahiwal</td>
<td>CNS0014619</td>
<td>50</td>
<td>1,605,207,524</td>
<td>1,594,140,047</td>
<td>99.31</td>
<td>11,067,477</td>
</tr>
<tr>
<td>Red Sindhi</td>
<td>CNS0014620</td>
<td>50</td>
<td>1,579,398,030</td>
<td>1,568,961,018</td>
<td>99.33</td>
<td>10,437,012</td>
</tr>
</tbody>
</table>

Table II. Summary statistics of variants in each sample.

<table>
<thead>
<tr>
<th>SNP Stats</th>
<th>Bhagnari</th>
<th>Cholistani</th>
<th>Sahiwal</th>
<th>Red Sindhi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SNPs</td>
<td>13,196,027</td>
<td>18,967,415</td>
<td>15,634,144</td>
<td>15,215,012</td>
</tr>
<tr>
<td>Total InDels</td>
<td>1563914</td>
<td>2263706</td>
<td>2028795</td>
<td>1850518</td>
</tr>
<tr>
<td>Singleton SNPs</td>
<td>1376894</td>
<td>3039282</td>
<td>2163090</td>
<td>1960479</td>
</tr>
<tr>
<td>Het/Hom</td>
<td>0.53</td>
<td>2.21</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>Ts/Tv Ratio</td>
<td>2.27</td>
<td>2.3</td>
<td>2.28</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Fig. 2. Number of single nucleotide polymorphism (SNP) distribution by chromosome (A) and number of insertion/deletions (InDels) distribution by chromosome (B).

Variant sets including SNPs and InDels were functionally annotated to their attribute and genes using SnpEff tool. In our dataset of total SNPs identified, 16,750,695 (41.53%) were discovered in between genes. 1,785,969 (4.43%) SNPs were in thousand base pair (bp) upstream region of genes and 1,826,857 (4.53%) were in thousand base-pair downstream region of genes (Table II). Due to distinct isoforms or overlapping genes, multiple functional effects were identified when compared the total number of SNPs and InDels used for variant. 58,970,854 transitions and 25,313,817 transversions were detected with Ts/Tv ratio of 2.3296 in SNPs. SnpEff identified 2,909 high impact (disruptive), 122,494 moderate and 239,647 low impact SNPs in all sequenced samples (Supplementary Table SII). Functional Annotation of InDels predicted the presence of 2,331,013 (40.78%) and 2,779,680 (48.63%) InDels in intergenic and intronic regions, respectively. Moreover, 285,210 (5.00%) and 273,913 (4.80%) indels were located within thousand base pair upstream and downstream genic regions (Table III). 5,155 high impact (disruptive), 2,359 moderate and 5,730 low impact InDels were identified in all sequenced samples (Supplementary Table SII). SnpEff categorized the identified SNPs as missense, nonsense and silent according to their functional class (Table V). Missense variations (nsSNPs) can possibly cause amino acid substitutions which phenotypically and functionally effect important traits of cattle. For further downstream analysis, missense variants were carried forward.

Identification and enrichment of functional SNPs in coding region

Phenotypic effects of the missense variants were identified by SIFT tool. SIFT uses sequence homology-based approach to characterize the effect of amino acid substitution on protein function. SIFT analysis predicted 154 deleterious missense variants overlapping 134 genes in coding sequence regions with a score between 0.00-0.04. Moreover, 325 missense variants in coding sequence regions were predicted as tolerated (Table VI and Supplementary Table S3). Even though a small number of deleterious
SNPs were identified, all these variants overlapped protein coding genes. Gene list comprising deleterious SNPs was submitted in ClueGO to identify the functional gene ontology (GO) biological processes (BP). Gene ontology (GO) network is represented in Figure 3. Enrichment analysis revealed presence of these genes in several biological processes. Highlighted biological processes in gene interaction network included lipid transport (GO:0010876), transmembrane transport (GO:0055085), response to organic cyclic compounds (GO:0014070), response to endoplasmic reticulum stress (GO:0034976), negative regulation of transport (GO:0051051), detection of chemical stimulus (GO:0009593), skin development (GO:0043588), endomembrane system organization (GO:0010256), cytoskeleton organization (GO:0007010), response to growth factors (GO:0070848) and DNA conformation change (GO:0071103). List of all biological pathways and genes related is present in Supplementary Table S3.

