The Journal of Agricultural Science

www.cambridge.org/ags

Copy number variation profiling in the genome of crossbred dairy cattle from Pakistan

Fakhar Un Nisa[1], Muhammad Asif[2,3], Qamar Un Nisa[4], Rubab Zahra Naqvi[2,3], Muhammad Saif Ur Rehman[1] and Zahid Mukhtar[2,3]

[1]Institute of Animal and Dairy Sciences, Faculty of Animal Husbandry, University of Agriculture, Faisalabad, Pakistan; [2]Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering College (NIBGE-C), Faisalabad, Pakistan; [3]Pakistan Institute of Engineering & Applied Sciences (PIEAS), Islamabad, Pakistan and [4]Department of Veterinary Pathology, University of Veterinary and Animal Sciences, Lahore, Pakistan

Animal Research Paper

Cite this article: Nisa FU, Asif M, Nisa QU, Naqvi RZ, Rehman MSU, and Mukhtar Z (2025). Copy number variation profiling in the genome of crossbred dairy cattle from Pakistan. *The Journal of Agricultural Science* **163**, 585–596. https://doi.org/10.1017/S0021859625100208

Received: 7 January 2025
Revised: 7 January 2025
Accepted: 14 April 2025
First published online: 31 July 2025

Keywords:
Copy number variations (CNV); CNV regions (CNVR); HandyCNV; PennCNV; SNP array

Corresponding authors: Fakhar Un Nisa; Emails: dr.fakharunnisa@gmail.com, fakhar.nisa@uaf.edu.pk; Zahid Mukhtar; Email: zahidmukhtar@yahoo.com

Abstract

The investigation of structural variants that may govern complex traits has significant importance. This is particularly true for the crossbred dairy cattle of Pakistan, which are deemed ideal for achieving optimal milk production and enhanced environmental adaptability in tropical conditions. This research detected and described copy number variation regions (CNVR) within the crossbred cattle genome. A GGP_HDv3_C chip containing 139,376 SNPs was utilized to genotype a cohort of 81 animals. In this study, 1055 CNVs were obtained after quality control, distributed across animals and encompassing all autosomes. From these, 268 CNVRs were detected, which covered 31.03 megabases, representing approximately 1.24% of the bovine genome. Functional analysis of these regions yielded 97 genes primarily associated with the immune and defense systems. Additionally, other observed categories encompassed production, health and reproduction. These findings enhanced the CNV map of bovines, offering the variant identification linked to traits subject to selection in both crossbred and indicine breeds of cattle.

Introduction

Copy number variations (CNVs) allude to genetic modifications that include deletions, duplications and insertions surpassing a size threshold of 50 base pairs (bp). These modifications intricately reshape the Deoxyribonucleic acid (DNA) architecture, exerting a profound influence on genomic diversity, a phenomenon readily apparent both within specific breeds and across diverse populations (Letaief *et al.*, 2017). CNVs have been observed to influence a greater proportion of genomic sequences compared to other types of genomic variations, such as single-nucleotide polymorphisms (SNPs) (Geistlinger *et al.*, 2018; Zhou *et al.*, 2016; Zhao *et al.*, 2013; Liu and Bickhart, 2012; Hou *et al.*, 2011; Zhang *et al.*, 2009). They can also affect the expression of adjacent genes, even when they may not be inherently connected through linkage disequilibrium (LD). CNVs and CNV regions (CNVRs) have been linked to both qualitative and quantitative traits across various animal species (da Silva *et al.*, 2016). Changes within CNVRs can appear as either copy number gains, copy number losses, or mixed types, involving both gain and loss simultaneously (Butty *et al.*, 2021).

CNVs have the potential to induce significant phenotypic variations through a diverse array of mechanisms, that is, through gene dosage effect, alterations in gene expression levels, gene blocking effects, gene fusion events, positional effects, the activation of previously dormant alleles, functional polymorphisms and the possibility of compounded effects (Zhang *et al.*, 2018b). They might represent the basis upon which evolutionary mechanisms can exert their influence (Emerson *et al.*, 2008). About 50% of recognized CNVs from humans involve protein coding regions, acknowledged for their roles in fundamental cellular processes, overall metabolism and the initiation of various diseases and disease susceptibility (Sebat *et al.*, 2004; Gupta *et al.*, 2015; Cooper *et al.*, 2011; Casey *et al.*, 2012; Almal and Padh, 2012; El-Sayed Moustafa *et al.*, 2012; Conrad *et al.*, 2010; Park *et al.*, 2015; Pinto *et al.*, 2010). Modifications in CNVs are identified in cancerous tissues (Gupta *et al.*, 2015; Park *et al.*, 2015; Ouyang *et al.*, 2014; Malek, 2013; Verma *et al.*, 2013) and have been linked to different other traits (Park *et al.*, 2015; Zarrei *et al.*, 2015; Butty *et al.*, 2021; Kang *et al.*, 2020; Di Gerlando *et al.*, 2019; Zhou *et al.*, 2018; da Silva *et al.*, 2016; Jakobsson *et al.*, 2008).

Historically, the identification of CNVs at the cytogenetic level utilized techniques such as Fluorescent In Situ Hybridization (FISH) and chromosomal karyotyping (Zhao *et al.*, 2013). The majority of extensive population-based studies for CNV detection primarily utilize two approaches: SNP genotyping panels and comparative genomic hybridization arrays (CGH) (Zhang *et al.*, 2014; Cicconardi *et al.*, 2013; Xu *et al.*, 2016). The merits and drawbacks linked to them are thoroughly discussed in the literature (Pinto *et al.*, 2011; Ionita-Laza *et al.*, 2009; Curtis *et al.*, 2009). Among these methods, the use of SNP arrays with varying densities is advantageous

in diverse livestock species because of their high-throughput nature and cost-effectiveness. Additionally, the incorporation of the B-allele frequency (BAF) and Log R Ratio (LRR) parameter further aids in result interpretation (Fadista *et al.*, 2008). Numerous computational tools and methods have been developed for the exhaustive analysis of CNVs on a genome-wide scale. Notably, the Hidden Markov Model (HMM) utilized in PennCNV stands out as a highly accurate approach for CNV detection, recognized for its heightened specificity and sensitivity (Pierce *et al.*, 2018; Geistlinger *et al.*, 2018; Zhang *et al.*, 2014).

