Research Article

Genetic Basis of Immunity in Indian Cattle as Revealed by Comparative Analysis of Bos Genome

Menaka Thambiraja[1], Shukkruthi K. Iyengar[1], Brintha Satishkumar[1], Sai Rohith Kavuru[1], Aakanksha Katari[1], Dheer Singh[2], Suneel K. Onteru[2], Ragothaman M. Yennamalli[1]

1. Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed to be University, India; 2. Molecular Endocrinology, Functional Genomics and Systems Biology Laboratory, National Dairy Research Institute, India

Indian cattle breeds (*Bos indicus*) are renowned for their greater resilience compared to *Bos taurus* in diverse environments and resistance to infections. However, the genomic signatures associated with disease-resistance traits in Indian cattle remain underexplored. A genome-wide comparison between *Bos indicus* and *Bos taurus* could uncover key immune markers related to disease-resistance traits. We conducted chromosome-by-chromosome analyses among Bos genomes using three pairwise combinations, namely *Bos taurus* (Hereford) vs. *Bos indicus* (Nelore breed), *Bos taurus* (Hereford) vs. *Bos indicus* (Gir), and *Bos indicus* (Nelore) vs. *Bos indicus* (Gir), to investigate genomic variations in immunity genes and quantitative trait loci (QTLs). The whole genome for Nelore, Gir, and Hereford breeds were compared using bioinformatics tools: SyMAP, GSAlign, and SyRI. Non-syntenic regions were associated with four immunity genes in Nelore: VSTM1L, Galectin-9, HDAC5, and POLH. Significant genomic differences were observed on chromosomes 1 and X across all inter-breed comparisons. Chromosome X of Nelore lost 50 Mbps of genomic regions compared to those of Hereford, although all genes remained intact. Further exploration of translocation events on Chromosome X in Nelore revealed chromosomes 13 and 11 had more single nucleotide variants and insertions, respectively. In total, 8, 16, and 209 innate immunity genes in Nelore were impacted by longer insertions, deletions, and substitutions respectively, and 62 innate immunity genes were affected by all three structural variations. QTL analysis indicated these variations were primarily linked to milk and health traits. This study highlights genomic variations in innate immunity genes in Nelore and associated QTLs.

 doi.org/10.32388/VZC1YY

Corresponding authors: Ragothaman M. Yennamalli, ragothaman@scbt.sastra.edu; Suneel K. Onteru, Suneel.Onteru@icar.gov.in

Introduction

Indicine cattle (members of *Bos indicus*), such as zebu cattle, native to the Indian subcontinent, have shown economic advantages over taurine cattle (members of *B. taurus*), particularly when raised under tropical conditions, and account for about 80% of India's total cattle population[1][2]. The native breeds are crucial to the country's dairy industry, have made India a global leader in milk production, and support over 80 million rural households. Currently, 50 distinct breeds of indicine cattle are registered in India, evolved through both natural selection and selective breeding for adaptation to different climatic conditions. These breeds can cope with extreme heat or similar adverse climatic conditions, survive despite poor nutrition, and are resistant to several diseases[3][4][5][6]. One notable example is 'Vechur' (named after a village of the same name), the smallest cattle breed in the world, found in Kottayam district of Kerala and well-known for its greater resistance to viral, bacterial, and parasitic diseases than that shown by other exotic breeds and crossbred cattle[7].

The innate immune response, which kicks in at the earliest stages of infection in mammals, protects them from pathogenic infections. The response is specific to the pathogen[8][9][10]. Understanding the differences between the immune systems of indicine and taurine cattle could provide valuable insights into disease resistance shown by members of *Bos indicus*, a trait that is essential for better management of herds[11][12]. Exploring the genomic regions responsible for variations in disease resistance and susceptibility opens new opportunities to identify significant genomic signatures[13].

Recent research on livestock improvement is increasingly focused on whole-genome data to detect genetic factors or signatures associated with important traits such as disease resistance by identifying such structural variations as insertions, deletions, duplications, inversions, and translocations. Two relevant sources of assembled whole-genome data at chromosome level cattle are (1) the first whole-genome sequence of *Bos indicus* cattle (Nelore breed) with 52X coverage by the SOLiD sequencing platform with a short-read length of 25–50 bases[14] and (2) the whole-genome sequencing data of *Bos taurus* cattle with multiple improvements, such as covering 91% of the genome, closing gaps, correcting errors, removing bacterial contigs, and identifying the first portion of Y chromosomes using whole-genome shotgun sequencing methods[15]. Recently, a comparative whole-genome analysis of several *Bos*

indicus breeds (Kangayam, Tharparkar, Sahiwal, Red Sindhi, and Hariana) and of *Bos taurus* cattle revealed over 155 million single nucleotide polymorphisms (SNPs) and more than one million InDels among a total of 17,252 genes. Among these genes, many involved in innate immune responses were associated with key pathways such as toll-like receptor signalling, retinoic acid-inducible gene I-like receptor signalling, NOD-like receptor signalling, and the Jak-STAT pathway. The same study identified many missense variants in genes such as TLR3 and TLR4, transcription factors such as IRFs (IRF3/IRF7) and NF-kB, and non-synonymous variants in genes such as MyD88, IRAK4, RIG-I, TRIM25, MAVS, NOD1, and NOD2[16]. Furthermore, draft de novo genome assemblies and mitochondrial genome assemblies of *Bos indicus* breeds such as Ongole, Kasargod Dwarf, Kasargod Kapila, and Vechur were generated using Illumina short-read technology. An analysis of the exon-intron structure of 15 key genes related to bovine traits including milk quality, metabolism, and immune response revealed structural variation in their exon-intron numbers. Notably, seven of these genes had fewer exons in Ongole than in *Bos taurus*. Among those genes, PFKP and GPX4 are involved in glycolysis regulation and immune response[17].

Despite these advances, a genome-wide comparison between completely assembled chromosome-level genomes of *Bos indicus* and *Bos taurus* it yet to be fully explored. To bridge that gap, we conducted a detailed chromosome-by-chromosome analysis of the *Bos* genomes available with the National Center for Biotechnology (NCBI), focusing specifically on *Bos indicus* cattle – zebu cattle such as Nelore (Bos_indicus_1.0, RefSeq: GCF_000247795.1) and Gir (ASM293397v1) – and *Bos taurus* cattle, namely Hereford (ARS-UCD2.0, RefSeq: GCF_002263795.3). We compared between members of three distinct pairs, namely *Bos taurus* (Hereford) and *Bos indicus* (Nelore), Hereford and Gir, and Nelore and Gir, to obtain fresh insights into the genetic basis of the disease resistance in *Bos indicus* cattle.

Materials and methods

The flowchart (Fig. 1) outlines the steps in a comparative study of the *Bos* genome. Initially, whole-genome sequence data were collected for three cattle breeds, namely Nelore, Gir (both from *Bos indicus*), and Hereford (*Bos taurus*). The data were compared for the three pairs specified earlier and also shown in Fig. 1, and genomic variations were identified using SyMAP[18], GSAlign, and SyRI. Because SyMAP provides data on syntenic regions and alignment statistics but without detailed data on variation, further analysis was performed using GSAlign and SyRI. Common chromosomal variations predicted by both GSAlign and SyRI were identified, and indicine protein-coding data were used to identify genes and

immunity genes associated with these variations. Lastly, quantitative trait loci (QTLs) were mapped to the genes, and immunity genes were identified within the common chromosomal variations.

Figure 1. Workflow.

Sample preparation

The whole-genome assembly data were downloaded from the NCBI in both GenBank and FASTA formats for the three breeds, namely Nelore (Accession ID: GCA_000247795.2), Gir (Accession ID: GCA_002933975.1), and Hereford (Accession ID: GCA_002263795.3). The assembly statistics were analysed for the number of chromosomes assembled, scaffold N50 values, GC percentage, and genome coverage. The data for all chromosomes from the assemblies of the three breeds were subsequently extracted for further analysis.

Comparative genomic analysis

To perform comparative genomic analysis among the selected breeds of cattle, we compared pairs made from three breeds, one from *Bos taurus*, namely Hereford, and two from *Bos indicus*, namely Nellore and Gir. The three pairs were Hereford–Nelore, Hereford–Gir, and Nelore–Gir. The comparisons focused on variation in terms of the length, number, and function of genetic variants as well as interchromosomal variation as revealed by comparing chromosomes within the same breed, particularly the X chromosome and autosomes, to detect the translocations of genomic elements from the X chromosome to other autosomes within the *Bos* genomes. For the chromosome-by-chromosome analysis, we used three bioinformatics tools, namely Synteny Mapping and Analysis Program (SyMAP), GSAlign[19], and Synteny and Rearrangement Identifier (SyRI)[20]. These tools were employed primarily to analyse the query genome against the reference genome of bovines chromosome by chromosome.

