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, Thanjavur, India.

[2]Molecular Endocrinology, Functional Genomics and Systems Biology Laboratory, Animal Biochemistry Division, ICAR-National Dairy Research Institute, Karnal, India.

$ Equal contribution

***Co-Corresponding Authors**

ORCID:

Ragothaman M. Yennamalli - 0000-0002-3327-1582

Suneel K. Onteru - 0000-0002-2834-314X

Dr. Suneel K. Onteru

Molecular Endocrinology, Functional Genomics and Systems Biology Laboratory, Animal Biochemistry Division, ICAR-National Dairy Research Institute, Karnal, India 132001.

Ph no: +91-8950735613, Email: suneelvet@gmail.com

Dr. Ragothaman M. Yennamalli

Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA Deemed to be University, Thanjavur, Tamil Nadu 613401, India.

Ph no: +91-8939321002, Email: ragothaman@scbt.sastra.edu

Abstract

Indian cattle breeds (*Bos indicus*) are renowned for their resilience in diverse environments and resistance to infections compared to *Bos taurus*. Although recent studies have explored the genetic basis of disease resistance in cattle, comprehensive analyses aimed at identifying the genomic signatures associated with disease-resistance traits in Indian cattle are still unclear. A genome-wide comparison between *Bos indicus* and *Bos taurus* could uncover key genomic elements related to disease-resistance traits or immunity. This study aimed to conduct chromosome-by-chromosome analyses among Bos genomes using three pairwise combinations: *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 focusing in immunity genes and quantitative trait loci (QTLs). Comparative genomic analysis was performed using SyMAP, GSAlign, and SyRI. An in-house Python script was utilized for gene annotation and visualization. Statistical differences in genomic variations among breed pairs were assessed using ANOVA. Gene Ontology (GO) and pathway enrichment analysis were carried out using DAVID for each variant type. The analysis revealed non-syntenic regions associated with four immunity genes in Nelore: VSTM1L, Galectin-9, HDAC5, and POLH. Significant genomic differences were observed on chromosomes 12, 24, 28, 29 and X. Chromosome X of the Nelore lost 50 Mbps of genomic regions compared to Hereford, although all genes remained intact. Further exploration of translocation events on Chromosome X in the Nelore revealed that chromosomes 28 had more single nucleotide variants (SNVs), insertions, and deletions. In total, 8, 16, and 209 innate immunity genes in Nelore were observed with longer insertions, deletions, and substitutions respectively with 62 genes were had all three variations with statistical significance. QTL analysis indicated these variations were primarily linked to milk and health traits. GO and KEGG enrichment analyses revealed that genes with substitution variants were significantly enriched in immune-related processes, including NF-kappa B signaling, cytokine production, and T cell receptor signaling. Several key immune genes, such as TLR4, TLR9, RIPK2, PLCG2, FADD, CLEC7A, RELA, CD40, TRIM25, IRAK1/2/3, NOD2, BCL2, and CARD9, were involved in multiple biological process, highlighting the impact of structural variations on immune function in Nelore cattle. Overall, this study highlights genomic variations in innate immunity genes in Nelore, its associated QTLs, biological process and pathways. This finding lays the groundwork for future research to identify the immune markers related to disease-resistance traits in indicine cattle.

Keywords: Hereford, Nelore, Gir, innate immunity genes, non-syntenic region, Common chromosomal variations, QTLs

Introduction

Indicine cattle (members of *Bos indicus*), such as zebu cattle, are native to the Indian subcontinent. Accounting for about 80% of India's total cattle population, they offer various economic advantages over taurine cattle (members of *B. taurus*), particularly when raised under tropical conditions [1, 2]. The native breeds are crucial to the country's dairy industry, which has made India a global leader in milk production, and supports 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 are registered in India. 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 the Kottayam district of Kerala. It is 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. This response is non-specific to the pathogen [8, 9, 10]. Therefore, 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.* This disease resistance trait is essential for better management of herds and breeding development strategies [11, 12]. In particular, exploring the genomic regions responsible for disease resistance and susceptibility variations opens new opportunities to identify significant genomic signatures for future molecular breeding studies [13].

Recent research on livestock improvement is increasingly focused on whole-genome data to detect genetic factors or genomic signatures associated with important traits such as disease resistance. This includes identifying such structural variations such as insertions, deletions, duplications, inversions, translocations, and other structural variations present on genomic regions by comparing indicine and taurine breeds. Two relevant sources of assembled whole-genome data at chromosome level of 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 identified a greater number of genetic variations. This study revealed over 155 million single-nucleotide polymorphisms

(SNPs) and more than one million InDels among a total of 17,252 genes [16]. 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 multiple studies exploring the genetic basis of disease resistance in cattle genomics [18, 19, 20, 21, 22], a genome-wide comparison between assembled chromosome-level genomes of *Bos indicus* and *Bos taurus* remains limited and requires additional investigation. To bridge that gap, we conducted a detailed chromosome-by-chromosome analysis of the Bos genomes available with the National Centre 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 members of three distinct pairs, namely *Bos taurus* (Hereford) and *Bos indicus* (Nelore), *Bos taurus* (Hereford) and *Bos indicus* (Gir), and *Bos indicus* (Nelore) and *Bos indicus* (Gir), to obtain fresh insights into the genetic basis of the disease resistance in *Bos indicus* cattle.

Materials and methods

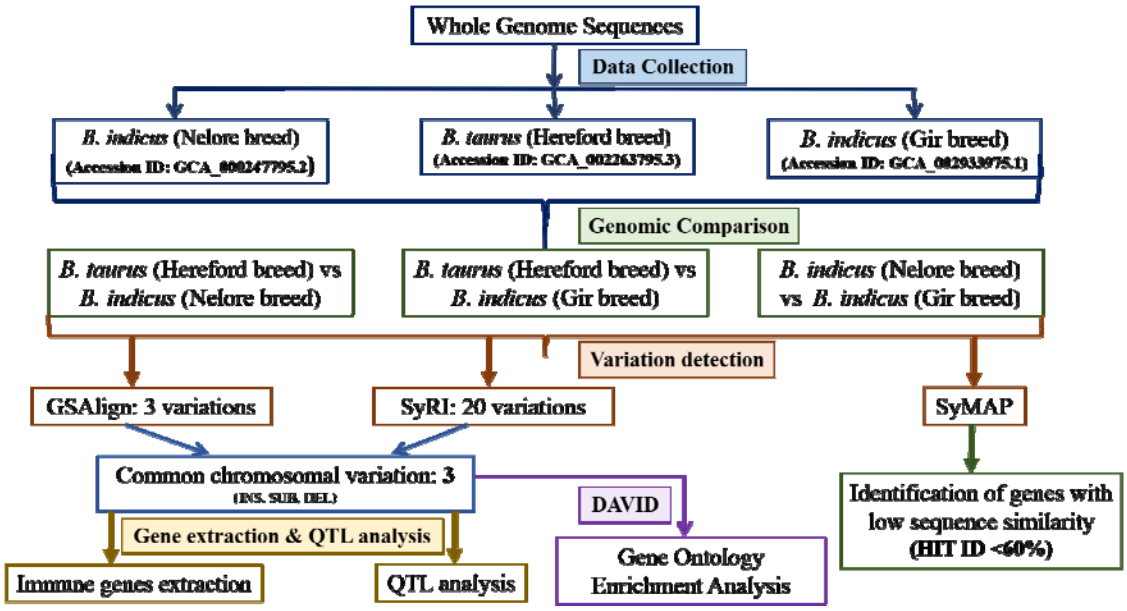

Fig. 1 Workflow used in this study.

General workflow

The flowchart (Fig. 1) outlines the steps in this 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 Synteny Mapping and Analysis Program (SyMAP) [23], GSAlign [24], and Synteny and Rearrangement Identifier (SyRI) [25]. 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 Nelore protein-coding data were used to identify genes and immunity genes associated with these variations. Quantitative trait loci (QTLs) were mapped to the genes, and immunity genes were identified within the common chromosomal variations. Gene Ontology (GO) and KEGG pathway enrichment analysis was performed for the genes carrying common chromosomal variations using DAVID [26, 27].

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 analyzed for the number of chromosomes assembled, scaffold N50 values, GC percentage, and genome coverage using QUAST [28]. The N50 values indicating length of the contig where all contigs of at least together to cover at least 50 % of the assembly in comparison to the reference genome, along with length greater than 100Mb was considered for the analysis. 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 conducted pairwise comparison using one from *Bos taurus*, namely Hereford, and two from *Bos indicus*, namely Nelore and Gir. The three pairwise comparison were Hereford–Nelore, Hereford–Gir, and Nelore–Gir. In each pairwise analysis, the genome of the former breed was designated as the reference, and the latter one as the query. The comparisons focused on variation in terms of the length, number, and function of genetic variants. In addition, interchromosomal variations were identified 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 comparative genomics tools, namely SyMAP, GSAlign, and SyRI. These tools were employed primarily to analyze the query genome against the reference genome of bovines' chromosome by chromosome.

Analysis using SyMAP

SyMAP was used for detecting and analyzing syntenic and non-syntenic regions between the query and the reference chromosomes of the selected genomes [23]. 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 the members of each of the three pairs of breeds as mentioned above and also to compare the 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 [24]. 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) and 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 [25]. As before, we analyzed the three pairs of chromosomes 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 [29]. 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 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 ensure high confidence variant detection, regions with poor assembly quality such as those containing excessive gaps or low alignment scores were excluded from the analysis. Furthermore, only chromosomal variations consistently identified by both GSAlign and SyRI were retained for downstream annotation, as SyMAP

does not provide variant-level information. This approach minimized technical artifacts and enhanced the biological and functional relevance of the results. Identical variants 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. Variant functional annotation was performed particularly in comparing *Bos taurus* and *Bos indicus* (Nelore), because annotations were unavailable for Gir.

Statistical Analysis

To evaluate the statistical significance of genomic variation across the three pairwise breed comparisons (Hereford vs. Nelore, Hereford vs. Gir, and Nelore vs. Gir), one-way analysis of variance (ANOVA) was performed. The test was applied separately for each type of variants: insertions, deletions, and substitutions identified using GSAlign and SyRI. ANOVA was used to determine whether the normalized number of variants per chromosome significantly different. For example, for normalized insertion counts in chromosome 1 in one pairwise comparison was checked for the same chromosome in other pairwise comparisons. A higher F-value indicates greater variance between groups relative to within-group variance. Statistical significance was assessed using corresponding P-values, with values less than 0.05 considered significant.

Genome annotation from common chromosomal variations

To annotate common chromosomal variations, we first identified structural variants detected by both GSAlign and SyRI. These overlapping variants were then analyzed using in-house Python scripts to locate genes within the *Bos indicus* reference genome (NCBI RefSeq assembly: GCF_000247795.1, Nelore) that harbored shared insertions, deletions, and substitutions. Although chromosome-level assembly data are available for the Gir breed, gene annotations have not yet been deposited in NCBI at the time of analysis. As a result, for the purpose of gene-level annotation of chromosomal variations, we focused on the pairwise comparison between *Bos taurus* (Hereford) and *Bos indicus* (Nelore), for which both complete assemblies and gene annotations are available. We applied eight specific criteria (outlined in Fig. 2) to filter and extract protein-coding genes associated with these variations in the Nelore genome. This approach resulted an annotated output containing detailed information on the genes associated with the variations.