Studies focused on identifying CNVs have been successfully conducted in economically significant animal species. These species encompass cattle (Jiang *et al.*, 2013; Xu *et al.*, 2014a; Bickhart *et al.*, 2012; Hou *et al.*, 2012a; Hou *et al.*, 2012b; Jiang *et al.*, 2012; Liu and Bickhart, 2012; Liu *et al.*, 2011; Liu *et al.*, 2010; Liu *et al.*, 2009), goat (Fontanesi *et al.*, 2010), sheep (Fontanesi *et al.*, 2011; Liu *et al.*, 2013) and buffalo (Ahmad *et al.*, 2023; Dash *et al.*, 2023; Kumar *et al.*, 2023; Liu *et al.*, 2019; Strillacci *et al.*, 2021; Yang *et al.*, 2023; Zhang *et al.*, 2022). A substantial quantity of CNVs have been observed in both indicine and taurine cattle breeds, particularly within genes and genomic regions influencing complex and quantitative traits (Zhang *et al.*, 2014; Cicconardi *et al.*, 2013; Bickhart *et al.*, 2012; Hou *et al.*, 2012c; Jiang *et al.*, 2012; Jiang *et al.*, 2013). Notably, there is a stronger overlap of CNVs reported among different taurine breeds compared to the overlap seen when indicine and taurine cattle are compared. Interestingly, indicine breeds displayed the greatest CNV diversity among all (Bickhart *et al.*, 2012).

In the context of cattle, CNVs have been associated with diverse traits, including parasite resistance (Hou *et al.*, 2012c), growth characteristics (Zhang *et al.*, 2018b; Xu *et al.*, 2014b), reproduction (Yue *et al.*, 2014; Sasaki *et al.*, 2016), milk production and composition (da Silva *et al.*, 2016; Gao *et al.*, 2017), milk somatic cell scores (Durán Aguilar *et al.*, 2017), meat quality (da Silva *et al.*, 2016; de Lemos *et al.*, 2018) and feed conversion ratios (de Almeida Santana *et al.*, 2016).

The identification of CNVs and CNVRs within crossbred cattle populations holds the potential to unveil specific genetic segments responsible for variations in critical economic traits (Liu *et al.*, 2024). A lot of CNV-related literature is available; however, only a few studies have been conducted to explore the crossbred cattle genomics of tropical regions like Pakistan (Chen *et al.*, 2024). The primary focus of the current study was to construct a comprehensive genome-wide CNV map for crossbred cattle employing SNP genotyping techniques, aiming to facilitate genetic enhancements and delve into the genetic foundations of improved production and environmental adaptability.

Previously, the signatures of selection and LD parameters were done using the same dataset (Nisa *et al.*, 2023; Nisa *et al.*, 2024). Sahiwal, a tropical dairy cattle, is well recognized for disease resistance and heat tolerance (Iqbal *et al.*, 2019), but the lower production is of great concern. The import of high-yielding dairy animals, like HF, is rising in Pakistan to mitigate the production-related issues. Crossbreeding Sahiwal and HF is a highly efficient method to bolster livestock productivity with improved sustainability and reproductive ability (Leroy *et al.*, 2016; Mbole-Kariuki *et al.*, 2014; Bebe *et al.*, 2003). This crossbreeding yields progeny that harness the benefits of hybrid vigour (Kumar et al., 2018). It may combine the high production yield of HF and the adaptability and heat tolerance of Sahiwal into a single individual, with improved production and adaptability.

Material and methods

Ethics statement

To guarantee the ethical and compassionate treatment of animals, the investigation outlined here received approval from the research ethics committee of the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan, on 10 June 2020. A qualified veterinarian supervised the blood collection process to minimize discomfort to the animals. Before sample collection, the researchers conducted a meeting with the farmers to elaborate on the objective of the investigation and secured verbal acknowledgment of consent.

Sample collection and data generation

The study sample comprised 81 crossbred cattle. The animals were selected based on varying proportions of HF and Sahiwal genetics across different lactations. Due to the inherent variability in crossbreeding, breed composition differed among individuals, with some possessing approximately 50% HF and 50% Sahiwal inheritance, while others had up to 31/32 HF ancestry, with the remainder from Sahiwal. Genotyping was done using the GGP_HDv3_C chip (GeneSeek® Genomic Profiler™) and commercially available services at GeneSeek (Neogen Corporation, Lincoln, NE, United States). The details about the blood sample collection, crossbred composition, DNA extraction, its qualitative and quantitative assessment and genotyping detail is mentioned in previous studies on the same dataset (Nisa *et al.*, 2024; Nisa *et al.*, 2023). The genotypes were originally discerned utilizing Illumina, Inc.'s Genome Studio. The examination was conducted based on the ARS-UCD1.2 bovine genome assembly.

Quality control (QC)

After genotyping, raw data consisted of 139,376 SNPs. QC was performed utilizing the PLINK v1.9 software as outlined by Purcell *et al.* (2007). This involved eliminating SNPs with a call rate of < 95%, a minor allele frequency (MAF) of < 0.02 and a Hardy-Weinberg equilibrium (HWE) of < 10E−05. Downstream analysis considered autosomal SNPs only.

Calling copy number variations (CNVs) and copy number variation regions (CNVRs)

In this study, the Genome Studio v2.0.5 software developed by Illumina was utilized to extract pertinent information, including BAF and LRR, from the signal intensity data of the genotyped samples. Notably, the genotyping data exhibited a minimal rate of missing values, boasting an impressive genotyping rate of 99.7%.

For the crucial task of CNV detection, the PennCNV programme was used (Wang *et al.*, 2007). This programme leverages the power of HMM for the accurate identification of CNVs. To facilitate the analysis, the Compile_pfb script within PennCNV was utilized. This script allowed the generation of a comprehensive genome-wide Population Frequency of B Allele (PFB) file, primarily derived from the BAF associated with each SNP.

To further refine the data analysis, Kcolumn, a Perl script within PennCNV, was employed. This script was instrumental in the segmentation and organization of the information pertaining to LRR, BAF and PFB. It is noteworthy that, due to the unavailability or incompleteness of complete pedigree information, the '–test' option was chosen for the CNV calling process,

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

ensuring the robustness of the analysis in situations where pedigree information was lacking or not fully utilized in the study.

CNVs were identified using the intensity files in the Perl script detect_cnv supplied by the PennCNV. QC for CNVs adhered to stringent criteria, necessitating a low LRR standard deviation (SD) of less than 0.3, with a minimal BAF drift of less than < 0.01 and a GC wave factor of less than 0.05.

It is important to highlight that although the PennCNV was basically designed for humans, essential modifications were incorporated while analysing to accommodate the extra chromosomes of the bovine genome. All other parameters and settings of PennCNV were retained at their default values during CNV calling.

The identified CNVs were subsequently categorized into discrete intervals, referred to as CNVRs. This choice was made to define CNVRs more naturally, encompassing intervals with overlapping CNVs that did not surpass the average size of the CNV+1SD. CNVRs were constructed using the CNVRuler programme with default parameters, as outlined by Kim *et al.* (2012b).