Analysis using SyMAP

SyMAP was used for detecting and analysing syntenic and non-syntenic regions between the query and the reference chromosomes of the selected genomes[19]. SyMAP integrates MUMmer, a high-performance alignment tool using Perl, with the core SyMAP program running on Java. SyMAP requires

genome sequences in FASTA format and corresponding annotation files in GFF3 format for both query and reference genomes. Because an annotation file for Gir genome was unavailable, we used *Bos indicus* (Nelore) as the query genome and *Bos taurus* (Hereford) as the reference genome. The chromosome-level FASTA files and the corresponding annotation files for both the query and the reference genomes were uploaded and SyMAP was run with default settings. SyMAP, in turn, generated synteny maps, dot plots of aligned chromosomes, and basic alignment statistics including coverage, annotations, and gene content. However, SyMAP does not produce a VCF (variant call format) file or any detailed variant information but merely provides a CSV (comma-separated values) file containing data on the percentage of hits, identity, similarity, and gene products from both the query and the reference genomes. Therefore, to identify genes with lower sequence similarity, we used a Python script created in house to extract genes showing less than 60% sequence similarity between the indicine and the taurine groups from the alignment output files.

Analysis using GSAlign

GSAlign efficiently aligns larger query and reference genomes by indexing the reference genome using the Burrows–Wheeler Transform (BWT) in Bowtie2. GSAlign employs a parallelized, divide-and-conquer approach to rapidly construct alignments and identify gapped regions. We used GSAlign to compare between the members of each of the three pairs of breeds as mentioned above and to compare between chromosomes of the same breed to investigate translocations of genetic elements. Specifically, we compared chromosomes X and Y with autosomes in the *Bos* genome to identify any translocation events. The input for GSAlign consisted of whole-genome data on the query and the reference genomes in FASTA format, with the reference file indexed in Bowtie2[21]. We executed the GSAlign script, specifying the appropriate query and reference genomes in FASTA format. GSAlign performed the alignment and generated output files in mutation annotation format (MAF) or clustal format (ALN), as well as VCF files containing the identified sequence variants. An in-house Python script was used to parse these output files to quantify variations such as single nucleotide variants (SNVs), insertions, and deletions.

Analysis using SyRI

SyRI was used for detecting genomic differences between related genomes using whole-genome assemblies. SyRI identifies syntenic regions, structural rearrangements, and local variations[19]. As before, we analysed the three pairs chromosome by chromosome. In each comparison, the former

genome served as the reference and the latter, as the query. We ran SyRI by specifying the appropriate query and the reference genomes in FASTA format within the script file. SyRI generated the alignment file in sequence alignment map (SAM) format using its integrated minimap2 tool and produced output files in tab-separated values (TSV) and VCF formats. The output included annotations on 20 different parameters and a visual representation of genomic variations generated using the integrated plotsr tool[22]. The output files were further processed using a Python script generated in house to count the occurrences of each variant across a given genome.

Validation

The data generated by all the three tools – SyMAP, GSAlign, and SyRI – were validated by cross-comparing the results from each tool and arriving at consensus results through the in-house Python script. Such validation was crucial to predicting and verifying genes and their functions within variable genomic regions. In comparing the genomes, SyMAP provides sequence-similarity information, GSAlign identifies three types of variations, and SyRI detects 20 distinct types of variations. To obtain high-confidence data, we focused on variations that were commonly identified by both GSAlign and SyRI (SyMAP does not provide variant information). Variations identified by two different methods are considered more reliable. Therefore, we focused on insertions, deletions, and substitutions identified by both GSAlign and SyRI for variant annotation, particularly in comparing *Bos taurus* and *Bos indicus* (Nelore), because annotations were unavailable for Gir.

Genome annotation from common chromosomal variations

To annotate common chromosomal variations, we first extracted the variations identified by both GSAlign and SyRI. Using in-house Python scripts, we then identified the genes in the indicine cattle genome (NCBI RefSeq assembly: GCF_000247795.1) that harboured the common chromosomal variations such as insertions, deletions, and substitutions as identified by both tools. Because the RefSeq and annotation for Gir were unavailable in the NCBI database, our focus was specifically on the pairwise comparison at the level of chromosomes between *Bos indicus* (Nelore) and *Bos taurus* (Hereford) for annotating these common variations. We applied eight specific criteria (outlined in Fig. 2) to determine the precise locations of these variants within genes. These criteria were used for filtering and extracting genes from the indicine protein-coding data associated with common chromosomal variations. This

process resulted in an annotated output containing detailed information on the genes affected by the variations.

Figure 2. Rules applied for extracting genes for common chromosomal variation from indicine protein-coding data. These rules were employed to detect genes within indicine protein-coding data that are linked to common chromosomal variations as predicted by two tools, GSAlign and SyRI. G_S, gene start; G_E, V_S, variation start; V_E, variation end. All establish the occurrence of variation in specific genomic regions at the beginning (1) or end (2) of a gene, within a gene (3), between two genes (4), upstream of a gene (5 and 7), and downstream of a gene (6 and 8).

Extracting immunity genes

Using a Python script, we extracted immunity-related genes by using both a keyword-based search and a gene list from the InnateDB database, which lists 1697 innate immunity genes. The following keywords were used for the search: *immunoglobulin, immunoreceptor, autoimmune, Toll-like receptor (TLR), IgG, autoimmune, autophagy, immunogen, immune, innate, T-cell, B-cell, lymphocyte, histocompatibility, CD24, CD4, LY96, IFIT3, PGLYRP1, NKG2D, UL16, leukocyte, cytokine, antimicrobial peptide, beta-defensin 2, IL15, IL2,* and *chemokine*.

Analysis of qualitative trait loci

Using the genomic variations identified between Nelore and Hereford, we proceeded to look at the QTLs associated with the genes that were found to have chromosomal variations including insertions, deletions, and substitutions, as identified by both GSAlign and SyRI. We used the QTL data of the Hereford breed (ARS_UCD1.2) in GFF format obtained from the cattle QTLdb of the Animal QTLdb database (www.animalgenome.org). The loci deposited in the database were related to characters associated with reproduction, production of milk, health, exterior, and meat and carcass. The locus of a qualitative trait is shown as a 4-bp-long segment, referred to as a QTL span. Using an in-house Python script, we identified the QTLs associated with various traits that were located in regions with common chromosomal variations in taurine breeds.

Results

Identification of non-syntenic regions between Nelore and Hereford using SyMAP

SyMAP analysis revealed both syntenic and non-syntenic regions, along with alignment statistics, between the genes of Nelore and those of Hereford. We focused on those genes located in non-syntenic regions with less than 60% sequence similarity. This analysis identified 13 genes across four chromosomes in Nelore that showed less than 60% sequence similarity when compared to Hereford. These genes included one gene each on chromosomes 7 and 23, six on Chromosome 18, and five on Chromosome 19 (Table 1). Notably, four of these genes in Nelore were associated with immune function: VSTM1L, Galectin-9, HDAC5, and POLH.

Chromosomes	Genes in B. indicus (Nelore breed)	Genes in B. taurus	Hit %Id	Hit %Sim
7	X-ray repair cross-complementing 4	adhesion G protein-coupled receptor E2	53	64
18	V-set and transmembrane domain-containing protein 1-like	kelch like family member 36	52	58
	protein phosphatase 1 regulatory subunit 37	retinyl ester hydrolase type 1,	58	64
	zinc finger protein 285	-	52	63
	Membrane-bound transcription factor peptidase,	Rho guanine nucleotide exchange factor 1,	57	59
	Annotation not available	basic helix-loop-helix and HMG-box containing 1	52	56
	D-box binding PAR bZIP transcription factor	SH3 and multiple ankyrin repeat domains 1	59	64
19	ribosomal protein S6 kinase B1	-	56	61
	galectin-9	galectin 9	48	57
	matrix metallopeptidase 28	intraflagellar transport 20	44	53
	histone deacetylase 5	CASC3, exon junction complex subunit	58	65
	matrix metallopeptidase 28	potassium voltage-gated channel subfamily H member 6	53	63
23	DNA polymerase eta, transcript variant X1, X2	kinesin light chain 4	54	64

Table 1. List of the chromosomes that have less than 60% sequence similarity between *B. indicus* (Nelore breed) and *B.taurus* predicted by SyMAP[18]

Interbreed genomic variations identified using GSAlign

The distribution of different variations including insertions, deletions, and substitutions along the chromosomes is shown in Fig. 3.

Figure 3. Interchromosomal variation between members of three pairs of cattle breeds: *B. taurus* (Hereford) and *B. indicus* (Nelore), *B. taurus* (Hereford) and *B. indicus* (Gir), and *B. indicus* (Nelore) and *B. indicu* (Gir), as identified by GSAlign. Black bars show the highest and the lowest numbers of chromosomal variations. Panels (A–I) represent the numbers of single nucleotide variations (SNVs), insertions, and deletions. Comparison between Hereford and Nelore: A, SNVs; B, insertions; C, deletions. Comparison between Hereford and Gir: D, SNVs, E, insertions; F, deletions. Comparison between Nelore and Gir: G, SNVs; H, insertions; I, deletions.