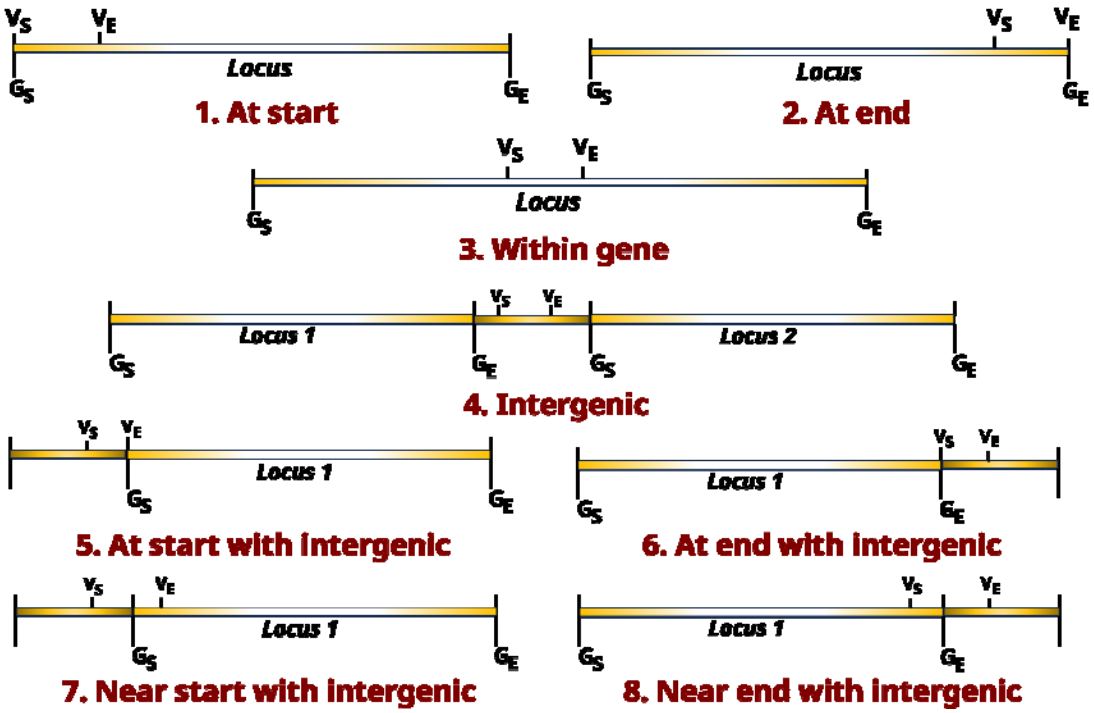

Fig. 2 Rules applied for extracting genes for common chromosomal variation from indicine protein-coding data. These rules were employed to detect genes from indicine protein-coding data that are linked to common chromosomal variations as predicted by two tools, GSAlign and SyRI. G_S, gene start; G_E, gene end; 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). Only protein-coding genes containing the variations within their genomic coordinates (rule no.3) were considered for further downstream anlysis. The analysis for variations located outside of the genes such as upstream, downstream, or in intergenic regions were not included in this study.

Extracting immunity genes

To identify the immune-related genes with structural variations, we employed a combined approach using both curated database and keyword-based search, implemented through an in-house Python script. First, we used the InnateDB database to identify innate immune genes with structural variations. This database contains a curated list of 1697 genes specifically involved in innate immunity. However, since InnateDB focuses solely on innate immunity, we aimed to capture other immune-related genes that may contribute to adaptive immunity or immune regulatory functions but are not included in the database. To achieve this, we conducted a keyword-based search on the annotated gene list derived from the

genome data. The following immune-related 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.* This combined approach allowed for a more comprehensive identification of genes involved in immune-related functions.

Analysis of quantitative trait loci

Using the genomic variations such as insertions, deletions, and substitutions, identified by both GSAlign and SyRI 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 QTL 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 reported QTLs associated with various traits in the chromosomal regions showing structural variants.

Gene Ontology Enrichment Analysis:

To investigate the functional relevance of genes with structural variants, we performed Gene Ontology (GO) enrichment analysis separately for genes impacted by insertions, deletions, and substitutions. The analysis was conducted using the DAVID Bioinformatics Resources (v6.8). For each variant category, a list of variant genes was submitted, and GO terms were evaluated under the categories of Biological Process (BP), Molecular Function (MF), Cellular Component (CC), KEGG Pathways, and InterPro protein domains. Only significantly enriched terms with a false discovery rate (FDR) < 0.05 were retained. We compared the significant GO terms across all three variant types to identify common functional signatures and selected those consistently enriched ($p < 0.05$) in each dataset. These significantly enriched terms were visualized using Python's matplotlib/seaborn.

Results

Reference Genome Statistics

We used three reference genomes for comparative variant discovery: *Bos taurus* (ARS-UCD1.2, GCF_002263795.3, Hereford breed), *Bos indicus* (Bos_Indicus, Nelore breed), and *Bos indicus* (ASM293397v1, Gir breed). The *Bos taurus* reference consisted of 1,958 contigs, with a total length of 2.77 Gb, an N50 of 103.3 Mb, and minimal gap content (1.02 N's per 100 kbp). The Bos_Indicus assembly contained only 32 contigs and had the highest N50 (106.3 Mb), but a higher proportion of gaps (7,664.85 N's per 100 kbp). ASM293397v1 was more fragmented (23,400 contigs), with moderate gap content (1,657 N's per 100 kbp) and an N50 of 105.7 Mb. All three assemblies have high contiguity based on their N50 values. Despite the limitation in *Bos indicus* assembly, we incorporate them to particularly identify indicine-specific variants. To mitigate the impact of low-quality regions, we excluded poorly aligned or highly gapped regions and retained only the variants identified by both GSAlign and SyRI tool for downstream annotation and analysis. This ensured the reliability and biological relevance of the identified variants.

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

Table 1: List of the chromosomes that have less than 60 % sequence identity between *B. indicus* (Nelore breed) and *B. taurus* predicted by SyMAP [23]. The sequence diversity in the non-syntenic regions is represented by differing functional annotation or unknown annotations.

Chromosomes	**Genes in *B. indicus* (Nelore breed)**	**Genes in *B. taurus* (Hereford breed)**	**Sequence identity (%)**	**Sequence similarity (%)**
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	*Annotation not available*	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	*Annotation not available*	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

SyMAP analysis revealed both syntenic and non-syntenic regions, along with alignment statistics, between the genomes of Nelore and those of Hereford. For downstream analysis, we focused on genes located in non-syntenic regions that exhibited less than 60% sequence identity. These gene candidates potentially are affected by significant variations and indicate divergence between breeds. This analysis identified 13 genes across four chromosomes in Nelore one gene each on chromosomes 7 and 23, six on Chromosome 18, and five on Chromosome 19 (Table 1). Among these, four genes in Nelore: VSTM1L, Galectin-9, HDAC5, and POLH were associated with immune-related functions. The remaining genes, though not directly linked to immunity, were retained for comparative and enrichment analyses to determine if broader functional categories were also impacted.

Interbreed genomic variations identified using GSAlign

The distribution of different variations including insertions, deletions, and substitutions along the chromosomes is shown in Fig. 3. 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. To normalize the variant count across chromosome, we calculated variant density, defined as the number of variants per megabase (Mb) genomic regions for each chromosome. Therefore, Chromosome 29 displayed the highest number of insertions (170.20 variants/Mb), Chromosome 24 shown highest number of substitutions (2924.84 variants/Mb), and Chromosome X, the highest number of deletions (108.63 variants/Mb) (Fig. 2: A, B, and C). The corresponding figures for Hereford and Gir were 673,875, 672,892, and 11,500,574 in Gir, with Chromosome X showing the highest number of insertions (306.01 variants/Mb) and the highest number of deletions (319.79 variants/Mb), and Chromosome 24 showing the highest number of substitutions (5085.31 variants/Mb) (Fig. 2: D, E, and F). 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 12 showing the highest number of insertions (281.94 variants/Mb), Chromosome 28 showing the highest number of substitutions (5574.32 variants/Mb), and Chromosome 29 showing the highest number of deletions (324.72 variants/Mb) (Fig. 2: G, H, and I).

355

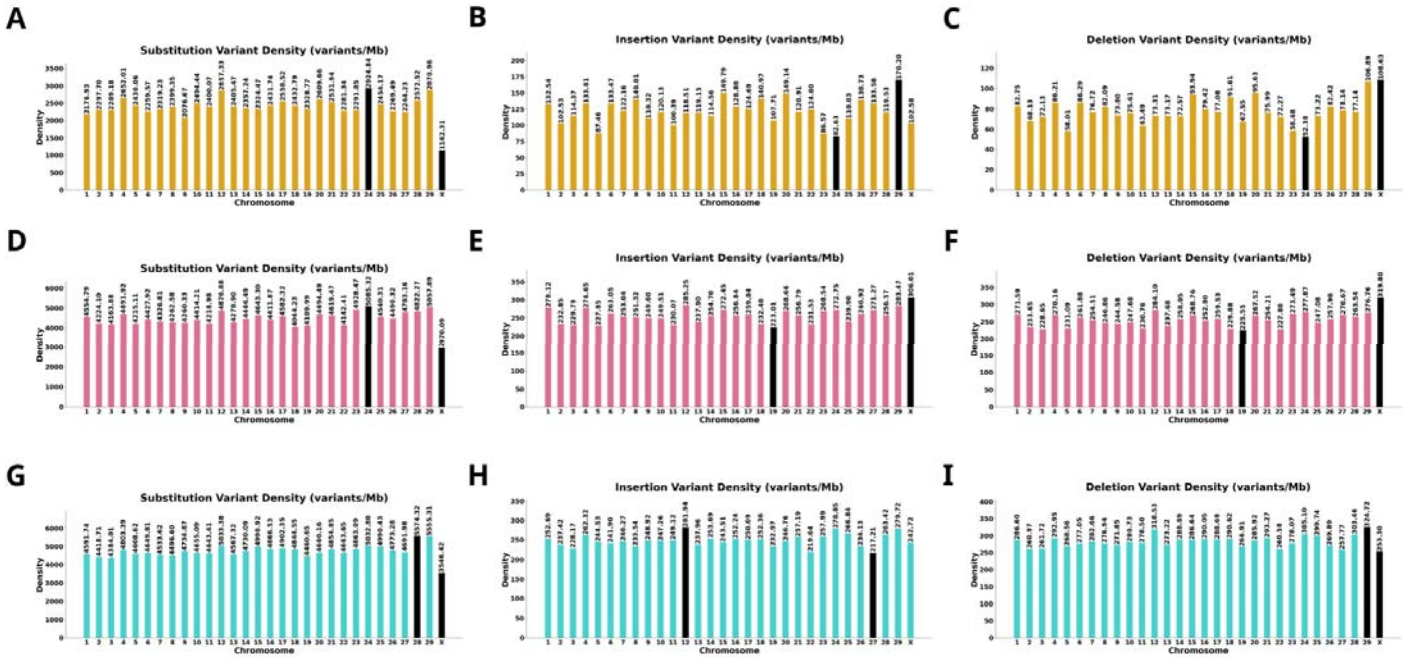

Fig. 3 Interbreed variant density (variants/Mb of genomic region) between members of three pairs of cattle breeds: *B. taurus* (Hereford) and *B. indicus* (Nelore), *B. taurus* (Hereford) and *B. indicus* (Gir), and *B. indicus* (Nelore) and *B. indicus* (Gir), as identified by GSAlign. Panels (A–I) represent the numbers of single nucleotide variations (SNVs), insertions, and deletions. Comparison between Hereford and Nelore (gold colored barplots): A, SNVs; B, insertions; C, deletions. Comparison between Hereford and Gir (pink colored barplots): D, SNVs, E, insertions; F, deletions. Comparison between Nelore and Gir (cyan colored barplots): G, SNVs; H, insertions; I, deletions. Black bars represent the minimum and the maximum variant counts.

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).