It is important to note that CNVRuler offers three distinct methods for defining CNVRs: CNVR, Reciprocal Overlap (RO) and Fragment. In this study, the CNVR method was selected. To ensure accuracy, a recurrence value of 0.3 was set to trim sparse regions of overlap, preventing the overestimation of CNVR size and frequency.

For validation purposes, the same CNVRs were also obtained using the HandyCNV package within R (Zhou *et al.*, 2021). Three categories of CNVRs were identified specifically: loss, gain and mixed.

Functional annotation

The automated annotation of genes located in the identified CNVs and CNVRs was conducted using the handyCNV package in R, specifically employing the call_gene function. However, before utilizing this function, the preparation of gene lists against the correct reference genome, namely ARS-UCD1.2, was imperative. This preparatory step was accomplished through the get_refgene function.

Two distinct gene lists were compiled. The first one drew upon data from UCSC, while the second was created using information sourced from the Ensembl Genome browser. After individually extracting information from both browsers, a comparative analysis was undertaken between the results to identify a set of consensus genes. Additionally, the Ensemble database, specifically Ensembl gene 110, was accessed via Biomart for the same annotation purpose.

Quantitative trait loci (QTL) detection within CNVRs

CNVRs were additionally scrutinized for their potential association with significant QTLs affecting various economically important traits. This evaluation utilized the CattleQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index). Genomic coordinates were employed to identify QTLs and genes that exhibited spatial overlap within CNVRs.

Additionally, annotation using Gene Ontology (GO) was performed using the DAVID platform (https://david.ncifcrf.gov/tools.jsp) (Huang *et al.*, 2021). This approach offered insights into the biological functions and pathways associated with the genes located within CNVs and CNVRs.

Comparison of CNVR with previous studies

To compare CNVRs found in this study with previously reported studies, autosomal CNVs from eight studies were retrieved from the Database of Genomic Variants Archive (DGVa) at EMBL-EBI (accessed on 30 September 2023). They were juxtaposed with CNVRs identified in this study. The study populations in these studies are mainly of taurine breeds; however, in 3 datasets, we got some samples from indicine breeds as well (Liu *et al.*, 2010; Hou *et al.*, 2011; Karimi *et al.*, 2017). One study detected CNV using CGH (Liu *et al.*, 2010), three studies used SNP Chip data (Hou *et al.*, 2011; Karimi *et al.*, 2017) and four studies used WGS data (Bickhart *et al.*, 2012; Boussaha *et al.*, 2015; Keel *et al.*, 2017; Mesbah-Uddin *et al.*, 2018). The studies encompassed a variable number of breeds, ranging from 1 to 21, with sample sizes ranging from 6 to 539. To compile the DGVa CNVR set, information including study details, type of CNV, chromosome, start and end position was extracted.

CNVs in these studies were identified using UMD3.1 and UMD3.1.1 assemblies of bovines. The coordinates from different assemblies were first converted to ARS-UCD1.2 using the LiftOver tool of UCSC Genome Browser (Navarro Gonzalez *et al.*, 2021). The minimum threshold for the ratio of bases requiring remapping was established at 0.4 (Butty *et al.*, 2020) and for all other LiftOver parameters, default values were applied.

After translation to ARS-UCD1.2 positions, CNVs that shared a minimum overlap of 1bp were merged. The DGVa CNVR set resulted in a total of 9243 CNVRs. CNVRs from our dataset were considered equivalent to those from the DGVa if the RO between them was at least 50%.

Results

In the current study, GGP_HDv3_C array data from 81 crossbred animals were employed to detect CNVs and CNVRs. The HMM within the PennCNV program was applied for this purpose. Initially, a total of 1206 CNVs were identified within the crossbred dataset. After a rigorous filtering process, 1055 CNVs were retained, distributed across animals and encompassing all autosomes (Supplementary File 1).

The observed CNV count per animal ranged from a minimum of 3 to a maximum of 37, with an average of 13.88 CNVs per animal. Btau11 displayed the highest number of CNVs, occurring at 97 distinct genomic locations. Conversely, Btau23 and Btau27 in the bovine genome exhibited the lowest number of CNVs, each containing only three CNVs.

The total regions displaying losses and gains in relation to the normal copy number (CNV = 2) amounted to 129 and 926, respectively. The size of the filtered CNVs displayed considerable variation, ranging from 2.9 kilobases (kb) to 1108.7 kb. The average CNV length was approximately 184.502 kb, with a median length of 133.472 kb. The relationship between CNV types and their respective lengths in kb was estimated (Figure 1), providing a visual representation of CNV distribution across the genome (Supplementary File 2). The box plot illustrates the distribution of CNV lengths across four CNV types (0, 1, 3 and 4). Notably, types 3 and 4 exhibit a broader range of CNV lengths and a higher number of outliers compared to types 0 and 1, indicating greater variability. The median CNV length is higher for type 3, while type 0 shows the least variability and few outliers, suggesting a more consistent CNV length distribution.

The distribution of CNV sizes is summarized (Table 1). Notably, nearly half (44%) of the CNVs fell within the size range of

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

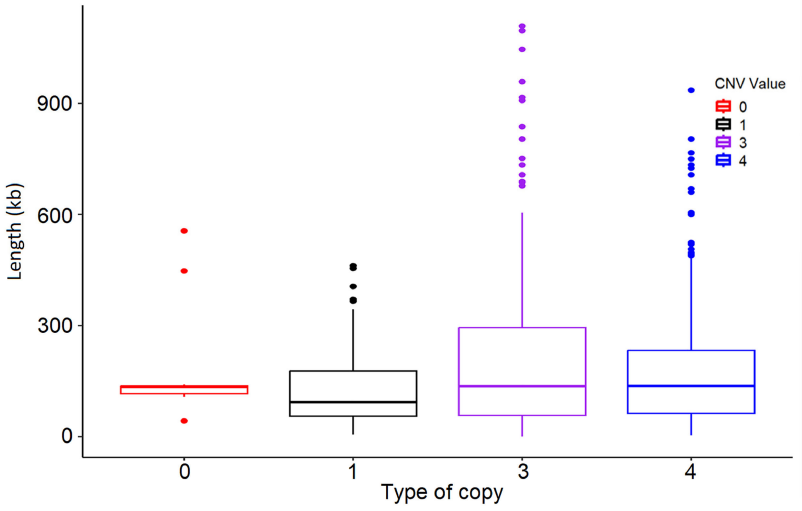

Figure 1. Box plot showing CNV lengths (kb) across four categories (0, 1, 3, 4). The boxes represent the interquartile range (IQR), with medians marked as horizontal lines; whiskers extend to 1.5×IQR and dots represent outliers. CNV categories vary in data distribution, with category 0 showing fewer data points and higher outliers compared to categories 3 and 4.