In comparing between Hereford and Nelore, we identified a total of 316,230 insertions, 205,618 deletions, and 6,161,296 substitutions across all 30 chromosomes in Nelore. Among them, Chromosome 1 displayed the highest number of insertions (21,012) and substitutions (345,118), and Chromosome X, the highest number of deletions (15,101) (Fig. 2: A, B, and C). The comparison between Nelore and Gir showed a total of 652,716 insertions, 739,807 deletions, and 12,331,978 substitutions in Gir, with Chromosome 1 showing the highest number of insertions (40,710), substitutions (739,768), and deletions (46,173) (Fig. 2: D, E, and F). The corresponding figures for Hereford and Gir were 673,875, 672,892, and 11,500,574 in Gir, with Chromosome 1 showing the highest number of insertions (44,092) and substitutions (722,089), and Chromosome X, the highest number of deletions (44,455) (Fig. 2: G, H, and I).

Genomic variations in Bos genomes identified using SyRI

Genomic variations across 20 variant annotation parameters were identified using SyRI through chromosome-by-chromosome comparisons between the reference and the query genomes (Table 2). The comparison between *Bos indicus* (Nelore) and *Bos taurus* (Hereford) served to identify a total of 16,199,633 variations across all 30 chromosomes, comprising 20 variants. The majority of these variations (15,543,630) were SNPs, distributed across all 30 chromosomes. Chromosome 1 recorded the highest number of variations (957,165) across 13 variants: CPG (102), CPL (33), DUP (208), DUPAL (269), HDR (2,233), INV (79), INVDP (154), NOTAL (1,222), SYN (498), TRANS (73), INVDPAL (191), INS (23,003), and SNP

(929,100). Chromosome X recorded 13,218 variations across four variant types: INVAL (375), INVTR (55), INVTRAL (106), and DEL (12,682). Chromosome 21 displayed a higher number of SYNAL variants (11,204), and Chromosome 6 showed an increase in TDM (6) and TRANSAL (139) variants. Chromosome 12 also recorded more TDM variants (6). More important, Chromosome X in *B. indicus* (Nelore) was missing a 50 Mbp genomic region, although none of the genes was lost, as can be seen from Fig. 4: C. Additionally, a 3.44 Mbp inversion was found at the head of Chromosome 23, and a 5.627 Mbp inversion on the tail of Chromosome 27 in *B. indicus* relative to *B. taurus*.

Figure 4. Synteny map of all 30 chromosomes, generated using SyRI, comparing between members of three pairs of breeds. Grey, conserved syntenic regions; yellow, inversions; green, translocations; and cyan, duplications. A–D, plots generated using plotsr[22]: A, *B. taurus* and *B. indicus* (Nelore); B, *B. taurus* and *B. indicus* (Gir); C, structural variations of Chromosome X of *B. taurus* and *B. indicus*; D; *B. indicus* (Nelore) and *B. indicus* (Gir).

SyRI Variants	Hereford breed vs Indicus breed	Hereford breed vs Gir breed	Nelore breed vs Gir breed
CPG	1458	2614	2750
CPL	428	254	2091
DUP	3300	2592	6436
DUPAL	4621	3304	8146
HDR	37088	30152	47015
INV	1145	2890	3001
INVAL	3375	3887	6161
INVDP	2129	1310	5147
INVTR	669	717	1346
NOTAL	17914	25116	38247
SYN	7022	7106	9983
SYNAL	53321	37580	55916
TDM	55	43	20
TRANS	1109	1596	2296
TRANSAL	2143	2794	4281
INVDPAL	2650	1666	6121
INVTRAL	1117	1276	2459
INS	329698	674040	556520
DEL	186761	674737	666349
SNP	15543630	14559218	16282402

Table 2. Bos genomes' comparison statistics as per SyRI

In the comparison between *B. indicus* (Gir) and *B. taurus*, 16,032,892 variations were identified covering 20 variants across all 30 chromosomes. Of these, 14,559,218 were SNPs. Chromosome 1 displayed a large number of variations (985,003) including INS (44,430), DEL (43,522), and SNP (897,051). Chromosome X recorded 12,467 variations across 15 variants: CPG (361), DUP (499), DUPAL (699), HDR (2623), INV (213), INVAL (528), INVDP (290), INVTR (98), NOTAL (2453), SYN (552), SYNAL (2839), TRANS (195), TRANSAL (478), INVDPAL (414), and INVTRAL (225). Unlike in Nelore, Chromosome X in Gir gained approximately 10 Mbp of the genomic region compared to that in taurine. Chromosome 6 showed an increase in the CPL variant (19), and Chromosome 4 had more TDM variant (4).

Comparing the two *B. indicus* breeds (Nelore and Gir) revealed 17,706,687 variations comprising 20 variants across all 30 chromosomes in Gir. Of these, 16,282,402 were SNPs. Chromosome 1 had the most variations (1,085,899) in eight variants: CPL (151), HDR (2798), 262 INV (174), SYN (624), SYNAL (3371), INS (36,392) DEL (43,567), and SNP (998,822). Chromosome X showed 10,660 variations across ten variants: CPG (580), DUP (1136), DUPAL (1789), INVAL (594), INVDP (964), INVTR (146), NOTAL (3128), TRANSAL (431), INVDPAL (1385), and INVTRAL (507). In addition, Chromosome X in Gir gained 60 Mbp of genomic material with intergene exchange. Chromosome 6 exhibited more TRANS (167) variants, whereas chromosomes 4 and 8 showed a greater number of TDM variants (3).

Analysis of interchromosomal variations between X or Y chromosome and autosomes within a breed using GSAlign

Interchromosomal genomic variations between autosomes and sex chromosomes (X or Y) were identified within *B. indicus* (Nelore and Gir) and *B. taurus* (Hereford) using GSAlign[19], (Fig. 5).

Figure 5. Sex chromosomes X or Y compared with autosomes between members of three pairs of breeds using GSAlign. Black bars show the highest and the lowest intrachromosomal variation among the breeds. A–C, Chromosome X compared with autosomes in Nellore. **A,** single nucleotide variations; **B,** insertions; **C,** deletions. D–F, Chromosome X compared with autosomes in Hereford. **D**, single nucleotide variations; **E**, insertions; and **F**, deletions. G–I, Chromosome X compared with autosomes in Gir. **G**, single nucleotide variations; **H**, insertions; **I**, deletions.

In Nelore (*B. indicus*), 80,495 insertions, 83,601 deletions, and 1,994,708 substitutions were identified across autosomes as interchromosomal variations. Chromosome 13 showed the highest number of SNVs (77,648), whereas Chromosome 11 showed the most insertions (3359) and deletions (3419). Similarly, in Hereford (*B. taurus*), a comparison between the X chromosome and autosomes revealed 129,623 insertions, 130,482 deletions, and 3,055,863 substitutions. Chromosome 1 recorded the largest number of variations, with 128,548 SNVs and 5438 insertions, whereas Chromosome 13 showed the most deletions (5228). In Gir (*B. indicus*), the analysis identified 137,587 insertions, 140,017 deletions, and 3,193,030 substitutions. Chromosome 3 showed the highest number of SNVs (122,486), insertions (5683), and deletions (5908).

Further analysis of Y chromosome variations in Nelore (Fig. 6) showed Chromosome 30 to have the highest number of SNVs (9290) and deletions (521) whereas Chromosome 17 showed the most insertions (427).

Figure 6. Chromosome Y and autosomes in *Bos indicus* (Nelore) compared using GSAlign. Black bars show the highest and the lowest intrachromosomal variation. **A**, single nucleotide variations; **B**, insertions; **C**, deletions.

Identification of high-confident variants

Variations involving insertions

We identified a significant number of insertions across various chromosomes. Chromosome 1 showed the highest number of insertions, with 3238 insertions occurring in intergenic regions, 1311 insertions affecting 294 genes, and 96 insertions affecting 27 immunity-related genes (Table 3 and Fig. 7).

Figure 7. Insertions in all 30 chromosomes in two genomes (*Bos indicus* and *B. taurus*). Circles indicate the highest number of insertions. **A**, number of insertions between the intergenic region and within gene region; **B**, number of insertions within gene region and genes affected by insertions; **C**, number of insertions in immunity genes and immunity genes affected by insertions.