![Figure 3](image)

**Fig. 3.** ClueGO gene ontology analysis of 134 genes with deleterious nsSNP identified in all samples. ClueGO identifies the enriched go terms and visualizes them in grouped annotation network. This network shows the relationship between the terms based on the similarity of their associated genes. Each node represents a gene ontology term and the associated genes.

### CONCLUSION

In silico approach to characterize genetic variations can assist us in predicting the consequences of mutations and explain their affecting role in biological mechanisms. In this study, we investigated whole genome sequence data of four different purpose cattle breeds in Pakistan (Cholistani, Sahiwal, Bhagnari and Red Sindhi) to identify genetic variations which lead toward deleterious effect on protein structure and functions. This analysis led to the discovery of 25,469,157 SNPs and 3,563,505 InDels in these breeds.

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 prime UTR variant</td>
<td>121,002</td>
<td>0.30%</td>
</tr>
<tr>
<td>5 prime UTR premature start codon gain variant</td>
<td>6,728</td>
<td>0.02%</td>
</tr>
<tr>
<td>5 prime UTR variant</td>
<td>43,109</td>
<td>0.11%</td>
</tr>
<tr>
<td>Downstream gene variant</td>
<td>1,826,857</td>
<td>4.53%</td>
</tr>
<tr>
<td>Initiator codon variant</td>
<td>18</td>
<td>0%</td>
</tr>
<tr>
<td>Intergenic region</td>
<td>16,750,695</td>
<td>41.53%</td>
</tr>
<tr>
<td>Intron variant</td>
<td>19,403,911</td>
<td>48.11%</td>
</tr>
<tr>
<td>Missense variant</td>
<td>122,494</td>
<td>0.30%</td>
</tr>
<tr>
<td>Noncoding transcript exon variant</td>
<td>27,818</td>
<td>0.07%</td>
</tr>
<tr>
<td>Splice acceptor variant</td>
<td>388</td>
<td>0.00%</td>
</tr>
<tr>
<td>Splice donor variant</td>
<td>684</td>
<td>0.00%</td>
</tr>
<tr>
<td>Splice region variant</td>
<td>36,706</td>
<td>0.09%</td>
</tr>
<tr>
<td>Start lost</td>
<td>239</td>
<td>0.00%</td>
</tr>
<tr>
<td>Stop gained</td>
<td>1,410</td>
<td>0.00%</td>
</tr>
<tr>
<td>Stop lost</td>
<td>192</td>
<td>0%</td>
</tr>
<tr>
<td>Stop retained variant</td>
<td>136</td>
<td>0%</td>
</tr>
<tr>
<td>Synonymous variant</td>
<td>202,961</td>
<td>0.50%</td>
</tr>
<tr>
<td>Upstream gene variant</td>
<td>1,785,969</td>
<td>4.43%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>InDels</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 prime UTR variant</td>
<td>22,164</td>
<td>0.39%</td>
</tr>
<tr>
<td>5 prime UTR variant</td>
<td>6,639</td>
<td>0.12%</td>
</tr>
<tr>
<td>Bidirectional gene fusion</td>
<td>15</td>
<td>0%</td>
</tr>
<tr>
<td>Conservative in frame deletion</td>
<td>497</td>
<td>0.01%</td>
</tr>
<tr>
<td>Conservative in frame insertion</td>
<td>509</td>
<td>0.01%</td>
</tr>
<tr>
<td>Disruptive in frame deletion</td>
<td>867</td>
<td>0.02%</td>
</tr>
<tr>
<td>Disruptive in frame insertion</td>
<td>546</td>
<td>0.01%</td>
</tr>
<tr>
<td>Downstream gene variant</td>
<td>285,213</td>
<td>4.99%</td>
</tr>
<tr>
<td>Frameshift variant</td>
<td>4,395</td>
<td>0.08%</td>
</tr>
<tr>
<td>Gene fusion</td>
<td>15</td>
<td>0%</td>
</tr>
<tr>
<td>Intergenic region</td>
<td>2,331,013</td>
<td>40.78%</td>
</tr>
<tr>
<td>Intragenic variant</td>
<td>3</td>
<td>0%</td>
</tr>
<tr>
<td>Intron variant</td>
<td>2,779,680</td>
<td>48.63%</td>
</tr>
<tr>
<td>Noncoding transcript exon variant</td>
<td>3,020</td>
<td>0.05%</td>
</tr>
<tr>
<td>Noncoding transcript variant</td>
<td>86</td>
<td>0.00%</td>
</tr>
<tr>
<td>Splice acceptor variant</td>
<td>439</td>
<td>0.01%</td>
</tr>
<tr>
<td>Splice donor variant</td>
<td>432</td>
<td>0.01%</td>
</tr>
<tr>
<td>Splice region variant</td>
<td>3,027</td>
<td>0.11%</td>
</tr>
<tr>
<td>Start lost</td>
<td>36</td>
<td>0.00%</td>
</tr>
<tr>
<td>Start retained variant</td>
<td>5</td>
<td>0%</td>
</tr>
<tr>
<td>Stop gained</td>
<td>55</td>
<td>0.00%</td>
</tr>
<tr>
<td>Stop lost</td>
<td>54</td>
<td>0.00%</td>
</tr>
<tr>
<td>Stop retained variant</td>
<td>7</td>
<td>0%</td>
</tr>
<tr>
<td>Transcript ablation</td>
<td>14</td>
<td>0%</td>
</tr>
<tr>
<td>Upstream gene variant</td>
<td>273,913</td>
<td>4.79%</td>
</tr>
</tbody>
</table>
Numerous SNPs, InDels and genes were found that have not been annotated yet. We annotated each variant’s possible functional role, allowing us to find numerous functionally significant candidate variants. Functional annotation revealed 122,943 missense (non-synonymous) SNPs in all samples 2,909 and 5,155 high impact (disruptive) SNPs and InDels in were also predicted by annotation. Analysis of missense variations revealed several genes with deleterious variation, involved in different biological and cellular functions. Go enrichment analysis revealed that genes harboring deleterious variations are significantly enriched in important biological processes, such as metabolic processes, development, transport, and homeostasis processes. Because this whole-genome sequencing study used just one animal of each breed, more research is needed to understand the exact dynamics of each gene–trait combination. The findings of this study can further be used as significant resource for further research on genomic characteristics to find variations in economically important traits and to develop precise genomic tools for cow breeding.