Table 1. Summary of CNV based on size in kilobases (kb)

Size (kb)	No. of CNVs	Percentages (%)	No. of CNVRs	Percentages (%)
0–100	463	43.96	176	65.67
101–200	241	22.88	45	16.79
201–300	137	13.01	20	7.46
301–400	90	8.54	14	5.22
401–500	61	5.79	11	4.10
>500	61	5.79	2	0.74

CNVs, copy number variations; CNVRs, copy number variation regions.

0 to 100 kb. CNVs in the 300–400 kb range were less common, while those exceeding 400 kb in size were relatively rare.

A summary plot of CNVs, displaying results categorized by length group, CNV type and chromosome, was generated using the HandyCNV tool (Figure 2). It depicts the distribution of CNVs across different chromosomes, categorized by CNV types (0, 1, 3, 4). Each line represents the number of CNVs for each chromosome, with different colours corresponding to distinct CNV values. The data show notable peaks in CNV counts for chromosomes 5, 14, 24 and 26, particularly for CNV types 3 (purple) and 4 (blue), suggesting a higher prevalence of these CNVs on these chromosomes. Meanwhile, CNV types 0 (red) and 1 (black) are less frequent and display lower variation across chromosomes (Supplementary File 3).

Likewise, each copy plot is differentiated based on its specific copy number (Figure 3). In it, the frequency and length distribution of CNVs across chromosomes for each CNV type (0, 1, 3, 4) using box plots is mentioned. The top panel (CNV type 0, red) shows a significant peak in frequency on chromosome 12. For CNV type 1 (second panel, black), chromosomes 12 and 25 exhibit elevated CNV frequencies and lengths. CNV type 3 (third panel, purple) displays a widespread distribution, with higher frequencies on chromosomes 5, 10 and 24, while CNV type 4 (bottom panel, blue) highlights chromosomes 5, 14, 24 and 26 as hotspots for CNV occurrence. The range of CNV lengths is greater for types 3 and 4, as indicated by the larger spread in the box plots (Summary plots are indicated in Supplementary File 4).

CNVRs are defined as genomic segments containing one or more CNVs that exhibit at least a single base pair of overlap. Consequently, CNVRs do not overlap with one another. We performed the merging of overlapping CNVs using two distinct approaches: CNVRuler and the HandyCNV package in R (Supplementary Files 5 and 6). As minor modifications were observed therefore in this study the CNVRs obtained using CNVRuler were mainly under consideration.

When employing the CNVRuler software (Kim *et al.*, 2012a), a total of 268 CNVRs were identified (Supplementary File 5). The majority of these CNVRs (65.67%) fell within the size range of 0 to 100 kb, with 16.79% ranging from 100 to 200 kb. Overall, the CNVRs ranged in size from 3.801 kb to 915.979 kb, with an average size of approximately 115.7949 kb. Among the 268 identified CNVRs, 212 represented gain events, 44 were indicative of loss events and 14 CNVRs comprised a combination of both gain and loss events. Detailed distributions of autosomal CNVRs are presented in Table 2 and Figure 4.

The cumulative length of the identified CNVRs amounted to 31.03 megabases (Mb), representing approximately 1.24% of the entire genome. It is important to note that the chromosome sizes were sourced from the most recent cattle assembly, ARS-UCD1.3.

The distribution of CNVRs across chromosomes exhibited variability, with the number of CNVRs per chromosome ranging from 0 on BTA23 and BTA24 to 20 on BTA7 and BTA19. The proportion of CNVRs as a fraction of the total chromosome length displayed a spectrum, ranging from 4.72% on BTA25 to 0% on BTA23 and BTA24.

Figure 5 is showing the CNVR map showing the distribution of CNVR across chromosomes (Supplementary File 7).

CNVR annotation

Annotations were performed separately using information sourced from both the Ensembl genome browser (Supplementary Files 8 and 9) and UCSC (Supplementary Files 10 and 11). Upon comparing the results, it became apparent that the outcomes from the two reference gene lists exhibited slight disparities. This divergence can likely be attributed to variations in methodologies and potential data sources employed by the custodians of each database when generating their gene annotations. Consequently, the position and quantity of genes may vary between different gene builds, even when referencing the same reference genome.