Chromosome number	Total number of Common variations	No. of. Insertions in intergenic	No. of. Insertions in "within gene"	No. of. Genes have insertions	No. of. Insertions in "at end with intergenic"	No. of. Genes have insertions in "at end with intergenic"	No. of. Insertions in immune genes	No. of. Immune genes have insertion
1	4536	3238	1311	294			33	12
2	2974	2065	909	255			28	8
3	3079	2177	903	296			39	15
4	3602	2403	1198	237			13	3
5	2347	1579	769	248			28	10
6	3254	2472	783	209			29	3
7	2883	2182	705	254			7	4
8	3380	2479	901	239			15	2
9	2349	1831	522	159			18	8
10	2649	1758	893	241			1	1
11	2288	1506	785	249	1	1	7	4
12	2181	1622	559	122			4	2
13	2142	1388	757	248			14	4
14	1925	1484	440	121			8	2
15	2814	1985	826	226			12	5
16	2282	1477	805	181			16	4
17	2027	1474	555	170			3	2
18	2083	1421	667	289			36	11
19	1500	905	607	243			2	2
20	2254	1836	418	106			1	1
21	1772	1333	440	147			9	4
22	1630	987	643	172			2	2
23	959	640	320	119			16	6
24	1078	754	326	95			6	2
25	1000	568	433	150			1	1
26	1511	930	582	129			35	2
27	1268	935	333	85			0	0
28	1270	779	494	108			4	1
29	1954	1388	567	149			6	2
30	3321	2782	539	96			15	4

Table 3. Occurrences of insertions found in the genomic region of *B. indicus* (Nelore breed) across chromosomes

For each chromosome, the GSAlign and SyRI data were compared to identify insertion sites that were identified by both the tools, which werethen compared to the annotated protein-coding data to identify insertions that are present within a coding region, between two coding regions, start, and end position of the gene

Chromosome X showed a large number of insertions, with 2782 insertions in intergenic regions and 539 insertions affecting 96 genes, including 18 insertions within 6 immunity-related genes. It is noteworthy that some chromosomes with fewer insertions affected a greater number of genes. For example, Chromosome 3 had 903 insertions affecting 296 genes, and Chromosome 18 had 667 insertions affecting 289 genes. Conversely, chromosomes with greater number of insertions often affected fewer genes, such as Chromosome 26, in which 47 insertions were confined to only three immunity genes.

A total of 54 genes were identified with insertions longer than 50 bp (Table 4 and Fig. 8). Notably, gene CACNG8 on Chromosome 18 showed the largest length of variation (LOV) at 420 bp, whereas gene SGO2 on Chromosome 2 had an insertion of 356 bp LOV. Gene WIPI1 on Chromosome 19 underwent two insertions: one of 109 bp and the other of 51 bp.

Figure 8. Characters of 30 genes with variations longer than 50 bp. The circle marks the highest number of insertions and the arrows show genes with multiple insertions and deletions.

Table 4. Extracting genes impacted by insertions with lengths of variation exceeding 50 bps.

For each chromosome, the genes affected by insertion variation where the length of the variation exceeded 50 base pairs were identified.

Across all chromosomes, a total of 343 immunity genes were affected by insertions; based on keyword search, 232 of these were classified as innate immunity genes and the remaining 111, as immunity genes. Of the total, 182 immunity genes experienced multiple insertions, whereas the remaining 161 experienced only one insertion. Overall, 78 immunity genes were affected by insertions longer than 10 bp; of these, 52 were innate immunity genes and 26 were keyword-based immunity genes. Notably, 10 genes were affected by insertions longer than 50 bp; of these, eight innate immunity genes – AP3B1 (2 insertions, 182 bp), PARD3 (24 insertions, 78 bp), ASCC3 (19 insertions, 75 bp), RICTOR (3 insertions, 75 bp), NOX4 (13 insertions, 69 bp), BANK1 (17 insertions, 68 bp), RFTN1 (16 insertions, 54 bp), and ANGPT1 (8 insertions, 50 bp) – and two keyword-based immunity genes (DOCK1, with 29 insertions, 98 bp and DOCK5, with 13 insertions, 69 bp) experienced multiple insertions. Some innate immunity genes were affected by a single insertion and yet showed variations longer than 10 bp, including MAPK14 (22 bp), DICER1 (19 bp), SRC (15 bp), CASP6 (13 bp), TPP2 (10 bp), IKBKE (10 bp), and RNASEL (10 bp).

Variations involving deletions

Chromosome 1 experienced the highest number of deletions, 946 in intergenic regions and 428 affecting 157 genes (Table 5 and Fig. 9). Among these 428, 39 deletions affected 11 immunity genes. However, Chromosome 4 showed a greater number of deletions affecting a total of 440 genes. Chromosome 18 experienced the highest number of deletions affecting immunity genes, with 31 deletions affecting 16 immunity-related genes.

Figure 9. Deletions in all 30 chromosomes in two breeds of cattle (Hereford and Nelore). Circles indicate the highest number of deletions. **A**, number of deletions between the intergenic region and within gene region; **B**, number of deletions within gene region and genes affected by the deletions; **C**, number of deletions in immunity genes and immunity genes affected by the deletions.

Chromosome number	Total number of Common variations	No. of. Deletions in intergenic	No. of. Deletions in "within gene"	No. of. Genes have Deletions	No. of. Deletions in immune genes	No. of. Immune genes have Deletion
1	1374	947	428	157	4	3
2	1041	684	357	144	14	4
3	1040	675	363	166	8	6
4	1232	792	440	145	3	1
5	755	469	286	121	9	5
6	1012	710	302	117	5	3
7	896	607	289	134	9	4
8	997	698	299	132	2	1
9	785	541	244	98	5	5
10	851	534	317	132	0	0
11	748	458	295	122	9	3
12	637	464	173	54	1	1
13	754	479	275	120	7	3
14	630	452	178	63	4	1
15	913	629	288	125	7	5
16	649	409	240	105	6	3
17	647	457	191	93	2	1
18	731	472	261	153	21	10
19	567	324	248	130	2	2
20	735	571	164	58	1	1
21	555	407	149	66	1	1
22	547	337	210	93	4	3
23	368	236	136	53	8	3
24	417	285	132	59	0	0
25	363	189	174	79	2	1
26	530	331	197	68	6	2
27	382	291	91	41	0	0
28	397	200	197	70	1	1
29	656	457	199	67	6	3
30	1913	1532	381	52	0	0

Table 5. Occurrences of deletions found in the genomic region of *B. indicus* (Nelore breed) across chromosomes.

For each chromosome, the GSAlign and SyRI data were compared to identify deletion sites that were identified by both the tools, which were then compared to the annotated protein-coding data to identify deletions that are present within a coding region, between two coding regions, start, and end position of the gene.

In addition, we identified 32 genes with deletions longer than 50 bp (Table 6 and Fig. 10). Gene GLTSCR1 on Chromosome 18 experienced the longest deletion, measuring 222 bp. Gene GMDS on Chromosome 23

experienced two deletions, with LOVs of 73 bp and 62 bp. Genes LOC109573590 and LOC109573589 on Chromosome 19 experienced deletions measuring 50 bp, and gene FAT3 on Chromosome 29 experienced deletions measuring 53 bp.

Figure 10. Genes on 30 chromosomes experiencing variations longer than 50 bp in two cattle breeds (Hereford and Nelore). Circles indicate the greatest number of deletions and arrows indicate genes experiencing multiple insertions and deletions.

doi.org/10.32388/VZC1YY

Chromosome number	Locus	Length	Start position of variations	End position of variations	Length of Variation (bp)
2	FMNL2	880	45675552	45675552	57
2	ATP13A2	1127	140139561	140139561	61
4	FAM133B	230	10114522	10114522	89
4	ZNF398	640	116553741	116553741	59
6	SEC24D	992	7376676	7376676	60
8	GLRA3	447	7101529	7101529	61
8	DPYSL2	572	77785234	77785234	54
12	STARD13	1127	27800830	27800830	71
12	UPF3A	512	85352478	85352478	67
13	LOC109567863	938	9749114	9749114	79
13	RIN2	711	39524556	39524556	67
13	LOC109567647	463	53770110	53770110	58
13	LOC109567562	514	82705387	82705387	54
14	SNTG1	517	20409272	20409272	67
15	TTC12	705	22295351	22295351	80
16	PM20D1	503	2629920	2629920	74
17	TMEM144	330	42725758	42725758	56
17	RAD9B	424	57165073	57165073	55
18	NECAB2	540	9508851	9508851	73
18	GLTSCR1	1470	54470335	54470335	222
19	ASIC2	542	16772459	16772459	64
19	LOC109573590	397	36731830	36731830	50
19	LOC109573589	933	36731830	36731830	50
21	SV2B	683	15272643	15272643	72
22	IFT122	1376	57821097	57821097	77
23	LOC109577323	462	20481417	20481417	64
23	F13A1	732	49703992	49703992	58
23	SLC22A23	528	51267159	51267159	57
23	GMDS	328	52498261	52498261	73
23	GMDS	328	52550614	52550614	62
24	WDR7	1487	58574505	58574505	68
26	XPNPEP1	666	30773879	30773879	67
26	VAX1	265	37976117	37976117	77
28	ARHGAP22	712	42828366	42828366	72
29	FAT3	3465	2074802	2074802	53

Table 6. Extracting genes impacted by deletions with lengths of variation exceeding 50 bps.

For each chromosome, the genes affected by deletion variation where the length of the variation exceeded 50 base pairs were identified.