Table 2: Bos genomes' comparison statistics as per SyRI

SyRI Variants	Hereford breed vs Nelore breed	Hereford breed vs Gir breed	Nelore breed vs Gir breed
Copy gain in non-reference genome (CPG)	1458	2614	2750
Copy loss in non-reference genome (CPL)	428	254	2091
Duplication (DUP)	3300	2592	6436
Duplication alignment (DUPAL)	4621	3304	8146
Highly diversed regions (HDR)	37088	30152	47015
Inversion (INV)	1145	2890	3001
Inversion alignment (INVAL)	3375	3887	6161
Inverted duplications (INVDP)	2129	1310	5147
Inverted translocation (INVTR)	669	717	1346
Not aligned regions (NOTAL)	17914	25116	38247
Syntenic regions (SYN)	7022	7106	9983
Syntenic alignment (SYNAL)	53321	37580	55916
Tandem repeat (TDM)	55	43	20
Translocation (TRANS)	1109	1596	2296
Translocation alignment (TRANSAL)	2143	2794	4281
Inverted duplication alignment (INVDPAL)	2650	1666	6121
Inverted translocation alignment (INVTRAL)	1117	1276	2459
Insertions (INS)	329698	674040	556520
Deletions (DEL)	186761	674737	666349
Substitutions Nucleotide Variations (SNV)	15543630	14559218	16282402

The comparison between *Bos indicus* (Nelore) and *Bos taurus* (Hereford) (Fig.4A) identified a total of 16,199,633 variations across all 30 chromosomes, comprising 20 variants. The majority of these variations (15,543,630) were SNVs, distributed across all 30 chromosomes. Chromosome 29 had the highest number of variations across 6 variants: DUP (1.99 variants/Mb), NOTAL (9.02 variants/Mb), SYN (3.43 variants/Mb), TRANS (0.76 variants/Mb), INS (171.47 variants/Mb), and DEL (96.65 variants/Mb). Chromosome 27 shown higher number of variants in CPL (0.28 variants/Mb), INVAL (4.16 variants/Mb), TDM (0,08 variants/Mb), SNV (7048.96 variants/Mb), DUPAL (2.81 variants/Mb). Chromosome X recorded higher variations across two variant types: INVTR (0.39 variants/Mb) and INVTRAL (0.76 variants/Mb). Chromosome 21 displayed a higher number in SYNAL variants (160.37 variants/Mb) and TRANSAL (1.48 variants/Mb). Chromosome 26 shown an increase in HDR (17.56 variants/Mb) and INVDPAL (1.31 variants/Mb). Chromosome 6 showed an

increase in INV (0.63 variants/Mb), Chromosome 12 showed an increase in CPG (0.80 variants/Mb), and Chromosome 17 showed an increase in INVDP (0.99 variants/Mb). 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. 4C. 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*.

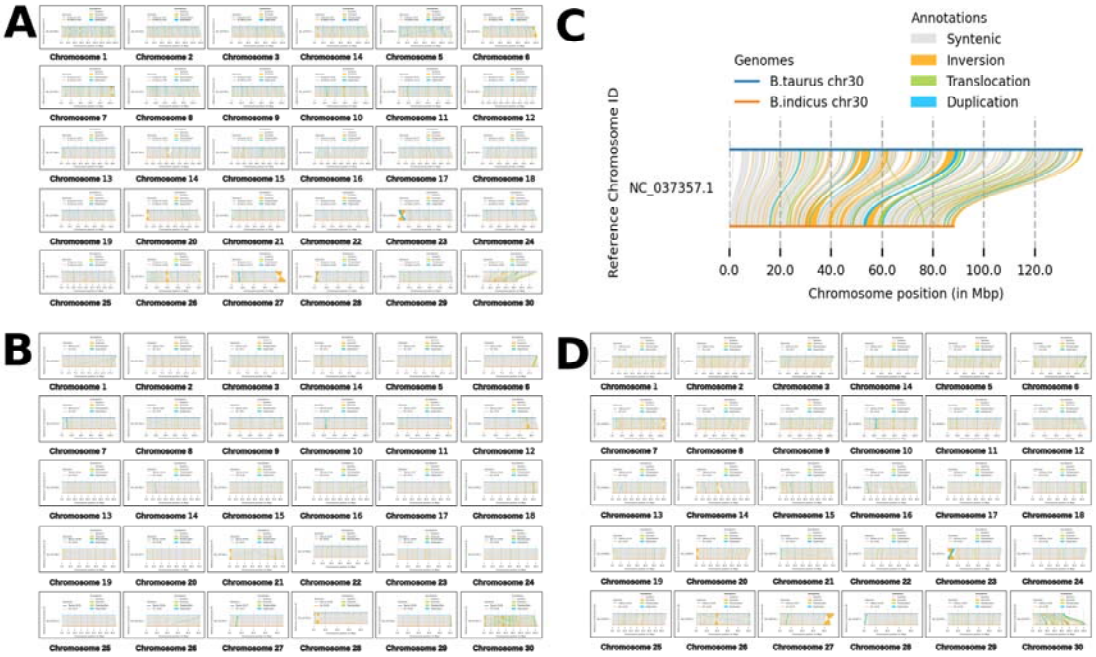

Fig. 4 Synteny map of all 30 chromosomes, generated using SyRI, comparing between members of three pairs of breeds. Grey, conserved syntenic regions; yellow, inversions; green, translocations; and cyan, duplications. A–D, plots generated using plotsr [22]: A, *B. taurus* and *B. indicus* (Nelore); B,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

In the comparison between *B. indicus* (Gir) and *B. taurus* (Hereford) is shown in Fig.4B, 16,032,892 variations were identified covering 20 variants across all 30 chromosomes. Similar to GSAlign, SyRI results were normalized for variant counts with respect to the chromosome length of the reference genome. Among these, 14,559,218 were SNVPs. Chromosome X had a large number of variants across the following 16 variants: CPG (2.59 variants/Mb), DUP (3.58 variants/Mb), DUPAL (5.02 variants/Mb), HDR (18.86 variants/Mb), INVAL (3.79 variants/Mb), INVDP (2.08 variants/Mb), INVTR (0.70 variants/Mb), NOTAL (17.64 variants/Mb), SYN (3.97 variants/Mb), SYNAL (20.42 variants/Mb), TRANS (1.40 variats/Mb), TRANSAL (3.44 variants/Mb), INVDPAL (2.97 variants/Mb), INVTRAL (1.61

variants/Mb), INS (292.39 variants/Mb), and DEL (300.20 variants/Mb). Unlike in Nelore, Chromosome X in Gir gained approximately 10 Mbp of the genomic region compared to that in taurine. Chromosome 17 showed an increase in the CPL variant (0.19 variants/Mb), Chromosome 18 increases in INV (1.62 variants/Mb), Chromosome 27 increases in TDM (0.06 variants/Mb), and Chromosome 29 had more SNV (6561.19 variants/Mb) variants.

Comparing the two *B. indicus* breeds such as Nelore and Gir breeds is shown in Fig. 4D revealed 17,706,687 variations comprising 20 variants across all 30 chromosomes in Gir. Of these, 16,282,402 were SNVs. Chromosome X showed higher variations across twelve variants: CPG (6.55 variants/Mb), DUP (12.83 variants/Mb), DUPAL (20.21 variants/Mb), HDR (21.89 variants/Mb), INVAL (6.71 variants/Mb), INVDP (10.89 variants/Mb), INVTR (1.64 variants/Mb), NOTAL (35.33 variants/Mb), TRANS (1.78 variants/Mb), TRANSAL (4.86 variants/Mb), INVDPAL (15.64 variants/Mb), and INVTRAL (5.72 variants/Mb). In addition, Chromosome X in Gir gained 60 Mbp of genomic material with intergenic regions. Chromosome 29 shown higher number of variations in SYNAL (24.45 variants/Mb) and SNVs (7224.98 variants/Mb). Chromosome 12 exhibited more DEL (283.04 variants/Mb), chromosome 15 had more SYN (5.14 variants/Mb), chromosome 20 had more INV (1.43 variants/Mb), chromosome 21 had more TDM (0.02 variants/Mb), and chromosome 28 had more INS (224.61 variants/Mb) variants.

Analysis of interchromosomal variations between X or Y chromosomes 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, (Fig. 5).

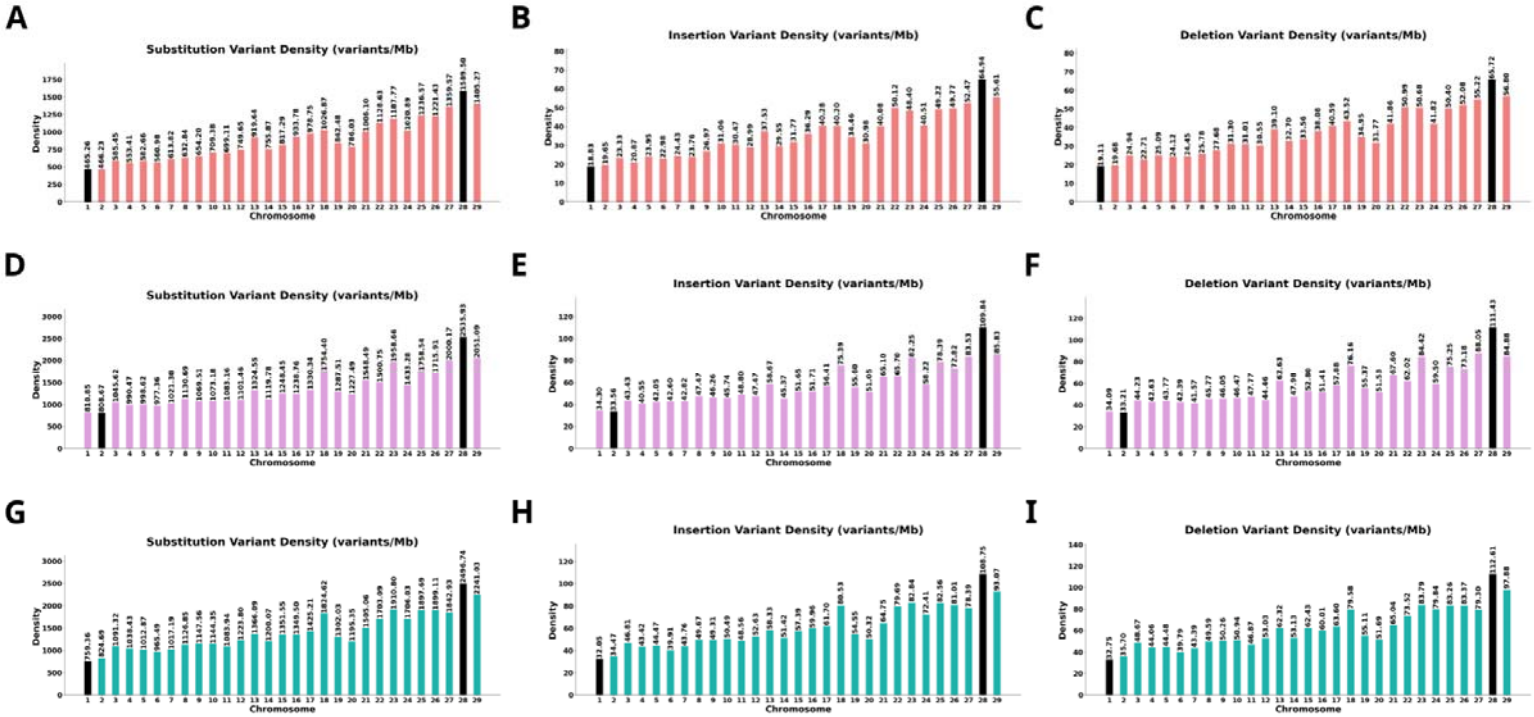

Fig. 5 Sex chromosomes X or Y compared with autosomes between members of three pairs of breeds using GSAlign. A-I represent the variant density maps, defines variant count per megabase (Mbp) of genomic regions. **A–C**, Chromosome X compared with autosomes in Nelore. **A,** single nucleotide variations; **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. Black bars represent the minimum and the maximum variant counts.