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

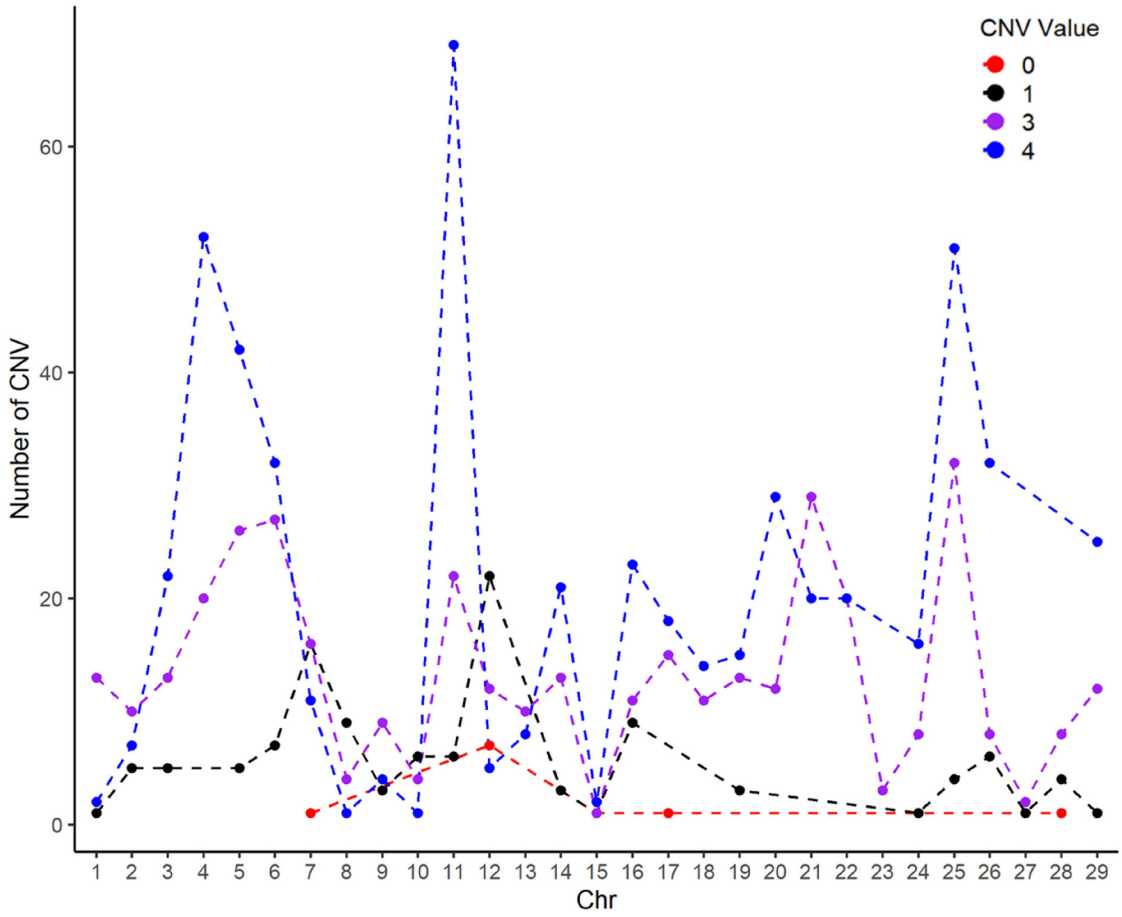

Figure 2. Line plot showing the distribution of CNVs across chromosomes, categorized by CNV types (0, 1, 3, 4). Each line represents the number of CNVs per chromosome, with distinct colours indicating different CNV values. The plot highlights variation in CNV counts across chromosomes and types.

Therefore, it is advisable to validate genes of interest by cross-referencing them in more than one database to ensure their reliability and robustness.

After comparing the annotation results from two different approaches, the list of consensus genes was obtained (Supplementary Files 12 and 13). The set of consensus genes common to both CNV results based on the common-gene-threshold criterion is 5%. A total of 97 genes were considered as the 'common high' that are present in both approaches more than 5% while the 637 genes were fell among the 'common low' that were present in both approaches but not crossing the threshold. The top 10 genes were searched in the cattle literature and found to be associated with relevant traits (Table 3). It can be observed here that the CNVs in our study are highly enriched with immune and defense genes, and the same findings can also be observed in other CNV studies in cattle populations (Liu and Bickhart, 2012; Goyache *et al.*, 2022; Jang *et al.*, 2021; Braga *et al.*, 2023).

For confirmation and validation, Ensemble Biomart (Ensemble Genes 110) was also used for the gene annotation. A total of 268 CNVRs containing 249 annotated genes, which can be classified further. Among annotated genes, 233 were protein coding, 2 as pseudogenes, 6 as microRNA, 3 as snRNA and rRNA ($n = 1$). While annotating these genes with GO terms, biological process components revealed that genes under CNVRs have reported functions related to immune response, production, reproduction, growth, heat stress and more. Many well-defined contrasting traits between indicine and taurine cattle, subject to natural and artificial selection for production, are governed by genes participating in diverse biological processes. These processes encompass reproduction, such as fertility, age of first oestrous, calving interval, (Sartori *et al.*, 2011), resilience against ecto- and endo-parasites (Piper *et al.*, 2009), adaptation to high temperatures (Beatty *et al.*, 2006), immunity to diseases (Brunelle *et al.*, 2008), as well as traits related to growth, carcass and meat quality (Bolormaa *et al.*, 2013).

Comparison of CNVR with previous studies

Upon comparison, only a limited number of overlapping CNVRs were observed between the CNVRs identified in this study and the DGVa CNVR set. Eleven overlapping CNVRs are identified. Five CNVRs from this study were overlapped with the study conducted by Hou *et al.* (2011), followed by three CNVRs overlapped from the studies of Mesbah-Uddin *et al.* (2018), Keel *et al.* (2017), Bickhart *et al.* (2012) and Karimi *et al.* (2017). Two overlapping CNVRs were observed with Liu *et al.* (2010), and only one overlapped CNVR with Boussaha *et al.* (2015) was observed.

Discussion

CNVs are key contributors to genomic structural variation, affecting gene function through changes in gene structure, dosage and regulation, with a larger impact on phenotypes than SNPs