Across all 30 chromosomes, 183 immunity genes were affected by deletions, of which 119 were identified as innate immunity genes, whereas 64 were keyword-based immunity genes. Among these, the innate immunity genes PRKCE and STIM1 each experienced 12 deletions, which was the maximum number. A total of 22 immunity genes were affected by deletions longer than 10 bp, of which 16 were innate immunity genes. No gene was affected by deletions longer than 50 bp. The keyword-based immunity gene BCL7B experienced two deletions, affecting 46 bp, and the innate immunity gene NFATC2 was

affected by two deletions spanning 39 bp. Several innate immunity genes were affected by single deletions longer than 10 bp, including MERTK (13 bp), GAS6 (18 bp), and C2 (28 bp). Notably, no immunity gene on Chromosome X was affected by deletions.

Furthermore, 11 innate immunity genes experienced multiple deletions, each longer than 10 bp: PTK2B (3 deletions affecting 33 bp), RFTN1 (8, 27 bp), ARHGAP15 (6, 24 bp), TRIM9 (2, 23 bp), PRKCE (12, 21 bp), FSTL1 (5, 19 bp), SMAD6 (2, 19 bp), STIM1 (12, 17 bp), DMBT1 (3, 17 bp), TRIM66 (2, 15 bp), ITGB3 (2, 15 bp), and ITPR1 (10, 11 bp).

Variations involving substitutions

Chromosome 1 showed the highest number of substitutions in intergenic regions, with a total of 122,906 substitutions. Chromosome 4 showed the highest number of substitutions within genes, totalling 70,079 substitutions (Table 7 and Fig. 11).

Figure 11. Substitutions across all 30 chromosomes in three breeds of cattle. Circles indicate the highest number of substitutions. **A**, substitutions between the intergenic region and within gene region; **B**, genes affected by substitutions; **C**, number of substitutions in immunity genes; **D**, immunity genes affected by substitutions.

doi.org/10.32388/VZC1YY

Chromosome number	Total number of Common variations	No. of. substitution in intergenic	No. of. substitution in "within gene"	No. of. Genes have substitution	No. of. substitution in "at start	No. of. Genes have substitution in at start of the locus	No. of. substitution in "at end"	No. of. Genes have substitution in at end of the locus	No. of. substitution in immune genes	No. of. Immune genes have substitution
1	176517	122906	54212	649					801	16
2	169343	112480	56883	708			1	1	2705	22
3	151944	99487	52673	978	1	1			2892	48
4	174424	104379	70079	557					1344	6
5	160380	101849	58634	944	3	3	1	1	1131	29
6	134491	89870	44640	423					1041	9
7	143012	94659	48411	984			3	3	771	28
8	143493	96798	46760	561					977	13
9	113634	82735	30937	376					838	13
10	141847	90039	51955	743	2	2	3	3	244	8
11	146664	93973	52922	768			3	3	987	27
12	130602	91076	39575	292					64	3
13	118758	75620	43213	647	4	4	1	1	622	10
14	107598	76541	31008	333			1	1	224	3
15	107493	69907	37807	751	1	1	1	1	1006	17
16	110144	68037	42132	466	2	2	1	1	547	11
17	103263	68454	34946	473			2	2	241	9
18	95599	58378	37538	995	1	1	3	3	1944	39
19	92065	52924	39307	997			1	1	231	19
20	99068	75082	24005	241					173	6
21	97891	69994	27876	379					279	9
22	85751	50099	35721	434			1	1	775	21
23	72530	46280	26420	547					1230	38
24	108402	75157	33376	256					364	3
25	66454	36872	29651	579					321	15
26	66923	40825	26085	269					667	3
27	55832	39689	16197	170					3	1
28	68044	38752	29566	238	2	2			429	4
29	83159	56455	26838	509	1	1	1	1	210	10
30	72766	54840	17925	304	1	1			374	21

Table 7. Occurrences of substitutions found in the genomic region of *B. indicus* (Nelore breed) across chromosomes.

For each chromosome, the GSAlign and SyRI data were compared to identify substitution sites that were identified by both the tools, which were then compared to the annotated protein-coding data to identify substitutions that are present within a coding region, between two coding regions, start, and end position of the gene.

Many genes were affected by substitutions: 995 genes on Chromosome 18 and 997 genes on Chromosome 19. Substitutions were also found both upstream and downstream of genes. Chromosomes 7, 10, 11, and 18 each experienced three substitutions affecting the downstream region of three genes, and Chromosome 13 experienced four substitutions affecting the upstream region of four genes. Chromosome 3 showed

the highest number of substitutions in immunity genes, with 78 immunity genes affected by 3497 substitutions.

A total of 5188 genes showed more than 50 substitutions each (Table 8), Chromosome 5 topping the list at 48,935 substitutions across 271 genes. Among the immunity genes, 301 were affected by substitutions across all 30 chromosomes: 209 were innate immunity genes and 92 were identified as keyword-based immunity genes, and all 301 experienced more than 50 substitutions. Chromosome 2 had the highest number of substitutions, with 3909 substitutions affecting 49 immunity genes. Gene SPAG16, which had been predicted, based on keywords, to be conferring immunity, experienced the highest number (1457) of substitutions, whereas the innate immunity gene PLXNA4 experienced 1338 substitutions. The innate immunity genes PRKCE, EDIL3, and PRKCA each experienced more than a thousand substitutions.

Table 8. Exploring the immune genes affected by substitution events with more than 50 substitutions.

For each chromosome, the immune genes impacted by substitution variation where the length of the variation exceeded 50 substitution counts were identified.

Notably, 93 immunity genes were affected by all three types of variations, namely insertions (INS), deletions (DEL), and substitutions (SUB). Of these, 62 were innate and 31 were predicted. Thus, we identified 12 innate immunity genes that had been affected by variations amounting to more than 500 bps: RFTN1 (16 INS affected by 54 bps, 8 DEL affected by 27 bps, and 433 SUB) on Chromosome 1 affected by 514 bps; ARHGAI (14 INS affected by 25 bps, 6 DEL affected by 24 bps, and 534 SUB) on Chromosome 2 affected by 583 bps; PLXNA4 (14 INS affected by 34 bps, 6 DEL affected by 9 bps, and 1338 SUB) on Chromosome 4 affected by 1381 bps; EDIL3 (17 INS affected by 44 bps, 4 DEL affected by 6 bps, and 1147 SUB) on Chromosome 7 affected by 1197 bps; AP3B1 (2 INS affected by 182 bps, 1 DEL affected by 1 bp, and 488 SUB) on Chromosome 10 affected by 671 bps; PRKCE (16 INS affected by 43 bps, 12 DEL affected by 21

bps, and 1184 SUB) on Chromosome 11 affected by 1248 bps; PARD3 (24 INS affected by 78bps, 2 DEL affected by 2 bps, and 703 SUB) on Chromosome 13 affected by 783 bps; ZFPM2 (12 INS affected by 18 bps, 5 DEL affected by 6 bps, and 649 SUB) and ANGPT1 (8 INS affected by 50 bps, 2 DEL affected by 3 bps, and 495 SUB) on Chromosome 14 affected by 673 bps; PTPRC (5 INS affected by 5 bps, 1 DEL affected by 1 bp, and 590 SUB) on Chromosome 16 affected by 596 bps; PRKCA (9 INS affected by 10 bps, 2 DEL affected by 2 bps, and 1089 SUB) on Chromosome 11 affected by 1101 bps; and TCF4 (3 INS affected 4 bps, 2 DEL affected by 2 bps, and 819 SUB) on Chromosome 24 affected by 825 bps.

Mapping of quantitative trait loci

We mapped quantitative trait loci (QTLs) on genes with common chromosomal variations, identified using both GSAlign and SyRI, across all 30 chromosomes.

For insertions, we mapped QTLs for 3609 genes, including 218 immunity genes (159 innate and 69 keyword-based). These genes were associated with various QTLs: health (889 genes), reproduction (1314 genes), milk (2129 genes), meat and carcass (1338 genes), and the exterior (680 genes). Among the immunity genes, 67 were linked to traits associated with health; 126, to milk; 76, to meat and carcass; 68, to production; 70 to reproduction; and 47, to exterior traits.

For deletions, we mapped QTLs on 2005 genes, including 131 immunity genes (92 innate and 47 keyword-based). These genes were associated with various traits: health (540 genes), reproduction (767 genes), milk (1,210 genes), meat and carcass (821 genes), and exterior traits (424 genes). Among the immunity genes, 42 were linked to traits associated with health; 76, to milk; 54, to meat and carcass; 49 to reproduction; and 32, to the exterior.

For substitutions, we mapped QTLs on 7883 genes, including 546 immunity genes (368 innate and 209 keyword-based). These genes were linked to health (1602 genes), reproduction (2582 genes), milk (4387 genes), meat and carcass (2364 genes), and the exterior (1188 genes). Additionally, among 3411 genes with more than 50 substitutions, including 213 immunity genes (153 innate and 70 keyword-based), we found associations with health (866 genes), reproduction (1287 genes), milk (2030 genes), meat and carcass (1311 genes), and the exterior (699 genes). Among these immunity genes, 140 were mapped to traits related to health; 285, to milk; 154, to meat and carcass; 166, to reproduction; and 85, to the exterior.