In Nelore (*B. indicus*), 80,495 insertions, 83,601 deletions, and 1,994,708 substitutions were identified between autosomes with Chromosome X. Chromosome 28 showed the highest number of SNVs (1589.50 variants/Mb), insertions (64.93 variants/Mb), and deletions (65.71 variants/Mb). 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 28 recorded the largest number of variations, with SNVs (2535.92 variants/Mb), insertions (109.84 variants/Mb), and deletions (111.43 variants/Mb). In Gir (*B. indicus*), the analysis identified 137,587 insertions, 140,017 deletions, and 3,193,030 substitutions. Chromosome 28 showed the highest number of SNVs (2496.73 variants/Mb), insertions (108.75 variants/Mb), and deletions (112.61 variants/Mb).

Further analysis of Y chromosome variations in Nelore (Fig. 6) showed Chromosome X to have the highest number of SNVs (104.38 variants/Mb) and deletions (5.85 variants/Mb) whereas Chromosome 17 showed the most insertions (5.54 variants/Mb).

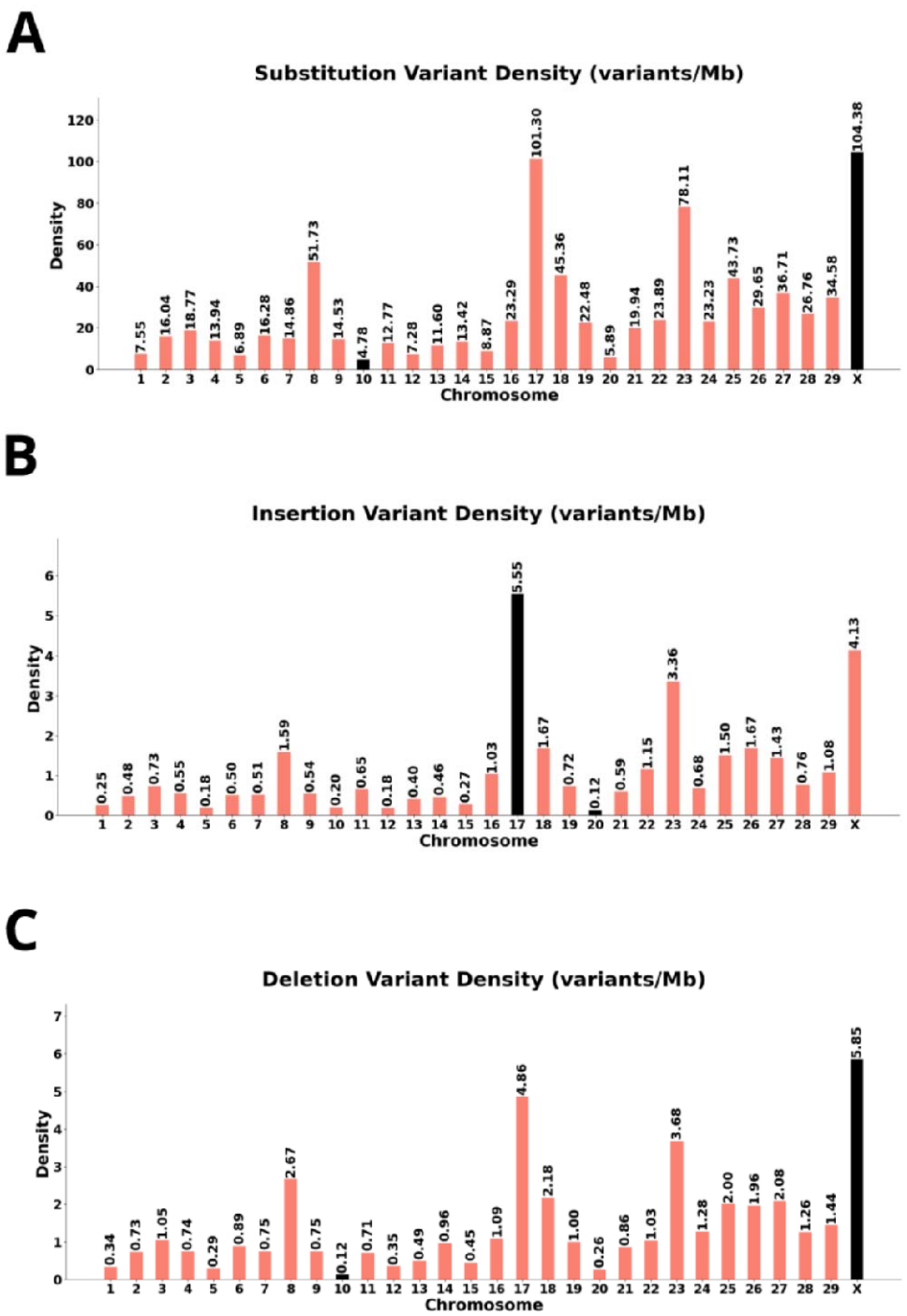

Fig. 6 Chromosome Y and autosomes in *Bos indicus* (Nelore) compared using GSAlign. A-C represent the variant density maps, defines variant count per megabase of genomic regions. **A**, single nucleotide variations; **B**, insertions; **C**, deletions. Black bars represent the minimum and the maximum variant counts.

Identification of high-confident variants

Statistical significance of variations

To assess the statistical significance of genomic variation across the three breed comparisons (Hereford vs. Nelore, Hereford vs. Gir, and Nelore vs. Gir), we performed one-way ANOVA on the normalized number of insertions, deletions, and substitutions identified by GSAlign and SyRI. The GSAlign analysis revealed highly significant differences among the breed pairs for all three variant types: substitutions (F-value = 387.18, p-value = $4.88 \times 10^{\square\square\square}$), insertions (F-value = 494.03, p-value = $3.17 \times 10^{\square\square\square}$), and deletions (F-value = 1239.32, p-value = $1.17 \times 10^{\square\square\square}$). Similarly, SyRI analysis also indicated statistically significant differences for insertions (F = 458.38, p-value = $6.28 \times 10^{\square\square\square}$), deletions (F = 1313.4, p-value = $1.01 \times 10^{\square\square\square}$), and to a lesser extent for substitutions (F = 8.02, p-value = $6.32 \times 10^{\square\square}$).

In ANOVA, a higher F-value reflects a greater difference in means between group relative to the variations within groups, suggesting stronger evidence of true differences. A smaller P-value (typically < 0.05) indicates strong statistical significance. All observed P-values in this analysis were extremely low, indicating highly significant differences in the number of variants among the breed comparisons. These findings confirm that structural variants, particularly insertions and deletions differ significantly among Nelore, Gir, and Hereford cattle, indicating considerable genomic divergence among the breeds.

Variations involving insertions

We identified a significant number of insertions across various chromosomes. As chromosome 29 showed higher insertions variant density, 1388 insertions occurring in intergenic regions, 567 insertions in 149 genes, and 6 insertions in two immunity-related genes (Table 3 and Fig. 7).

Chromosome X showed 2782 insertions in intergenic regions and 539 insertions in 96 genes, including 18 insertions in six 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 in 296 genes, and Chromosome 18 had 667 insertions in 289 genes. Conversely, chromosomes with a greater number of insertions often affected fewer genes, such as Chromosome 26, in which 47 insertions were confined to only three immunity genes.

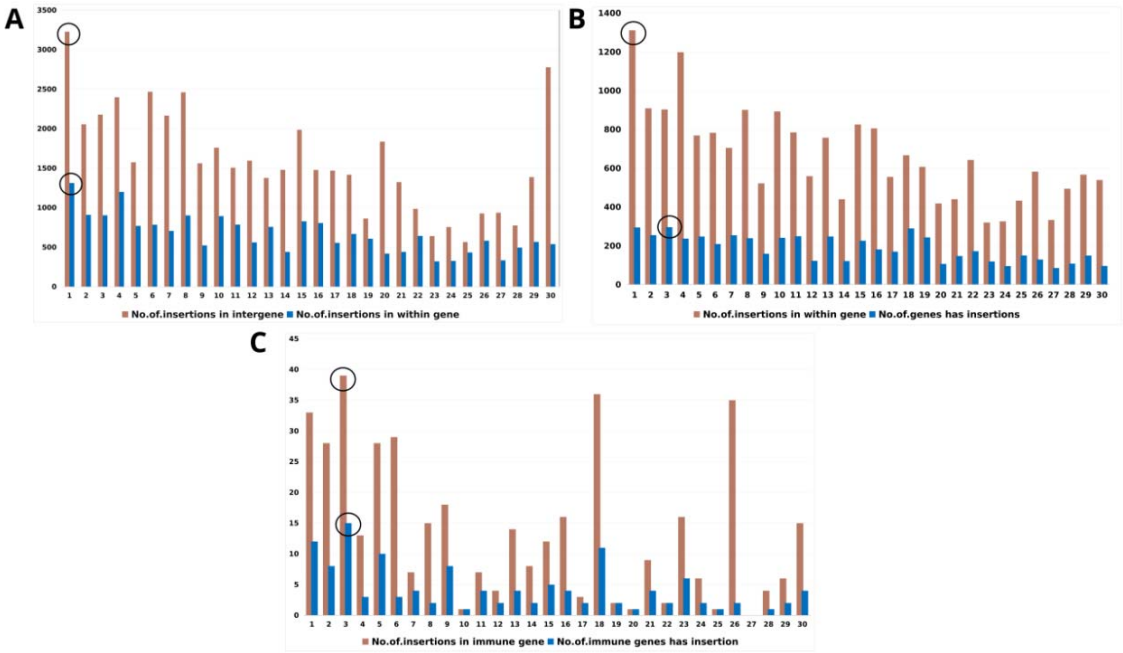

Fig. 7 Insertions in all 30 chromosomes in two genomes (*Bos indicus* and *Bos taurus*). Circles indicate the highest number of insertions. **A**, number of insertions between the intergenic region and within gene region; **B**, number of insertions within gene region and genes affected by insertions;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

Since variations greater than 50 base pairs (bp) can significantly affect gene expression and function by altering gene dosage, disrupting coding regions, or interfering with long-range regulation, we focused on genes impacted by variants exceeding 50bps long [30]. As a result, a total of 54 genes were identified with insertions longer than 50 bp (Supplementary Table 1 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.

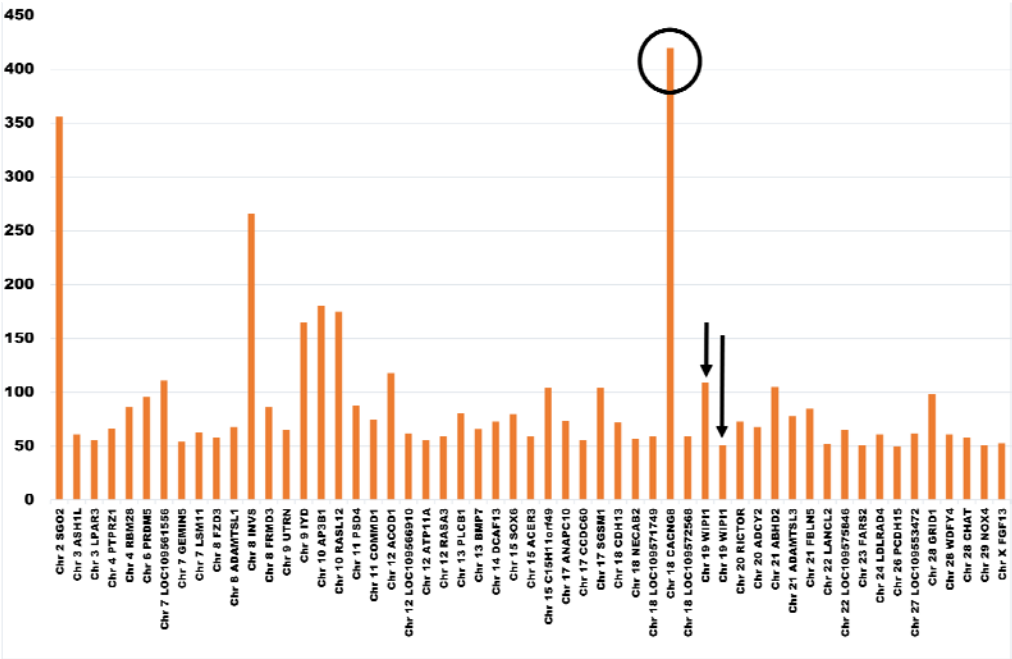

Fig. 8 Characters of 30 genes with variations longer than 50 bp. The circle marks the highest number of insertions and the arrows show genes with multiple insertions and deletions.