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

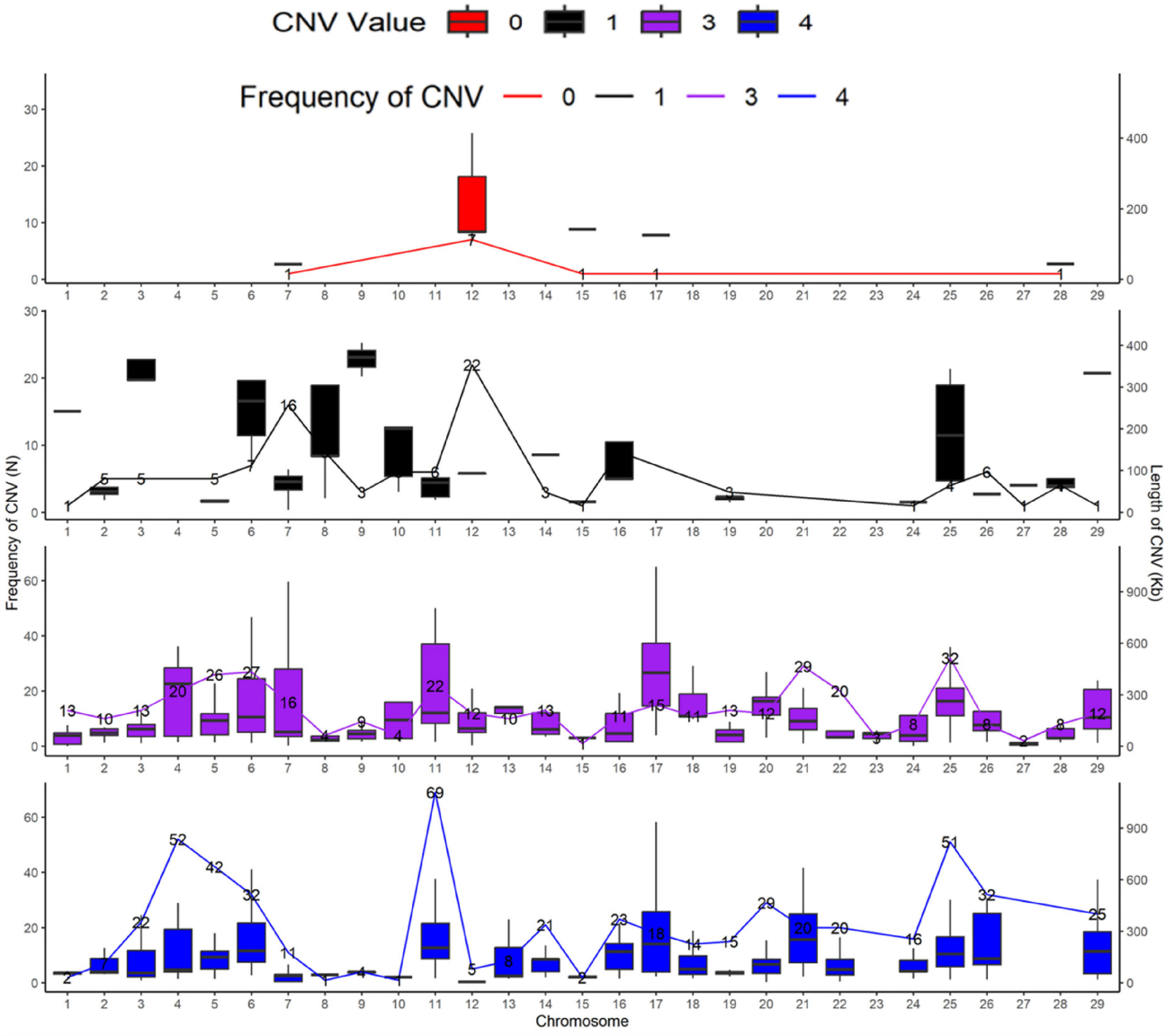

Figure 3. Prevalence and type of CNVs across autosomes for four types (0, 1, 3, 4) represented as boxplots (lengths in kb) and lines with points (frequency per chromosome). Boxplots show medians, interquartile ranges and whiskers, while numbers on points indicate CNV counts.

(Dang *et al.*, 2024; Xu *et al.*, 2014b; Zhang *et al.*, 2009). In livestock, CNVs influence economically important traits and disease conditions, making them valuable molecular markers (Cheng *et al.*, 2020; Liu *et al.*, 2011). Recent studies have extensively explored CNV diversity in both Bos taurus, Bos indicus and their hybrids (Benfica *et al.*, 2024a; Benfica *et al.*, 2024b; Cai *et al.*, 2024; Dang *et al.*, 2024; Delledonne *et al.*, 2024; Du *et al.*, 2024; Liu *et al.*, 2024; Maezawa *et al.*, 2024; Oliveira *et al.*, 2024; Wang *et al.*, 2024).

The present study aimed to generate a genome-wide CNV map of crossbred dairy cattle in Pakistan. Our results revealed widespread CNVRs, with 1055 CNVs and 268 CNVRs detected using the PennCNV software (Wang *et al.*, 2007). PennCNV was chosen for its ability to utilize all available information for each SNP, including the LRR, BAF, PFB and the distance between neighbouring SNPs. Dang *et al.* (2024) detected 16,507 CNVs and 3,728 CNVRs, accounting for 0.61% of the reference genome in Yunling cattle and Benfica *et al.* (2024b) found 3,161 CNVs and 561 CNVRs covering 3.99% of the Nellore autosomal genome. A similar study on Nellore cattle also indicated 14,914 CNVs and 1,884 CNVRs (Benfica *et al.*, 2024a). 870 CNVRs were reported in Holstein cattle (Oliveira *et al.*, 2024), 755 CNVRs, accounting for approximately 3.24% of the genome in Pingliang Red Cattle (Wang *et al.*, 2024). Similarly the Delledonne *et al.* (2024) reported 123,814 CNVs and 1,397 CNVRs in Holstein cattle.

The observed CNV count per animal is 3–37, with an average of 13.88 CNVs. There are variations in these values in the literature, ranging from 13 to 51 with an average of 32.5 (Delledonne *et al.*, 2024). The crossbred cattle exhibited a relatively high number of CNVs per individual compared to breed groups from other regions. Several factors may contribute to this observation, such as the inadequate representation of Sahiwal or indigenous cattle in the utilized SNP chips for bovines. This lower resolution could potentially introduce bias into the results, especially when compared to studies that did not include indicine cattle in their analysis. Differences in the abundance of CNVs across diverse cattle populations have been previously noted. Specifically,

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

Table 2. No. of CNVRs, proportional length of CNVRs on each autosome using HandyCNV and CNVRuler

		Using HandyCNV R package			Using CNVRuler software		
Chr	Length of chromosome (bp)	Number of CNVR	Total length	Percentages	No of CNVRs	Length of CNVR (bp)	Percentage (%)
1	158,534,110	11	825674	0.5208	11	825663	0.5208
2	136,231,102	12	1049179	0.7701	12	942570	0.6918
3	121,005,158	15	2199660	1.8178	16	1464406	1.2102
4	120,000,601	7	957035	0.7975	7	754556	0.6287
5	120,089,316	15	2703362	2.2511	16	1622408	1.3510
6	117,806,340	12	2955760	2.5089	13	2150648	1.8255
7	110,682,743	16	2829493	2.5563	20	1834475	1.6574
8	113,319,770	9	807918	0.7129	9	807909	0.7129
9	105,454,467	10	1383083	1.3115	11	1325965	1.2573
10	103,308,737	5	740258	0.7165	5	740253	0.7165
11	106,982,474	12	2763340	2.5829	13	1857848	1.7365
12	87,216,183	7	3101114	3.5556	8	1247488	1.4303
13	83,472,345	7	976205	1.1694	7	751223	0.8999
14	82,403,003	11	1130045	1.3713	11	992346	1.2042
15	85,007,780	5	282879	0.3327	5	282874	0.3327
16	81,013,979	10	1687131	2.0825	10	1039649	1.2832
17	73,167,244	4	1760103	2.4055	4	826996	1.1302
18	65,820,629	8	1568565	2.3830	8	1426884	2.1678
19	63,449,741	20	1505495	2.3727	20	1505475	2.3727
20	71,974,595	11	1433422	1.9915	11	972810	1.3516
21	69,862,954	6	1593890	2.2814	7	946613	1.3549
22	60,773,035	7	1311229	2.1575	7	1079395	1.7761
23	52,498,615	2	133721	0.2547	0		0
24	62,317,253	8	912841	1.4648	0		0
25	42,350,435	11	2601858	6.1436	11	2002098	4.7274
26	51,992,305	8	901651	1.7342	8	814694	1.5669
27	45,612,108	3	94299	0.2067	3	94296	0.2067
28	45,940,150	7	1323741	2.8814	7	1323734	2.8814
29	51,098,607	10	2341046	4.5814	10	1399748	2.7393
	2,489,385,779	269	43873997	1.7624	270	31033024	1.2466

CNVs, copy number variations; CNVRs, copy number variation regions.