Discussion

The current study sought to explore genomic variations between two *Bos indicus* breeds (Nelore and Gir) and one *Bos taurus* breed (Hereford) to identify unique features related to immunity and QTLs in indicine cattle.

Identification of non-syntenic immunity genes in indicine and taurine cattle

Our synteny analysis revealed reduced sequence similarity across 13 genes in indicine cattle compared to their taurine counterparts. The variations occurred in four immunity-related genes, underscoring the potential for distinct immune responses by indicine cattle. One such gene, VSTM1L, is expressed in multiple immune cells, including T cells, B cells, natural killer cells, and dendritic cells. This gene plays a vital role in regulating immune cell activation, differentiation, and the expression of immunity-related proteins such as antibodies and cell surface receptors[23]. Another gene, Galectin-9, which is evolutionarily conserved, has been implicated in innate immune responses to bacterial infections and also plays a role in modulating cytotoxic immune responses by suppressing the activity of T cells and natural killer cells.[24]. Additionally, Histone deacetylase 5 (HDAC5) was identified as a gene involved in the regulation of cytokine signalling pathways and immune cell activity, as well as the development of immune-related diseases[25]. DNA polymerase eta (POLH), another gene identified in this study, is a specialized polymerase crucial to error-free bypass of DNA damage, is highly expressed in rapidly proliferating immune cells such as B and T lymphocytes, and plays a key role in the development and function of the immune system[26]. These findings highlight the genetic distinctions between indicine and taurine cattle, particularly in genes linked to immunity, suggesting the genetic make-up of the immune system in indicine cattle.

Comparison of chromosomes in different breeds

Indicine cattle showed a significantly higher number of variations on chromosomes 1 and X than those seen in taurine cattle. Similarly, comparisons between Nelore and Gir, both indicine breeds, also showed a higher number of variations on Chromosome 1. These findings suggest that compared to other chromosomes, chromosomes 1 and X exhibit notable genomic differences. Furthermore, substitutions were more frequent than insertions and deletions across all comparisons. A particularly noteworthy finding was the loss of approximately 50 Mbp of genomic regions on Chromosome X in Nelore. Despite 381 deletions across 52 genes and 1532 deletions in intergenic regions in Chromosome X, no deletions

were observed in immunity-related genes within Nelore. Chromosome X is the second largest in the genome and harbours many genes that can influence various phenotypes[27][28], a feature attributed to higher linkage disequilibrium than that seen in autosomes[29]. To investigate whether the lost genomic regions from Chromosome X could have translocated to other autosomes, we compared Chromosome X with other chromosomes in Nelore. The comparison revealed that compared to Chromosome X, Chromosome 13 accounted for a higher number of substitutions, and Chromosome 11 accounted for a greater number of insertions and deletions. In addition, intrachromosomal insertions, deletions, and substitutions between Chromosome X and autosomes within each breed suggest a possibility of genome dose compensation among cattle breeds. A genome-wide association study (GWAS) aimed at identifying chromosome regions linked to sexual precocity in Nelore cattle found that the highest proportion of genetic variation for early pregnancy and scrotal circumference was concentrated on specific chromosomes. Large effects for early pregnancy were located on chromosomes 5, 6, 7, 14, 18, 21, and 27, whereas those for scrotal circumference were located on chromosomes 4, 8, 11, 13, 14, 19, 22, and 23[30]. Notably, the region on Chromosome 13 associated with scrotal circumference had previously been linked to the phenotype of fat layer thickness in Nelore cattle[29].

Genetic variations in immunity genes in Nelore

By identifying common chromosomal variations, we were able to pinpoint high-confidence variants, particularly highlighting the impact on immunity genes. Notably, 78 immunity genes were affected by insertions longer than 10 bp. Among these, 52 were classified as innate immunity genes, as identified through the InnateDB database, and the remaining 26 were identified as immunity-related genes based on keyword searches. Ten immunity genes were affected by longer insertions, greater than 50 bp, and eight of these were innate immunity genes, namely AP3B1, PARD3, ASCC3, RICTOR, NOX4, BANK1, RFTN1, and ANGPT1. Among these, AP3B1 is probably involved in regulating coat colour in cattle and other domesticated mammals[31][32][33]. Several genes that activate the immune system, such as CD59[34], CDH9[35], PROCR[36], RLBP1[37], BOLA[38], and STOM[32], respond to environmental stress and are strongly associated with responses to parasitic diseases as well[39]. Nelore macrophages control the intracellular replication of *Brucella abortus* more effectively, suggesting that, compared to Holstein, Nelore may possess greater natural resistance to brucellosis[40].

Additionally, we also identified 22 immunity genes affected by deletions longer than 10 bp, 16 of these being innate immunity genes. Regarding variations in the substitutions, 301 immunity genes were

affected across all chromosomes, including 209 innate immunity genes and 92 immunity genes identified through keyword-based searches. Each of the 301 immunity genes experienced more than 50 substitutions. Notably, the innate immunity genes PRKCE, EDIL3, and PRKCA were particularly affected: each had experienced over a thousand substitutions.

It is also noteworthy that 93 immunity genes were affected by all the three types of variations, namely insertions, deletions, and substitutions. Of these, 62 were innate immunity genes, of which the following 12 had experienced variations longer than 500 bp: RFTN1, ARHGAI, PLXNA4, EDIL3, AP3B1, PRKCE, PARD3, ZFPM2, ANGPT1, PTPRC, PRKCA, and TCF4. We identified variations in genes such as IFNGR1, IFNAR1, IGSF21, IL23R, and others in Nelore. Some of these gene variations have been linked to resistance to viral infections in mice and humans[41], although information on bovines remains scarce. A comparative analysis of innate and adaptive immunity in European breeds (Holstein, Brown Swiss, and Hereford) and zebu breeds (Gir, Nelore, and Guzera) was conducted by evaluating their peripheral blood leukocyte profiles, including monocytes, eosinophils, lymphocytes, CD4+ and CD8+ T cells, and CD21+ B cells. Zebu breeds, particularly Nelore, showed higher frequencies of cells involved in innate immunity, such as monocytes and non-T, non-B cells, which may explain the greater resistance shown by Nelore to some infectious and parasitic diseases[40]. Additionally, BOLA genes located on Chromosome 23, part of the bovine major histocompatibility complex (MHC) class II gene family, are promising candidates for the immune response and adaptation to tropical environments[42].

Identification of QTLs by mapping immunity genes

We mapped genes affected by common chromosomal variations to taurine QTLs and identified associations with 3609 genes, including 218 immunity genes affected by insertions, 2005 genes including 131 immunity genes affected by deletions, and 7883 genes including 546 immunity genes affected by substitutions. A total of 36 QTLs were identified across chromosomes 1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 18, 19, 20, 21, 22, 24, 25, and 26 (UMD 3.1), including eight candidate genes related to feed efficiency in Nelore[43]. Additionally, 2531 CNVRs (95.5%) in Nelore overlapped by at least one base with QTLs from the Cattle Animal QTLdb. In most cases, these CNVRs were completely contained within the QTL regions. Among them, 482 polymorphic CNVRs overlapped with 2310 QTLs, corresponding to 282 traits, including several related to milk production. Furthermore, ten CNVRs overlapped with regions linked to meat tenderness in Nelore[44][45]. Notably, we observed that many of the immunity genes mapped to QTLs were associated with milk, health, reproduction, meat and carcass quality, and the

exterior. Specifically, we identified 67 genes affected by insertions, 42 genes affected by deletions, and 140 genes affected by substitutions that were mapped to health-related QTLs. These findings suggest that variations in immunity genes are critical not only to health but may also influence other economically important traits in cattle, such as milk production and meat quality. We also identified 46 genes within selection signature regions—genes that modulated 24 key biological processes in Nelore heifers[46]. Five of the genes – TBC1D23, NIT2, CELA1, BMPR1B, and HEXB – affect immune system pathways and inflammatory responses, and SPTBN1 is involved in the MAPK signalling pathway. These pathways are activated by inflammatory interleukins and cytokines released from damaged tissues[47][48][49]. Banos et al.[47] also observed correlations between traits related to immunity and those related to reproduction in dairy cattle.