**ACKNOWLEDGMENTS**

The authors are highly grateful to Director NIGAB and NARC for providing all the necessary facilities.

**Supplementary material**

There is supplementary material associated with this article. Access the material online at: https://dx.doi.org/10.17582/journal.pjz/20211221081250

**Statement of conflict of interest**

The authors have declared no conflict of interest.

**REFERENCES**


Fernandes Júnior, G.A., de Oliveira, H.N., Carvalheiro,


Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider,


Supplementary Material

In Silico Analysis of Genome Wide Non-Synonymous Single Nucleotide Polymorphisms in Indigenous Cattle Breeds of Pakistan

Umer Farooq¹,², Nimra Murtaza², Abubakar Siddique¹, Bilal Saleem¹, Obaid Ur Rehman¹, Nageen Zahra¹, Muhammad Uzair¹, Muhammad Naeem Riaz¹,²* and Muhammad Ramzan Khan¹*¹

¹National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Centre, Park Road, Islamabad, Pakistan
²Department of Biosciences, COMSATS University Islamabad, Park Road, Islamabad, Pakistan
³Animal Biotechnology Program, Animal Sciences Institute, National Agricultural Research Centre, Park Road, Islamabad, Pakistan

Supplementary Table S1. Distribution of SNPs and InDels across chromosomes.

Supplementary Table S2. Number of variants by their impact.

Supplementary Table S3. Results of the deleterious nsSNP analysis by SIFT tool.

Supplementary Table S4. Gene Ontology biological processes (bp) enrichment by ClueGo.

* Corresponding author: mrkhan@PARC.gov.pk; naeemriaz@parc.gov.pk
0030-9923/2022/0001-0001 $ 9.00/0

Copyright 2022 by the authors. Licensee Zoological Society of Pakistan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).