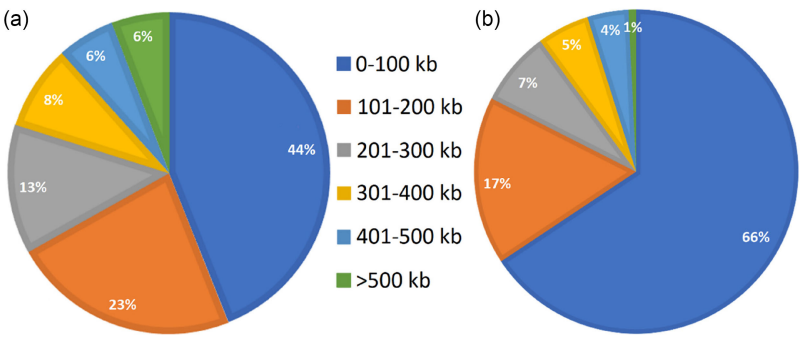

Figure 4. Distribution of CNVs (a) and CNVRs (b) in different distance categories (0–100 kb, 101–200 kb, 201–300 kb, 301–400 kb, 401–500 kb and >500 kb) across autosomes. (a) The majority of CNVs (44%) fall within the 0–100 kb range. (b) Detailed breakdown of CNVR size ranges, highlighting that 65.67% of the CNVRs are within the 0–100 kb category, followed by 16.79% in the 101–200 kb range.

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

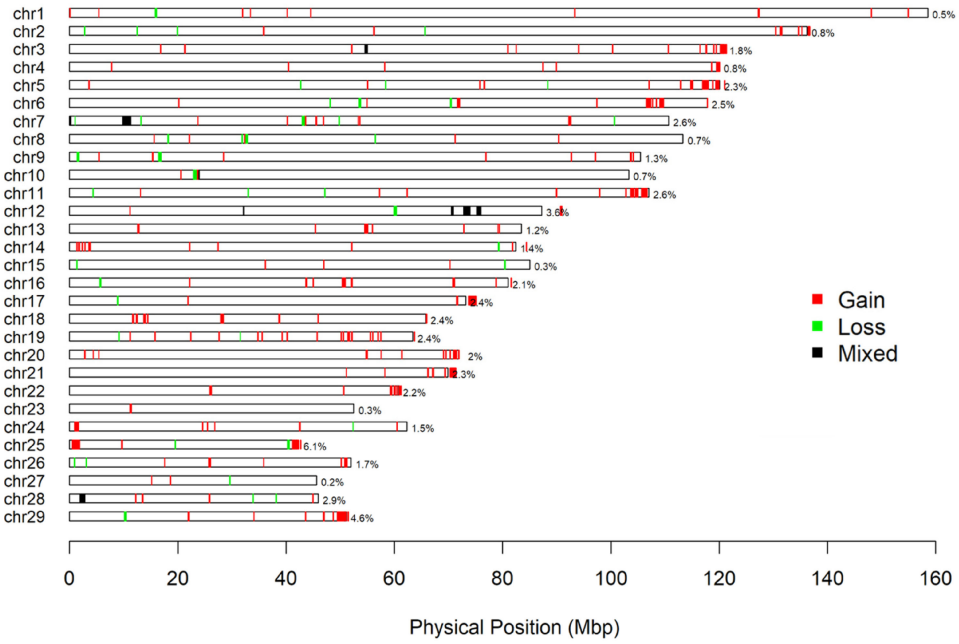

Figure 5. Genome-wide CNVR map illustrating the distribution of CNVRs across autosomes. Each horizontal bar represents a chromosome, with CNVRs categorized as gains (red), losses (green) and mixed events (black) based on their type. The percentages indicate the proportion of the chromosomes covered by CNVRs. The physical positions of CNVRs are displayed along the x-axis in megabase pairs (Mbp).

Table 3. Top 10 highly common genes and their association with economic traits

Gene	Reported functions
NOXA1	Role in hypoxic adaptation (Zhang *et al.*, 2020)
VIPR2	Candidate genes affecting fat percentage with an important role in milk synthesis (Zhang *et al.*, 2018a; Capomaccio *et al.*, 2015).
TUBB4B	encodes different sub-families of tubulin (Laskowski et al., 2017)
PDGFA	Increased expression in bovine tuberculosis (Meade *et al.*, 2007)
ARSA	Involved in different reproductive traits (Forde *et al.*, 2013)
MAPK8IP2	Involved in different reproductive traits (Fayad *et al.*, 2007)
NDOR1	Suggested to play a role in the bioreduction of anti-cancer drugs in humans (Paine *et al.*, 2000; Froese *et al.*, 2008)
CHKB	Growth traits (Goshu *et al.*, 2018)
CPT1B	Involved in lipid metabolism and Bovine Respiratory Disease Susceptibility (He *et al.*, 2022)
ODF3B	Lymphoblastoid cells (Ryu *et al.*, 2014)
SCO2	Downregulation of this gene is associated with fat gain and increased insulin resistance (Hill *et al.*, 2017; Gershoni *et al.*, 2021).

indicine and African taurine breeds exhibit a higher CNV abundance compared to European taurine breeds, a characteristic attributed to their breed divergence and population history (Liu *et al.*, 2011). These findings highlight the impact of factors such as changes in historical effective population size, gene flow and selection processes on the varying CNV abundance observed in distinct populations.

Thus, it is reasonable to posit that the sustained small effective population size over numerous generations in this group may have prompted a relaxation of purifying selection against mildly deleterious CNVs. Consequently, such relaxation could contribute to the accumulation of a substantial number of CNV events. This aligns with findings from Upadhyay *et al.* (2017) suggesting that genetically isolated small populations may accumulate an abundance of CNVs. However, it is noteworthy that in different studies, deletions were primarily observed (Upadhyay *et al.*, 2017, Oliveira *et al.*, 2024, Tao *et al.*, 2007), whereas the current study predominantly identified gain events. Nevertheless, it is essential to acknowledge that the present study is limited by a low sample size, and larger samples from other indigenous breeds are necessary to further explore this hypothesis.

CNVRs were generated using the two available in silico molecular techniques, that is, CNVRuler and HandyCNV package of R. Some differentiating points were observed within the two (Table 2). The contributing reason to this may be the type of algorithms used for the detection, as well as the technology. These methodologies vary in coverage range and their capabilities to identify and pinpoint CNV breakpoints (Zhan *et al.*, 2011). The functional analysis of the regions encompassed by CNVRs unveiled genes linked to complex traits.

In our analysis of CNVs within the dataset, we observed a notable prevalence of gain events. Here, it is essential to consider the influence of biological variation. Gain events can naturally occur more frequently than loss events in certain genomic regions or within specific populations due to inherent biological diversity. In some instances, these gain events might offer a selective advantage, thus driving their increased occurrence.