Conclusion

A comprehensive comparative genomic analysis of the *Bos* genome is presented, emphasizing the identification of syntenic and non-syntenic regions, genomic variations between breeds, translocations in sex chromosomes, and genomic variations in innate immunity genes and QTLs. Our findings highlight significant genomic variations on chromosomes 1 and X seen in comparisons between two members of each pair of breeds, and non-syntenic regions identified in four immunity genes in Nelore, namely VSTM1L, Galectin-9, HDAC5, and POLH. More specifically, chromosomes 13 and 11 in Nelore recorded a greater number of SNVs and insertions, respectively, suggesting possible translocations involving Chromosome X. The chromosome-by-chromosome analysis comparing Nelore with Hereford revealed variations in innate immunity genes in Nelore, along with their association with QTLs related to milk and health. These findings not only extend our understanding of the genetic basis for disease resistance in Indian cattle but also provide valuable information for developing future breeding strategies that could enhance immunity and overall productivity of indicine cattle.

Statements and Declarations

Funding

Indian Council of Agricultural Research-National Agricultural Science Fund (F. No.NASF/SUTRA-02/2022-23/50) to RMY, DS, and SKO.

Acknowledgements

The authors thank ICAR-National Agricultural Science Fund, SASTRA Deemed to be university, and ICAR-National Dairy Research Institute. The authors thank Mr. Pulkit Anupam Srivastava for discussions related to the python scripts.

References

1. ^*Serrano GM, Egito AA, McManus C, Mariante A da (2004). "Genetic diversity and population structure of Brazilian native bovine breeds." Pesquisa Agropecuária Brasileira. 39(6): 543–549. doi:10.1590/s0100-204x2004000600005.*
2. ^*Verma S, Thakur A, Katoch S, Shekhar C, Wani AH, Kumar S, Dohroo S, Singh G, Sharma M (2017). "Differences in innate and adaptive immune response traits of Pahari (Indian non-descript indigenous breed) and Jersey crossbred cattle." Veterinary Immunology and Immunopathology. 192: 20–27. doi:10.1016/j.vetimm.2017.09.003.*
3. ^*P. Hansen (2004). "Physiological and cellular adaptations of zebu cattle to thermal stress." Animal Reproduction Science. doi:10.1016/s0378-4320(04)00066-1.*
4. ^*Okeyo Mwai, Olivier Hanotte, Young-Jun Kwon, Seoae Cho (2015). "- invited review - african indigenous cattle: Unique genetic resources in a rapidly changing world." Asian-Australasian Journal of Animal Sciences. 28 (7): 911–921. doi:10.5713/ajas.15.0002r.*
5. ^*Pramod RK, Velayutham D, PK S, PS B, Zachariah A, Zachariah A, B C, SS S, P G, Dhinoth Kumar B, Iype S, Gupta R, Santhosh S, Thomas G (2018). "Complete mitogenome reveals genetic divergence and phylogenetic relationships among Indian cattle (Bos indicus) breeds." Animal Biotechnology. 30(3): 219–232. doi:10.1080/10495398.2018.1476376.*
6. ^*Sharma R, Kishore A, Mukesh M, Ahlawat S, Maitra A, Pandey AK, Tantia MS (2015). "Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers." BMC Genetics. 16(1). doi:10.1186/s12863-015-0221-0.*
7. ^*Shivakumara PN, Aravindakshan TV, Naicy T, Anilkumar K, Uma R (2018). "Molecular characterization and differential mRNA expression profiling of Toll - like receptor - 2 gene in Vechur (Bos indicus) and crossbred (Bos indicus x Bos taurus) cattle of Kerala in response to anthrax vaccination." Meta Gene. 16: 15–20. doi:10.1016/j.mgene.2018.01.003.*

8. ^ Charles A. Janeway Jr., Ruslan Medzhitov (2002). "Innate immune recognition." Annual Review of Immunology. 20 (1): 197–216. doi:10.1146/annurev.immunol.20.083001.084359.
9. ^ Takeda K, Kaisho T, Akira S (2003). "Toll-like receptors." Annual Review of Immunology. 21(1): 335–376. doi:10.1146/annurev.immunol.21.120601.141126.
10. ^ Anamika Mishra, Periyasamy Vijayakumar, Ashwin Ashok Raut (2017). "Emerging avian influenza infections: Current understanding of innate immune response and molecular pathogenesis." International Reviews of Immunology. 36 (2): 89–107. doi:10.1080/08830185.2017.1291640.
11. ^ Janez Ferluga, Hadida Yasmin, Mohammed N. Al-Ahdal, Sanjib Bhakta, Uday Kishore (2020). "Natural and trained innate immunity against mycobacterium tuberculosis." Immunobiology. 225 (3): 151951. doi:10.1016/j.imbio.2020.151951.
12. ^ Vlasova AN, Saif LJ (2021). "Bovine Immunology: Implications for Dairy Cattle." Frontiers in Immunology. 12. doi:10.3389/fimmu.2021.643206.
13. ^ Amod Kumar, Gyanendra Kumar Gaur, Manjit Panigrahi, Muhasin Asaf V. N., Bhuvana Priya, Jay Prakash Gupta, Ravi Kumar Gandham (2022). "Comparative gene expression profile in circulating pbmcs ofbos indicusand crossbred cattle to understand disease tolerance mechanism." Animal Biotechnology. 34 (4): 1594–1602. doi:10.1080/10495398.2022.2043883.
14. ^ anavez, F. C., Luche, D. D., Stothard, P., Leite, K. R., Sousa-Canavez, J. M., Plastow, G., Meidanis, J., Souza, M. A., Feijao, P., Moore, S. S., & Camara-Lopes, L. H. (2012). Genome sequence and assembly of Bos indicus. Journal of Heredity, 103(3), 342–348. doi:10.1093/jhered/esr153.
15. ^ Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009). "A whole-genome assembly of the domestic cow, Bos Taurus." Genome Biology. 10(4). doi:10.1186/gb-2009-10-4-r42.
16. ^ Vijayakumar P, Singaravadivelan A, Mishra A, Jagadeesan K, Bakyaraj S, Suresh R, Sivakumar T (2022). "Whole-genome comparative analysis reveals genetic mechanisms of disease resistance and heat tolerance of tropicalBos indicuscattle breeds." Genome. 65(4): 241–254. doi:10.1139/gen-2021-0030.
17. ^ Abhisek Chakraborty, Manohar S. Bisht, Rituja Saxena, Shruti Mahajan, Joby Pulikkan, Vineet K. Sharma (2023). "Genome sequencing and de Novo and reference-based genome assemblies of Bos indicus breeds." Genes & Genomics. 45 (11): 1399–1408. doi:10.1007/s13258-023-01401-w.
18. ^ a, b Soderlund C, Nelson W, Shoemaker A, Paterson A (2006). "SyMAP: A system for discovering and viewing syntenic regions of FPC maps." Genome Research. 16(9): 1159–1168. doi:10.1101/gr.5396706.

19. ^ a b c d Hsin-Nan Lin, Wen-Lian Hsu (2020). "GSALIGN: An efficient sequence alignment tool for intra-species genomes." BMC Genomics. 21 (1). doi:10.1186/s12864-020-6569-1.

20. ^ Manish Goel, Hequan Sun, Wen-Biao Jiao, Korbinian Schneeberger (2019). "SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies." Genome Biology. 20 (1). doi:10.1186/s13059-019-1911-0.

21. ^ Goel, M. et al. (2019) 'Syri: Finding genomic rearrangements and local sequence differences from whole-genome assemblies', Genome Biology, 20(1). doi:10.1186/s13059-019-1911-0.

22. ^ a b Goel, M. and Schneeberger, K. (2022) 'PLOTSR: Visualizing structural similarities and rearrangements between multiple genomes', Bioinformatics, 38(10), pp. 2922–2926. doi:10.1093/bioinformatics/btac196.

23. ^ Oludada OE, Idowu TO, Jeon Y, Choi I (2022). "VSTM5 is a novel immune checkpoint that promotes oral tolerance of cell-mediated and antibody responses." Biochemical and Biophysical Research Communications. 635: 283–290. doi:10.1016/j.bbrc.2022.10.057.

24. ^ Schlichtner S, Meyer NH, Yasinska IM, Aliu N, Berger SM, Gibbs BF, Fasler-Kan E, Sumbayev VV (2021). "Functional role of galectin-9 in directing human innate immune reactions to Gram-negative bacteria and T cell apoptosis." International Immunopharmacology. 100: 108155. doi:10.1016/j.intimp.2021.108155.

25. ^ Paul V. Licciardi, Tom C. Karagiannis (2012). "Regulation of immune responses by histone deacetylase inhibitors." ISRN Hematology. 2012: 1–10. doi:10.5402/2012/690901.

26. ^ Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., & Reynaud, C.-A. (2005). Contribution of DNA polymerase η to immunoglobulin gene hypermutation in the mouse. The Journal of Experimental Medicine, 201 (8), 1191–1196. doi:10.1084/jem.20050292.

27. ^ Russell E Lyons, Nguyen To Loan, Leanne Dierens, Marina R S Fortes, Matthew Kelly, Sean S McWilliam, Yutao Li, Rowan J Bunch, Blair E Harrison, William Barendse, Sigrid A Lehnert, Stephen S Moore (2014). "Evidence for positive selection of taurine genes within a QTL region on chromosome x associated with testicular size in Australian brahman cattle." BMC Genetics. 15 (1). doi:10.1186/1471-2156-15-6.