Across all chromosomes, a total of 343 immunity genes were affected by insertions. Among them, 232 were classified as innate immunity genes, identified using the InnateDB database, and the remaining 111 were classified as other immune-related genes, determined using a keyword-based search. Of the total, 182 immunity genes experienced multiple insertions, whereas the remaining 161 experienced only one insertion. Moreover, 78 immunity genes were affected by insertions longer than 10 bp; 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 X had the highest number of deletions per Mb of genomic regions, 1532 deletions in intergenic regions and 381 deletions in 52 genes, although no immune-related genes had deletions (Table 4 and Fig. 9). However, Chromosome 4 showed a greater number of deletions, a total of 440 genes. Chromosome 18 had the highest number of deletions in immune genes, with 31 deletions in 16 immunity-related genes.

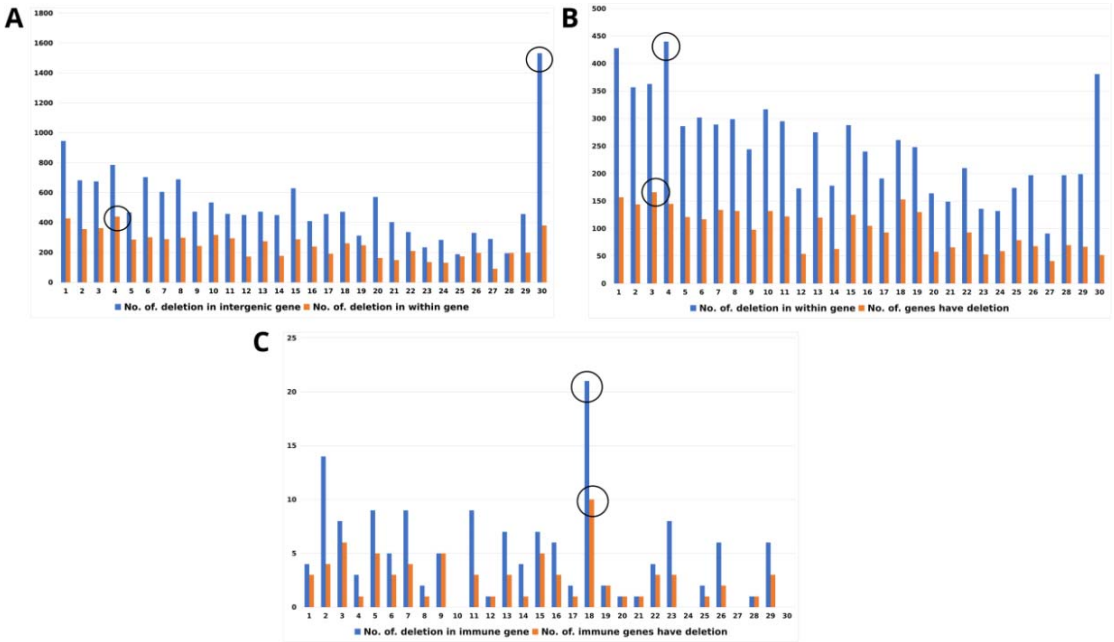

Fig. 9 Deletions in all 30 chromosomes in two breeds of cattle (Hereford and Nelore). Circles indicate the highest number of deletions. **A**, number of deletions between the intergenic region and within gene region; **B**, number of deletions within gene region and genes affected by the deletions; **C**,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

In addition, we identified 32 genes with deletions longer than 50 bp (Supplementary Table 2 and Fig. 10). Gene GLTSCR1 on Chromosome 18 had the longest deletion, measuring 222 bp. Gene GMDS on Chromosome 23 had two deletions, with LOVs of 73 bp and 62 bp. Genes LOC109573590 and LOC109573589 on Chromosome 19 had deletions measuring 50 bp, and gene FAT3 on Chromosome 29 had deletions measuring 53 bp.

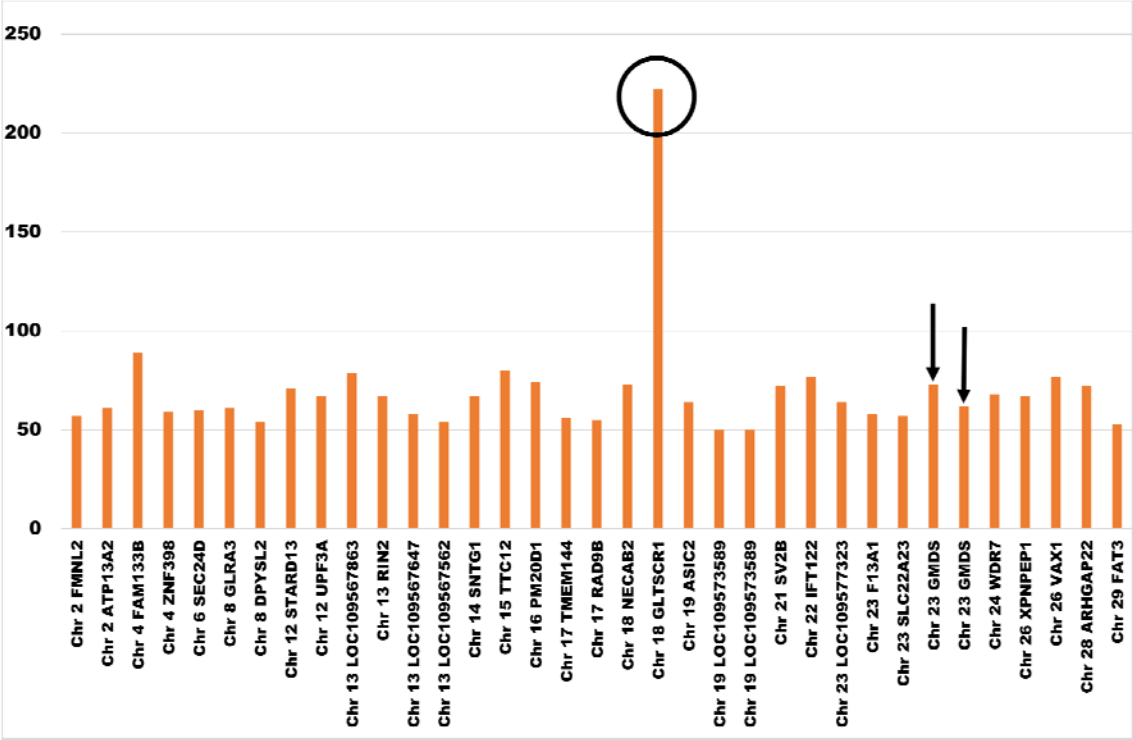

Fig. 10 Genes on 30 chromosomes experiencing variations longer than 50 bp in two cattle breeds (Hereford and Nelore). Circles indicate the greatest number of deletions and arrows indicate genes experiencing multiple insertions and deletions.

Across all 30 chromosomes, 183 immunity genes had 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 had 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 had two deletions, affecting 46 bp, and the innate immunity gene NFATC2 had two deletions spanning 39 bp. Several innate immunity genes had single deletions longer than 10 bp, including MERTK (13 bp), GAS6 (18 bp), and C2 (28 bp). Notably, no immunity gene on Chromosome X had deletions.

Furthermore, 11 innate immunity genes had 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 24 showed the highest number of substitutions density per Mb, a total of 75,157 substitutions are present in intergenic regions, 33,376 substitutions in 256 genes, and 364 substitutions in three immune-related genes (Table 3 and Fig. 11).

Table 3: 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.

Chromosome number	Total number of Common variations	Number o substitution in intergenic	Number of substitution in "within gene"	Number of Genes have substitution	Number of substitution in "at start"	Number of Genes have substitution in at start of the locus	Number of substitution in "at end"	Number of Genes have substitution in at end of the locus	Number of substitution in immune genes	Number 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

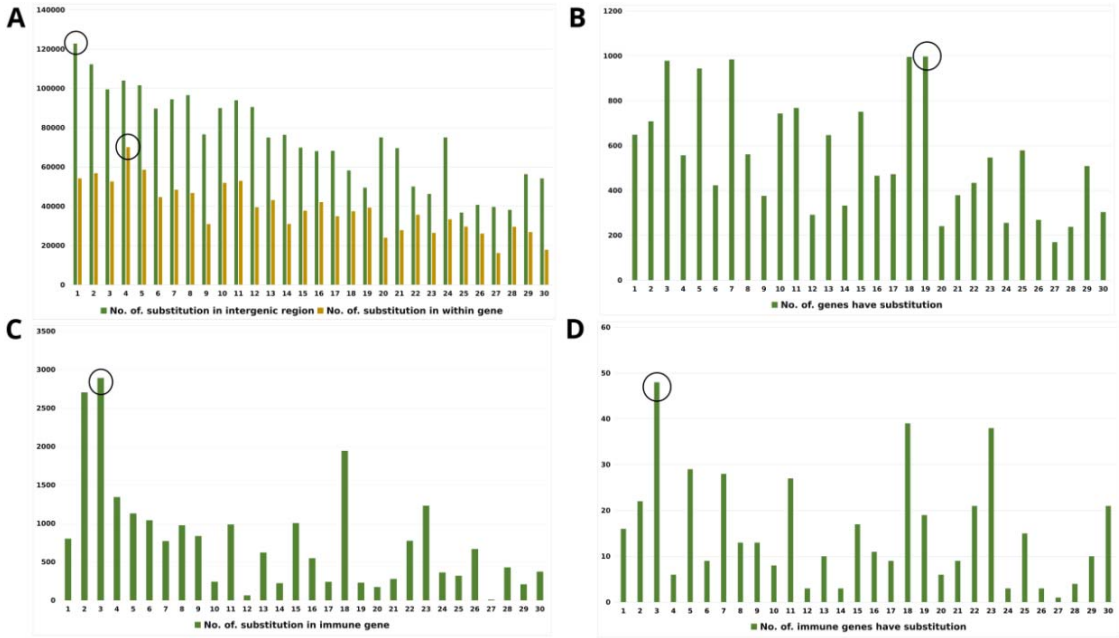

Fig. 11 Substitutions across all 30 chromosomes in three breeds of cattle. Circles indicate the highest number of substitutions. **A**, substitutions between the intergenic region and within gene region; **B**, genes affected by substitutions; **C**, number of substitutions in immunity genes; **D**, immunity genes affected by substitutions. The color legend indicates the location of the variation in 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 had three substitutions affecting the downstream region of three genes, and Chromosome 13 had 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. Chromosome 5 topping the list at 48,935 substitutions across 271 genes. Among the immunity genes, 301 had substitutions across all 30 chromosomes (Supplementary Table 3): 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 in 49 immunity genes. Gene SPAG16, which had been predicted, based on keywords, to be conferring immunity, had the highest number (1457) of substitutions, whereas the innate immunity gene PLXNA4 had 1338 substitutions. The innate immunity genes PRKCE (1184 substitutions), EDIL3 (1147 substitutions), and PRKCA (1089 substitutions) each have more than a thousand substitutions.

Notably, 93 immunity genes had all three types of variations, namely insertions (INS), deletions (DEL), and substitutions (SUB). Of these, 62 were classified as innate immunity genes based on the InnateDB database, while the remaining 31 genes were matched through immune keyword-based search. 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 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. In addition, 37 genes with variants exceeding 50 bp in length were mapped to various QTLs. Among the immune-related genes, MUC16 was mapped to milk and production QTLs, while RICTOR was associated with milk, health, and meat and carcass QTLs, as shown in Supplementary Table 3. 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. The QTLs mapped to the gene impacted by deletions variant exceeding 50 bps long are mentioned in the supplementary Table 4.

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 are mentioned in supplementary Table 5, 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.

GO term enrichment and Pathway analysis of genes affected by variants

For all the genes that had variations, we performed GO term enrichment analysis to identify significantly over represented biological processes (BP), molecular functions (MF), cellular components (CC), KEGG pathways, and InterPro protein domains.

For the genes with insertions, we observed enriched biological processes that included immune-related as well (Fig. 12A). For example, along with intracellular signal transduction (GO:00355556) and others, we observed immunity related GO enriched terms such as Wnt signaling pathway (GO:0016055) and endocytosis (GO:0006897), which are implicated in immunity as per previous reports [31, 32]. The KEGG pathway enrichment results (Fig. 12B) reveal calcium signaling pathway, pathways in cancer, Rap1/Ras signaling pathway and others. We observed pathways involve in immune response such as MAPK signaling, Wnt signaling, Human papillomavirus infection, inflammatory mediator regulation of TRP channels, endocytosis, Fc gamma R-mediated phagocytosis, T-cell receptor signaling, Yersinia infection, salmonella infections, and bacterial invasion of epithelial cells.

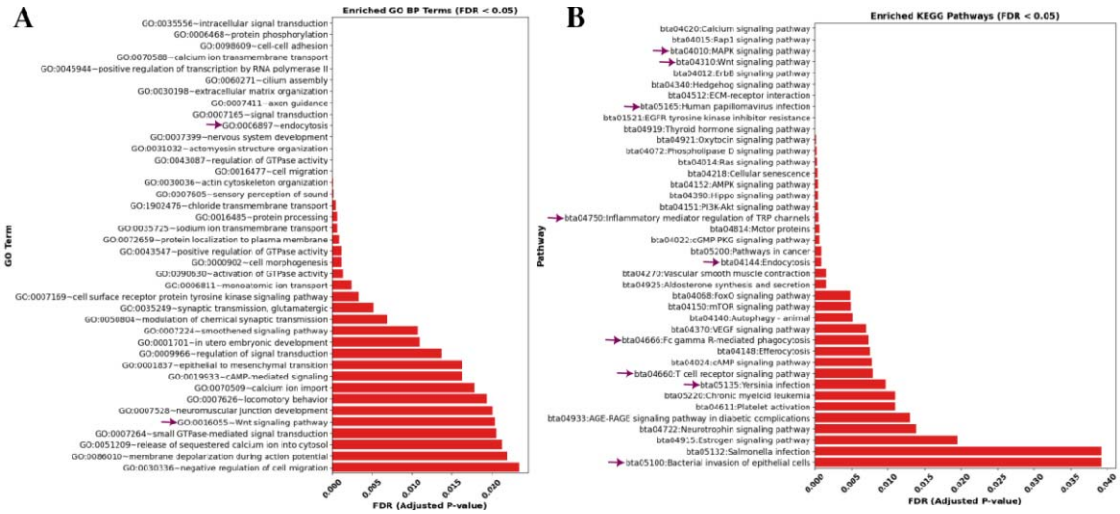

Fig. 12 GO term and pathway enrichment for insertion variations in Nelore breed. The horizontal bar length indicates the false discovery rate with adjusted p-values. The violet arrows indicate the biological process and KEGG pathways enriched in immune-related functions. **(A) GO term for biological process (BP) enrichment analysis** result for genes with insertions. **(B) KEGG pathway enrichment analysis** result for the genes with insertions. These results were filtered based on FDR < 0.05.

Additionally, the GO term enriched MF (Supplementary Figure 1A) includes functions such as protein binding, calmodulin binding, protein serine kinase activity, and various receptor activity such as macrophage colony-stimulating factor receptor (GO:0005011) involved in immune response and platelet-derived growth factor alpha receptor activity. Similarly, GO term enriched CC were significant for locations in the cytosol, cytoplasm, nucleoplasm, synapse, plasma membrane and early endosome (GO:0005769) involve in both innate and adaptive immunity (Supplementary Figure 1B). InterPro protein domains significantly enriched in kinase-like domains, PH-like domains, WD40 repeat-like domains, and immunoglobulin (Ig) domains such as Ig_I-set (Supplementary Figure 1C).

For the genes with deletion variations, we observed enriched biological processes (Fig. 13A) such as protein phosphorylation (GO:0006468), and others. The immune-related functions are also enriched in endocytosis and B-cell receptor signaling pathway (GO:0050853) (Fig. 13A). The KEGG pathways (Fig. 13B) enrichment analysis reveal pathways involve in immune-related functions such as MAPK signaling, human papillomavirus infection, Wnt signaling, T-cell receptor signaling, Virion-Lassa virus and SFTS virus, bacterial invasion of epithelial cells, and chemokine signaling.

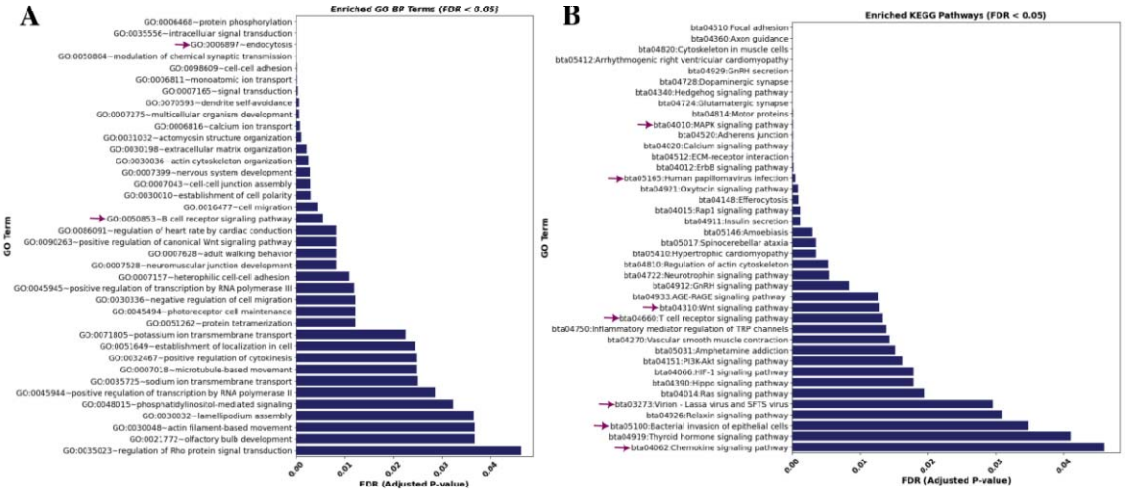

Fig. 13 GO Biological Process (BP) enrichment analysis result for the deletion variations in Nelore breed. The horizontal bar length indicates the false discovery rate with adjusted p-values. The violet arrows indicate the biological process and KEGG pathways enriched in immune-related functions. **(A) GO term for biological process (BP) enrichment analysis** result for genes with insertions. **(B) KEGG pathway enrichment analysis** result for the genes with insertions. These results were filtered based on FDR < 0.05.

Moreover, the GO term enriched MF (Supplementary Figure 2A) involves kinase activity, protein binding, calmodulin binding, and receptor activities such as macrophage colony-stimulating factor receptor similar to those identified in insertions. Similarly, GO term enriched CC were significant for locations involved in immune response such as endosome membrane (GO:0010008), early endosome, late endosome (GO:0005770), immunological synapse (GO:0001772) (Supplementary Figure 2B). Enriched InterPro protein domain includes kinase-like, EGF-like, PH-like, and immunoglobulin-related domains such as Ig_sub2 and Ig_E-set (Supplementary Figure 2C).

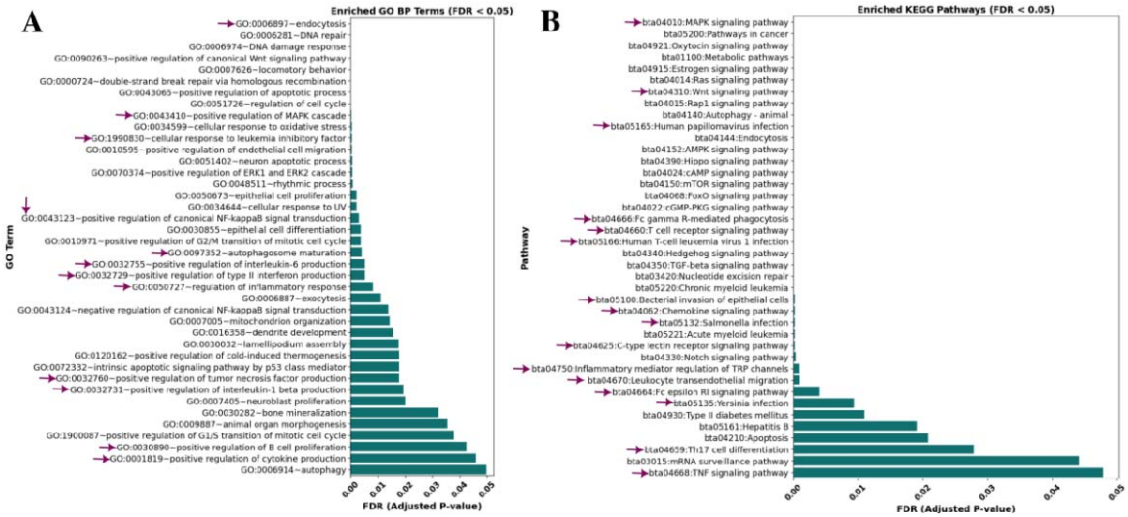

Fig. 14 GO: Biological Process (BP) enrichment analysis result for the substitution variations in Nelore breed. The horizontal bar length indicates the false discovery rate with adjusted p-values. The violet arrows indicate the biological process and KEGG pathways enriched in immune-related functions. **(A) GO term for biological process (BP) enrichment analysis** result for genes with insertions. **(B) KEGG pathway enrichment analysis** result for the genes with insertions. These results were filtered based on FDR < 0.05.

For the genes with substitution variations, we observed enriched biological processes (Fig. 14A) in immune-related functions such as endocytosis, cellular response to leukemia inhibitory factor (GO:1990830), autophagosome maturation (GO:0097352), positive regulation of MAPK cascade (GO:0043410), NF-kappaB signal transduction (GO:0043123), interleukin-6 productions (GO:0032755), type II interferon production (GO:0032729), tumor necrosis factor production (GO:0032760), interleukin-1 beta production (GO:0032731), B-cell proliferation (GO:0030890), cytokine production (GO:0001819), and regulation of inflammatory response (GO:0050727). The KEGG pathway (Fig. 14B) enrichment reveal the pathways involve in immune-related functions such as MAPK signaling, Wnt signaling, human papillomavirus infection, Fc gamma R-mediated phagocytosis, T-cell receptor signaling, human T-cell leukemia virus 1 infection, bacterial invasion of epithelial cells, chemokine signaling, salmonella infection, C-type lectin receptor signaling, Fc epsilon RI signaling, inflammatory mediator regulation of TRP channels, yersinia infection, Th17 cell differentiation, and leukocyte transendothelial migration.

In addition, The GO term enriched MF (Supplementary Figure 3A) involves macrophage colony-stimulating factor receptor activity, nuclear receptor activity, Hsp90 binding, p53 binding, and DNA helicase activity. Similarly, GO term enriched CC were significant for

locations involve in immune response such as autophagosomes (GO:005776), phagocytic vesicles (GO:0045335), immunological synapses, endocytic vesicles (GO:030139), early endosome membrane (GO:0031901) and lysosomal membranes (GO:0005765) (Supplementary Figure 3B). Enriched InterPro protein domain includes immunoglobulin-related domains such as Ig_I-set, Ig_E-set, and Ig_sub2 (Supplementary figure 3C).