28. ^ T. Johnson, M. Keehan, C. Harland, T. Lopdell, R.J. Spelman, S.R. Davis, B.D. Rosen, T.P.L. Smith, C. Couldrey (2019). "Short communication: Identification of the pseudoautosomal region in the Hereford Bovine Reference Genome Assembly ars-UCD1.2." Journal of Dairy Science. 102 (4): 3254–3258. doi:10.3168/jds.2018-15638.

29. ^ a b Gregório Miguel Ferreira de Camargo, Laercio R Porto-Neto, Matthew J Kelly, Rowan J Bunch, Sean M McWilliam, Humberto Tonhati, Sigrid A Lehnert, Marina R S Fortes, Stephen S Moore (2015). "Non-synonymous mutations mapped to chromosome x associated with Andrological and growth traits in beef cattle." BMC Genomics. 16 (1). doi:10.1186/s12864-015-1595-0.

30. ^Irano, N., de Camargo, G. M., Costa, R. B., Terakado, A. P., Magalhães, A. F., Silva, R. M., Dias, M. M., Bignardi, A. B., Baldi, F., Carvalheiro, R., de Oliveira, H. N., & de Albuquerque, L. G. (2016). Genome-wide association st udy for indicator traits of sexual precocity in Nellore cattle. PLOS ONE, 11(8). doi:10.1371/journal.pone.015950 2.

31. ^Da Wei Huang, Brad T Sherman, Richard A Lempicki (2008). "Systematic and integrative analysis of large gene lists using David Bioinformatics Resources." Nature Protocols. 4 (1): 44–57. doi:10.1038/nprot.2008.211.

32. a, bJaemin Kim, Olivier Hanotte, Okeyo Ally Mwai, Tadelle Dessie, Salim Bashir, Boubacar Diallo, Morris Aga ba, Kwondo Kim, Woori Kwak, Samsun Sung, Minseok Seo, Hyeonsoo Jeong, Taehyung Kwon, Mengistie Ta ye, Ki-Duk Song, Dajeong Lim, Seoae Cho, Hyun-Jeong Lee, Duhak Yoon, Sung Jong Oh, Stephen Kemp, Hak-Kyo Lee, Heebal Kim (2017). "The genome landscape of indigenous African cattle." Genome Biology. 18 (1). d oi:10.1186/s13059-017-1153-y.

33. ^George E. Liu, Derek M. Bickhart (2012). "Copy number variation in the cattle genome." Functional & Integ rative Genomics. 12 (4): 609–624. doi:10.1007/s10142-012-0289-9.

34. ^E. K. F. Chan, S. H. Nagaraj, A. Reverter (2010). "The evolution of tropical adaptation: Comparing taurine a nd zebu cattle." Animal Genetics. 41 (5): 467–477. doi:10.1111/j.1365-2052.2010.02053.x.

35. ^Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, Gredler B, Bapst B, Flury C (2017). "Population structure and genomic inbreeding in nine Swiss dairy cattle populations." Genetics Se lection Evolution. 49(1). doi:10.1186/s12711-017-0358-6.

36. ^Sorbolini S, Marras G, Gaspa G, Dimauro C, Cellesi M, Valentini A, Macciotta NP (2015). "Detection of selecti on signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but diffe rent selection histories." Genetics Selection Evolution. 47(1). doi:10.1186/s12711-015-0128-2.

37. ^Smetko A, Soudre A, Silbermayr K, Müller S, Brem G, Hanotte O, Boettcher PJ, Stella A, Mészáros G, Wurzin ger M, Curik I, Müller M, Burgstaller J, Sölkner J (2015). "Trypanosomosis: potential driver of selection in Afri can cattle." Frontiers in Genetics. 6. doi:10.3389/fgene.2015.00137.

38. ^O. Oddgeirsson, S. P. Simpson, A. L. G. Morgan, D. S. Ross, R. L. Spooner (1988). "Relationship between the b ovine major histocompatibility complex (BOLA), erythrocyte markers and susceptibility to mastitis in Icela ndic cattle." Animal Genetics. 19 (1): 11–16. doi:10.1111/j.1365-2052.1988.tb00783.x.

39. ^Chugang Mei, Linsheng Gui, Jieyun Hong, Sayed Haidar Abbas Raza, Chen Aorigele, Wanqiang Tian, Matt hew Garcia, Yaping Xin, Wucai Yang, Song Zhang, Linsen Zan (2021). "Insights into adaption and Growth Ev olution: A Comparative Genomics study on two distinct cattle breeds from northern and southern China." Molecular Therapy - Nucleic Acids. 23: 959–967. doi:10.1016/j.omtn.2020.12.028.

40. ^ a, b A.A. Macêdo, A.P.V. Marciano, L.M. Rocha, J.R.F. Alves-Júnior, A.M.C. Faria, J.F.F. Bittar, M.S.S. Araújo, R.L. S antos, O.A. Martins-Filho (2013). "Comparative phenotypic profile of subpopulations of peripheral blood leu kocytes in European (Bos Taurus Taurus) and zebu cattle (Bos Taurus indicus)." Genetics and Molecular Re search. 12 (4): 6838–6849. doi:10.4238/2013.december.19.2.

41. ^ Sadler AJ, Williams BR (2008). "Interferon-inducible antiviral effectors." Nature Reviews Immunology. 8 (7): 559–568. doi:10.1038/nri2314.

42. ^ Amanda Marchi Maiorano, Diercles Francisco Cardoso, Roberto Carvalheiro, Gerardo Alves Fernandes Júni or, Lúcia Galvão de Albuquerque, Henrique Nunes de Oliveira (2022). "Signatures of selection in Nelore cattl e revealed by whole-genome sequencing data." Genomics. 114 (2): 110304. doi:10.1016/j.ygeno.2022.110304.

43. ^ Priscila SN de Oliveira, Aline SM Cesar, Michele L do Nascimento, Amália S Chaves, Polyana C Tizioto, Ry mer R Tullio, Dante PD Lanna, Antonio N Rosa, Tad S Sonstegard, Gerson B Mourao, James M Reecy, Dorian J Garrick, Maurício A Mudadu, Luiz L Coutinho, Luciana CA Regitano (2014). "Identification of genomic regi ons associated with feed efficiency in nelore cattle." BMC Genetics. 15 (1). doi:10.1186/s12863-014-0100-0.

44. ^ Silva VH, Regitano LC, Geistlinger L, Pértille F, Giachetto PF, Brassaloti RA, Morosini NS, Zimmer R, Coutin ho LL (2016). "Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle." PLOS ONE. 11(6): e0157711. doi:10.1371/journal.pone.0157711.

45. ^ Tizioto PC, Decker JE, Taylor JF, Schnabel RD, Mudadu MA, Silva FL, Mourão GB, Coutinho LL, Tholon P, So nstegard TS, Rosa AN, Alencar MM, Tullio RR, Medeiros SR, Nassu RT, Feijó GL, Silva LO, Torres RA, Siqueira F, Regitano LC (2013). "Genome scan for meat quality traits in Nelore beef cattle." Physiological Genomics. 4 5(21): 1012–1020. doi:10.1152/physiolgenomics.00066.2013.

46. ^ Donicer E. Montes, Camila U. Braz, André M.F. Ribeiro, Lígia Cavani, Marina M.D. Barbero, Lucia G. Albuqu erque, Rogério A. Curi, Henrique N. Oliveira (2019). "Selection signatures in candidate genes and QTL for rep roductive traits in Nellore heifers." Animal Reproduction Science. 207: 1–8. doi:10.1016/j.anireprosci.2019.06. 004.

47. ^ a, b Georgios Banos, Eileen Wall, Michael P. Coffey, Ainsley Bagnall, Sandra Gillespie, George C. Russell, Tom N. McNeilly (2013). "Identification of immune traits correlated with dairy cow health, reproduction and pro ductivity." PLoS ONE. 8 (6): e65766. doi:10.1371/journal.pone.0065766.

48. ^ Linlin Chen, Huidan Deng, Hengmin Cui, Jing Fang, Zhicai Zuo, Junliang Deng, Yinglun Li, Xun Wang, Ling Zhao (2017). "Inflammatory responses and inflammation-associated diseases in organs." Oncotarget. 9 (6): 7204–7218. doi:10.18632/oncotarget.23208.

49. ^*Park J, Min JS, Kim B, Chae UB, Yun JW, Choi MS, Kong IK, Chang KT, Lee DS (2015). "Mitochondrial ROS govern the LPS-induced pro-inflammatory response in microglia cells by regulating MAPK and NF-κB pathways." Neuroscience Letters. 584: 191–196. doi:10.1016/j.neulet.2014.10.016.*

Declarations

Funding: Indian Council of Agricultural Research-National Agricultural Science Fund (F. No.NASF/SUTRA-02/2022-23/50) to RMY, DS, and SKO.

Potential competing interests: No potential competing interests to declare.