Shared Enrichment Across Variant Types

In addition to variant-specific enrichment, we identified several biological processes and KEGG pathways (FDR < 0.05) that were commonly enriched across all three variant types such as insertions, deletions, and substitutions. These shared enriched terms for biological process and KEGG pathways are visualized as heatmaps (Supplementary Figure 4A and 4B). The common enriched biological processes include intracellular signal transduction, protein phosphorylation, cell migration, activation of GTPase activity, and endocytosis. The KEGG pathway analysis reveals the shared enrichment in several signaling and immune-related pathways, including calcium signaling, MAPK signaling, Wnt signaling, oxytocin signaling, inflammatory mediator regulation of TRP channels, T-cell receptor signaling, and bacterial invasion of epithelial cells.

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. While the *Bos taurus* ARS-UCD1.2 reference genome is of high contiguity and reference-grade quality (1,958 contigs, N50 = 103.3 Mb, and only 1.02 N's per 100 kbp), we also incorporated two *Bos indicus* genomes such as Bos_Indicus (32 contigs, N50 = 106.3 Mb, 7,664.85 N's/100 kbp) and ASM293397v1 (23,400 contigs, N50 = 105.7 Mb, 1,657 N's/100 kbp). Although these indicine assemblies exhibit limitations such as increased gap content or fragmentation, we included these genomes to identify indicine specific variant in the high confidence regions. We acknowledge that regions of lower assembly quality could influence variant detection in some loci. To address this, we applied good quality control measures such as excluded low-confidence and highly gapped regions, and focused on confidently aligned segments across all three genomes. In addition, we filter the high confidence variants which identified by two different tool such as GSAlign and SyRI. This approach helped ensure that the identified structural variants reflect true biological differences rather than assembly artifacts. The ANOVA result indicates the variants were statistically significant across three pairwise comparison.

Our synteny analysis identified 13 genes in indicine cattle with less than 60% sequence identity compared to their taurine counterparts, based on pairwise alignment metrics generated by SyMAP [33, 34]. Among these, four genes were immune-related, suggesting potential divergence in immune function between breeds. 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 [35]. 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. [36]. 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 [37]. 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 [38]. 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.

Compared to taurine Hereford breed, Nelore cattle showed a significantly higher number of insertions on chromosomes 29, substitutions on chromosome 24 and deletions on chromosome X. Gir cattle exhibited a genomic differences on chromosome X with higher number of insertions and deletions, as well as substitutions on chromosome 24. These findings suggest chromosomes X exhibits notable genomic differences when compared to Hereford cattle. Although both Nelore and Gir are indicine breeds, the comparison between them revealed a higher number of insertions on chromosome 12, deletion on chromosome 29, and substitution on chromosome 28. Overall, substitutions were observed more frequently 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. Chromosome X is the second largest in the genome and harbours many genes that can influence various phenotypes [39, 40], a feature attributed to higher linkage disequilibrium than that seen in autosomes [41]. 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 28 accounted for a higher number of substitutions, 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 [40]. Notably, the region on Chromosome 13 associated with scrotal circumference had previously been linked to the phenotype of fat layer thickness in Nelore cattle [41].

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 color in cattle and other

domesticated mammals [27, 43, 44]. Several genes that activate the immune system, such as CD59 [45], CDH9 [46], PROCR [47], RLBP1 [48], BOLA [49], and STOM [44], respond to environmental stress and are strongly associated with responses to parasitic diseases as well [50]. Nelore macrophages control the intracellular replication of *Brucella abortus* more effectively, suggesting that, compared to Holstein, Nelore may possess greater natural resistance to brucellosis [51]. 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 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 [52], 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 [53]. 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 [49].

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 [54]. 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 [55, 56]. 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 [57]. Five of the genes – TBC1D23, NIT2, CELA1, BMPR1B, and HEXB – affect immune system pathways and inflammatory responses, and SPTBN1 is involved in the MAPK signaling pathway. These pathways are activated by inflammatory interleukins and cytokines released from damaged tissues [58, 59, 60]. Banos et al also observed correlations between traits related to immunity and those related to reproduction in dairy cattle.

The GO enrichment analysis revealed that genes with substitution variants were highly enriched in immune-related biological processes, including positive regulation of NF-kappa B signaling and the production of key cytokines such as interleukin (IL)-6, IL-1 beta, interferon (IFNs), tumor necrosis factor (TNF) and B cell proliferation. We identified 15 immune-related genes were involved in multiple biological processes. Specifically, TLR4 and TLR9 were associated with the regulation of IL-6, IFN-gamma, TNF, and B cell proliferation. RIPK2, CARD9, and NOD2 were involved in regulating IL-6, IFN-gamma, TNF, and NF-kB signaling, while FADD contributed to the regulation of IL-6, NF-kB, and TNF. Likewise, CLEC7A and RELA were involved in the positive regulation of IL-6, IFN-gamma, and NF-kB. CD40 played a role in NF-kB activation and B cell proliferation. TNF involved in regulating IL-6, IFN-gamma, and TNF production. Members of IRAK family (IRAK1/2/3) were also involved in NF-kB-related pathways, and TRIM25 regulate the production of IFN-gamma, NF-kB, and IL-6. Furthermore, BCL2 involved in B cell signaling and proliferation, which had both deletions and substitutions. Similarly, PLCG2, observed with both deletions and substitutions, was involved in B cell signaling as well as the positive regulation of IL-6 and TNF production. Collectively, these findings suggest that structural variants, particularly substitutions, significantly impact immune-related genes involved in NF-kB signaling, cytokine production, and B cell activation. In addition, genes carrying all three types of variants (substitutions, insertions, and deletions) were enriched in the biological process of endocytosis.

Consistent with the GO results, KEGG pathway enrichment analysis showed that genes with substitution variants were highly enriched in several immune-related pathways. These included the MAPK signaling pathway, Wnt signaling, Fc gamma R-mediated phagocytosis, T-cell receptor signaling, C-type lectin receptor signaling, chemokine signaling, Fc epsilon RI signaling, and pathways related to infections such as human papillomavirus, human T-cell leukemia virus 1, salmonella, yersinia, and bacterial invasion of epithelial cells. Other enriched pathways included inflammatory mediator regulation of TRP channels and Th17 cell differentiation. Genes with all three types of variants were also enriched in the T-cell receptor (TCR) signaling pathway. A total of 12 immune-related genes: CBLB, PIK3R1, AKT3, PAK3, NFATC2, MAPK14, NFKB1, VAV2, PTPRC, CD28, SOS1, and CARD11 were involved in this pathway. T cells are critical components of the adaptive immune system, and TCR signaling plays a central role in T cell development, homeostasis, activation, and tolerance. Proper signaling through the TCR is essential for initiating immune responses, and defects or dysregulation in this pathway have been implicated in the development of various immune-related diseases [61]. The identified genes, along with their associated biological processes and pathways, represent promising candidates for further investigation to better understand their role in the disease resistance traits of indicine 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 immune-related genes and its associated QTLs, biological functions and pathways. Our findings highlight significant genomic variations on chromosomes 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. The chromosome-by-chromosome analysis comparing Nelore with Hereford revealed 62 immune-related genes in Nelore having insertions, deletions, and substitution, along with their association with QTLs related to milk and health. GO and KEGG enrichment analyses showed significant enrichment of immune-related biological processes and pathways, especially those involving NF-kappa B signaling, cytokine production, and T cell receptor signaling. Key immune genes such as TLR4, TLR9, RIPK2, PLCG2, FADD, CLEC7A, RELA, CD40, TRIM25, IRAK1/2/3, NOD2, BCL2, and CARD9 were identified as central players involved in multiple biological process. Altogether, these findings not only extend our understanding of the genetic basis for disease resistance in Indian cattle (Nelore breed) but also provide valuable information for developing future breeding strategies that could enhance the immunity and overall productivity of indicine cattle.

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

Acknowledgment: 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.

Data Availability: The datasets used in this analysis were sourced from previously submitted datasets in the NCBI database, as cited in the manuscript. The output data relevant to this study is provided within the manuscript, and all generated output files and utilized in-house Python script are available in the GitHub repository: https://github.com/raghuyennamalli/cattle_immunoinformatics

References

[1] Serrano, G. M., Egito, A. A., McManus, C., & Mariante, A. da. (2004). Genetic diversity and population structure of Brazilian native bovine breeds. *Pesquisa Agropecuária Brasileira*, *39*(6), 543–549. https://doi.org/10.1590/s0100-204x2004000600005

[2] Verma, S., Thakur, A., Katoch, S., Shekhar, C., Wani, A. H., 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. https://doi.org/10.1016/j.vetimm.2017.09.003

[3] HANSEN, P. (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science*. https://doi.org/10.1016/s0378-4320(04)00066-1

[4] Mwai, O., Hanotte, O., Kwon, Y.-J., & Cho, S. (2015). - invited review - african indigenous cattle: Unique genetic resources in a rapidly changing world. *Asian-Australasian Journal of Animal Sciences*, *28*(7), 911–921. https://doi.org/10.5713/ajas.15.0002r

[5] Sharma, R., Kishore, A., Mukesh, M., Ahlawat, S., Maitra, A., Pandey, A. K., & Tantia, M. S. (2015). Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. *BMC Genetics*, *16*(1). https://doi.org/10.1186/s12863-015-0221-0

[6] Pramod, R. K., Velayutham, D., P. K., S., P. S., B., Zachariah, A., Zachariah, A., B., C., S. S., 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. https://doi.org/10.1080/10495398.2018.1476376

[7] Shivakumara, P. N., Aravindakshan, T. V., 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. https://doi.org/10.1016/j.mgene.2018.01.003

[8] Janeway, C. A., & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, *20*(1), 197–216. https://doi.org/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. https://doi.org/10.1146/annurev.immunol.21.120601.141126

[10] Mishra, A., Vijayakumar, P., & Raut, A. A. (2017). Emerging avian influenza infections: Current understanding of innate immune response and molecular pathogenesis. *International Reviews of Immunology*, *36*(2), 89–107. https://doi.org/10.1080/08830185.2017.1291640

[11] Ferluga, J., Yasmin, H., Al-Ahdal, M. N., Bhakta, S., & Kishore, U. (2020). Natural and trained innate immunity against mycobacterium tuberculosis. *Immunobiology*, *225*(3), 151951. https://doi.org/10.1016/j.imbio.2020.151951

[12] Vlasova, A. N., & Saif, L. J. (2021). Bovine immunology: Implications for dairy cattle. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.643206

[13] Kumar, A., Kumar Gaur, G., Panigrahi, M., V. N., M. A., Priya, B., Gupta, J. P., & Gandham, R. K. (2022). Comparative gene expression profile in circulating pbmcs of *Bos*

indicus and crossbred cattle to understand disease tolerance mechanism. *Animal Biotechnology*, *34*(4), 1594–1602. https://doi.org/10.1080/10495398.2022.2043883

[14] Canavez, 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. https://doi.org/10.1093/jhered/esr153.

[15] Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A., & Salzberg, S. L. (2009). A whole-genome assembly of the domestic cow, *Bos Taurus*. *Genome Biology*, 10(4). https://doi.org/10.1186/gb-2009-10-4-r42.

[16] Vijayakumar, P., Singaravadivelan, A., Mishra, A., Jagadeesan, K., Bakyaraj, S., Suresh, R., & Sivakumar, T. (2022a). Whole-genome comparative analysis reveals genetic mechanisms of disease resistance and heat tolerance of tropical *Bos indicus*cattle breeds. *Genome*, *65*(4), 241–254. https://doi.org/10.1139/gen-2021-0030

[17] Chakraborty, A., Bisht, M. S., Saxena, R., Mahajan, S., Pulikkan, J., & Sharma, V. K. (2023). Genome sequencing and de Novo and reference-based genome assemblies of *Bos indicus* breeds. *Genes & Genomics*, 45(11), 1399–1408. https://doi.org/10.1007/s13258-023-01401-w

[18] Koufariotis, L.T. et al. (2018) 'Variance explained by whole genome sequence variants in coding and regulatory genome annotations for six dairy traits', *BMC Genomics*, 19(1). doi:10.1186/s12864-018-4617-x.

[19] Álvarez Cecco, P. et al. (2022) 'Genome□wide scan for signatures of selection in the Brangus Cattle Genome', *Journal of Animal Breeding and Genetics*, 139(6), pp. 679–694. doi:10.1111/jbg.12733.

[20] Naji, M.M. et al. (2021) 'Investigation of ancestral alleles in the Bovinae subfamily', *BMC Genomics*, 22(1). doi:10.1186/s12864-021-07412-9.

[21] Dorji, J. et al. (2024) 'Ancestral alleles defined for 70 million cattle variants using a population-based likelihood ratio test', *Genetics Selection Evolution*, 56(1). doi:10.1186/s12711-024-00879-6.

[22] Utsunomiya, Y.T. et al. (2013) 'Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods', *PLoS ONE*, 8(5). doi:10.1371/journal.pone.0064280.

[23] 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. https://doi.org/10.1101/gr.5396706

[24] Lin, H.-N., & Hsu, W.-L. (2020). GSALIGN: An efficient sequence alignment tool for intra-species genomes. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-020-6569-1

[25] Goel, M., Sun, H., Jiao, W.-B., & Schneeberger, K. (2019). Syri: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1911-0

[26] Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). David: A web server for functional enrichment analysis and functional

annotation of gene lists (2021 update). *Nucleic Acids Research*, 50(W1). https://doi.org/10.1093/nar/gkac194

[27] Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using David Bioinformatics Resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

[28] Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). Quast: Quality Assessment Tool for Genome Assemblies. *Bioinformatics*, 29(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

[29] 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.

[30] Hu, Y., Xia, H., Li, M., Xu, C., Ye, X., Su, R., Zhang, M., Nash, O., Sonstegard, T. S., Yang, L., Liu, G. E., & Zhou, Y. (2020). Comparative analyses of copy number variations between *Bos taurus* and *Bos indicus. BMC Genomics*, 21(1). https://doi.org/10.1186/s12864-020-07097-6

[31] Gleeson, P. A. (2014). The role of endosomes in innate and adaptive immunity. *Seminars in Cell & Developmental Biology*, 31, 64–72. https://doi.org/10.1016/j.semcdb.2014.03.002

[32] Staal, F. J., Luis, T. C., & Tiemessen, M. M. (2008). Wnt signalling in the immune system: WNT is spreading its wings. *Nature Reviews Immunology*, 8(8), 581–593. https://doi.org/10.1038/nri2360

[33] Tian, W. and Skolnick, J. (2003) 'How well is enzyme function conserved as a function of pairwise sequence identity?', Journal of Molecular Biology, 333(4), pp. 863–882. doi:10.1016/j.jmb.2003.08.057.

[34] Apatoff, A., Kim, E. and Kliger, Y. (2006) 'Towards alignment independent quantitative assessment of homology detection', PLoS ONE, 1(1). doi:10.1371/journal.pone.0000113.

[35] Oludada, O. E., Idowu, T. O., 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. https://doi.org/10.1016/j.bbrc.2022.10.057

[36] Schlichtner, S., Meyer, N. H., Yasinska, I. M., Aliu, N., Berger, S. M., Gibbs, B. F., Fasler-Kan, E., & Sumbayev, V. V. (2021). Functional role of galectin-9 in directing human innate immune reactions to gram-negative bacteria and T cell apoptosis. *International Immunopharmacology*, *100*, 108155. https://doi.org/10.1016/j.intimp.2021.108155

[37] Licciardi, P. V., & Karagiannis, T. C. (2012). Regulation of immune responses by histone deacetylase inhibitors. *ISRN Hematology*, *2012*, 1–10. https://doi.org/10.5402/2012/690901

[38] 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. https://doi.org/10.1084/jem.20050292.

[39] Lyons, R. E., Loan, N. T., Dierens, L., Fortes, M. R., Kelly, M., McWilliam, S. S., Li, Y., Bunch, R. J., Harrison, B. E., Barendse, W., Lehnert, S. A., & Moore, S. S. (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). https://doi.org/10.1186/1471-2156-15-6

[40] Johnson, T., Keehan, M., Harland, C., Lopdell, T., Spelman, R. J., Davis, S. R., Rosen, B. D., Smith, T. P. L., & Couldrey, C. (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. https://doi.org/10.3168/jds.2018-15638

[41] de Camargo, G. M., Porto-Neto, L. R., Kelly, M. J., Bunch, R. J., McWilliam, S. M., Tonhati, H., Lehnert, S. A., Fortes, M. R., & Moore, S. S. (2015). Non-synonymous mutations mapped to chromosome X associated with Andrological and growth traits in beef cattle. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1595-0

[42] 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 study for indicator traits of sexual precocity in Nellore cattle. *PLOS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0159502.

[43] Liu, G. E., & Bickhart, D. M. (2012). Copy number variation in the cattle genome. *Functional & Integrative Genomics*, *12*(4), 609–624. https://doi.org/10.1007/s10142-012-0289-9

[44] Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., Agaba, M., Kim, K., Kwak, W., Sung, S., Seo, M., Jeong, H., Kwon, T., Taye, M., Song, K.-D., Lim, D., Cho, S., Lee, H.-J., Yoon, D., … Kim, H. (2017). The genome landscape of indigenous African cattle. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1153-y

[45] Chan, E. K., Nagaraj, S. H., & Reverter, A. (2010). The evolution of tropical adaptation: Comparing taurine and zebu cattle. *Animal Genetics*, *41*(5), 467–477. https://doi.org/10.1111/j.1365-2052.2010.02053.x

[46] 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 Selection Evolution*, *49*(1). https://doi.org/10.1186/s12711-017-0358-6

[47] Sorbolini, S., Marras, G., Gaspa, G., Dimauro, C., Cellesi, M., Valentini, A., & Macciotta, N. P. (2015). Detection of selection signatures in Piemontese and marchigiana cattle, two breeds with similar production aptitudes but different selection histories. *Genetics Selection Evolution*, *47*(1). https://doi.org/10.1186/s12711-015-0128-2

[48] Smetko, A., Soudre, A., Silbermayr, K., Müller, S., Brem, G., Hanotte, O., Boettcher, P. J., Stella, A., Mészáros, G., Wurzinger, M., Curik, I., Müller, M., Burgstaller, J., & Sölkner, J. (2015). Trypanosomosis: Potential driver of selection in African cattle. *Frontiers in Genetics*, *6*. https://doi.org/10.3389/fgene.2015.00137

[49] Oddgeirsson, O., Simpson, S. P., Morgan, A. L., Ross, D. S., & Spooner, R. L. (1988). Relationship between the bovine major histocompatibility complex (BOLA), erythrocyte markers and susceptibility to mastitis in Icelandic cattle. *Animal Genetics*, *19*(1), 11–16. https://doi.org/10.1111/j.1365-2052.1988.tb00783.x

[50] Mei, C., Gui, L., Hong, J., Raza, S. H., Aorigele, C., Tian, W., Garcia, M., Xin, Y., Yang, W., Zhang, S., & Zan, L. (2021). Insights into adaption and Growth Evolution: A Comparative Genomics study on two distinct cattle breeds from northern and southern China. *Molecular Therapy - Nucleic Acids*, *23*, 959–967. https://doi.org/10.1016/j.omtn.2020.12.028

[51] Macêdo, A. A., Marciano, A. P. V., Rocha, L. M., Alves-Júnior, J. R. F., Faria, A. M. C., Bittar, J. F. F., Araújo, M. S. S., Santos, R. L., & Martins-Filho, O. A. (2013a). Comparative phenotypic profile of subpopulations of peripheral blood leukocytes in European (*Bos taurus* Taurus) and zebu cattle (*Bos taurus indicus*). *Genetics and Molecular Research*, *12*(4), 6838–6849. https://doi.org/10.4238/2013.december.19.2

[52] Sadler, A. J., & Williams, B. R. (2008). Interferon-inducible antiviral effectors. *Nature Reviews Immunology*, 8(7), 559–568. https://doi.org/10.1038/nri2314

[53] Maiorano, A. M., Cardoso, D. F., Carvalheiro, R., Júnior, G. A., de Albuquerque, L. G., & de Oliveira, H. N. (2022). Signatures of selection in Nelore cattle revealed by whole-genome sequencing data. *Genomics*, *114*(2), 110304. https://doi.org/10.1016/j.ygeno.2022.110304

[54] de Oliveira, P. S., Cesar, A. S., do Nascimento, M. L., Chaves, A. S., Tizioto, P. C., Tullio, R. R., Lanna, D. P., Rosa, A. N., Sonstegard, T. S., Mourao, G. B., Reecy, J. M., Garrick, D. J., Mudadu, M. A., Cootinho, L. L., & Regitano, L. C. (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genetics*, *15*(1). https://doi.org/10.1186/s12863-014-0100-0

[55] Tizioto, P. C., Decker, J. E., Taylor, J. F., Schnabel, R. D., Mudadu, M. A., Silva, F. L., Mourão, G. B., Coutinho, L. L., Tholon, P., Sonstegard, T. S., Rosa, A. N., Alencar, M. M., Tullio, R. R., Medeiros, S. R., Nassu, R. T., Feijó, G. L., Silva, L. O., Torres, R. A., Siqueira, F., Regitano, L. C. (2013). Genome scan for meat quality traits in Nelore beef cattle. *Physiological Genomics*, *45* (21), 1012–1020. https://doi.org/10.1152/physiolgenomics.00066.2013

[56] Silva, V. H., Regitano, L. C., Geistlinger, L., Pértille, F., Giachetto, P. F., Brassaloti, R. A., Morosini, N. S., Zimmer, R., & Coutinho, L. L. (2016a). Genome-wide detection of cnvs and their association with meat tenderness in Nelore cattle. *PLOS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0157711

[57] Montes, D. E., Braz, C. U., Ribeiro, A. M. F., Cavani, L., Barbero, M. M. D., Albuquerque, L. G., Curi, R. A., & Oliveira, H. N. (2019). Selection signatures in candidate genes and QTL for reproductive traits in Nellore heifers. *Animal Reproduction Science*, *207*, 1–8. https://doi.org/10.1016/j.anireprosci.2019.06.004

[58] Park, J., Min, J.-S., Kim, B., Chae, U.-B., Yun, J. W., Choi, M.-S., Kong, I.-K., Chang, K.-T., & Lee, D.-S. (2015). Mitochondrial Ros govern the LPS-induced pro-inflammatory response in microglia cells by regulating MAPK and NF-KB PATHWAYS. *Neuroscience Letters*, *584*, 191–196. https://doi.org/10.1016/j.neulet.2014.10.016

[59] Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z., Deng, J., Li, Y., Wang, X., & Zhao, L. (2017). Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget*, *9*(6), 7204–7218. https://doi.org/10.18632/oncotarget.23208

[60] Banos, G., Wall, E., Coffey, M. P., Bagnall, A., Gillespie, S., Russell, G. C., & McNeilly, T. N. (2013). Identification of immune traits correlated with dairy cow health, reproduction and productivity. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0065766

[61] Thaventhiran, T. (2013). T cell co-inhibitory receptors-functions and signalling mechanisms. *Journal of Clinical & Cellular Immunology*, 01(S12). https://doi.org/10.4172/2155-9899.s12-005