Furthermore, genomic regions that undergo duplication or gain events may contain genes or sequences that confer advantageous traits, such as enhanced disease resistance or improved adaptability. This phenomenon could be attributed to positive selective pressure acting on these regions, thereby leading to a higher frequency of gain events.

Annotations were performed separately using information sourced from both Ensembl genome browser (Birney *et al.*, 2004) and UCSC (Karolchik *et al.*, 2003) using the HandyCNV package of R. Upon comparison of the gene list from both genome browsers, 97 common-high and 637 common-low genes were obtained. The top 10 major genes are found to be involved in

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

different economically important traits like milk production, growth traits, adaptation, disease resistance and immunity (Table 3). This may explain the increased production, heat tolerance and disease resilience abilities of crossbreds, which is the underlying reason for their production. Gene enrichment and QTLs play crucial roles in major functional regions of the genome. Gene Ontology analyses for the detected CNVRs revealed enrichment in important GO terms, highlighting some relevant traits. For instance, GO:0030879 is involved in mammary gland development, directly influencing the milk production of the animals. Similarly, GO:0051879 is mainly linked with heat shock proteins, playing a crucial role in multiple types of stresses, including heat stress. Another intriguing term is GO:0071456, related to hypoxic adaptation, suggesting that these animals can effectively adapt to environments with limited oxygen supply, such as high elevations or hypoxic conditions (Table 3).

FISH and quantitative Polymerase Chain Reaction are well-acknowledged methods for the confirmation and validation of CNVs, offering high specificity and accuracy (Bickhart *et al.*, 2012). However, these analyses are recognized for being expensive, time-demanding and consuming a substantial amount of biological material. Therefore, this study opted for an in silico method to identify CNVRs while minimizing the reliance on extensive laboratory resources (Bickhart *et al.*, 2012).

The CNVs and CNVRs discovered in this study lay the groundwork for future research on CNVs in other Pakistani cattle breeds and Zebu cattle worldwide. Subsequent investigations should explore the impact of incorporating CNV information in genomic selection for crossbred dairy cattle in Pakistan. Furthermore, it is highly recommended to conduct CNV-based Genome-Wide Association Studies focusing on critical traits in these cattle. This holistic approach will contribute valuable insights to the field of cattle genomics and enhance our understanding of the genetic basis of important traits in diverse cattle populations.

Conclusion

This study marks the inaugural genome-wide detection of CNVs in crossbred dairy cattle in Pakistan. The genes identified within these CNV regions illuminate potential biological processes that may underlie indigenous cattle's adaptability and disease resistance. QTL analyses revealed significant overlaps between many CNVRs and QTLs associated with economically important traits in cattle, including lactation, fertility, stimulus recognition and health. These findings present viable candidates for further validation in the population. Given the preliminary nature of this report, it is strongly recommended that high-density SNP arrays, whole-genome sequencing, or resequencing data from key indigenous cattle breeds with larger sample sizes be collected and utilized to construct a comprehensive genome-wide map of CNVs in indigenous cattle.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0021859625100208

Acknowledgements. We acknowledge the Higher Education Commission, Pakistan for providing funding to Fakhar un Nisa for her PhD Studies.

Author contributions. Conceptualization: Fakhar un Nisa, Qamar un Nisa. Data curation: Fakhar un Nisa, Muhammad Asif. Formal analysis: Fakhar un Nisa, Rubab Zahra Naqvi. Investigation: Fakhar un Nisa, Muhammad Saif Ur Rehman. Methodology: Fakhar un Nisa. Project administration: Muhammad Asif, Zahid Mukhtar. Resources: Muhammad Asif. Software: Fakhar un Nisa. Supervision: Zahid Mukhtar. Visualization: Fakhar un Nisa. Writing – original draft: Fakhar un Nisa. Writing – review and editing: Zahid Mukhtar, Muhammad Saif Ur Rehman.

Funding statement. This study is funded by the PAKISTAN AGRICULTURAL RESEARCH COUNCIL, AGRICULTURAL LINKAGES PROGRAMME (ALP) with Project Identification No. AS 016 titled 'Development and application of genomic selection in foreign and local cattle breeds for improvement in dairy-related traits'.

Competing interests. The authors declare that they have no competing interests.

Ethical standards. To guarantee the ethical and humane treatment of animals, the investigation outlined in this research paper received approval from the Research Ethics Committee of the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan on 10 June 2020. A qualified veterinarian supervised the blood collection process to minimize distress and harm to the animals. Prior to collecting any samples, the researchers conducted a meeting with the farm owners where the animals were housed. During this meeting, they explained the study's purpose and obtained verbal informed consent.

References

Ahmad SF, Chandrababu Shailaja C, Vaishnav S, Kumar A, Gaur GK, Janga SC, Ahmad SM, Malla WA and Dutt T (2023) Read-depth based approach on whole genome resequencing data reveals important insights into the copy number variation (CNV) map of major global buffalo breeds. *BMC Genomics* **24**, 616.

Almal SH and Padh H (2012) Implications of gene copy-number variation in health and diseases. *Journal of Human Genetics* **57**, 6–13.

Beatty D, Barnes A, Taylor E, Pethick D, McCarthy M and Maloney S (2006) Physiological responses of Bos taurus and Bos indicus cattle to prolonged, continuous heat and humidity. *Journal of Animal Science* **84**, 972–985.

Bebe BO, Udo HM, Rowlands GJ and Thorpe W (2003) Smallholder dairy systems in the Kenya highlands: breed preferences and breeding practices. *Livestock Production Science* **82**, 117–127.

Benfica LF, Brito LF, Do Bem RD, De Oliveira LF, Mulim HA, Braga LG, Cyrillo JN, Bonilha SF and Mercadante MEZ (2024a) Detection and characterization of copy number variation in three differentially-selected Nellore cattle populations. *Frontiers in Genetics* **15**, 1377130.

Benfica LF, Brito LF, Do Bem RD, Mulim HA, Glessner J, Braga LG, Gloria LS, Cyrillo JN, Bonilha SF and Mercadante ME (2024b) Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nellore cattle. *BMC Genomics* **25**, 54.

Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M and Taylor JF (2012) Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Research* **22**, 778–790.

Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V and Cutts T (2004) An overview of Ensembl. *Genome Research* **14**, 925–928.

Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, Tier B, Barendse W, Reverter A and Goddard ME (2013) Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. *Genetics Selection Evolution* **45**, 1–12.

Boussaha M, Esquerré D, Barbieri J, Djari A, Pinton A, Letaief R, Salin G, Escudié F, Roulet A and Fritz S (2015) Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PLoS One* **10**, e0135931.

Braga LG, Chud TC, Watanabe RN, Savegnago RP, Sena TM, Do Carmo AS, Machado MA, Panetto JCDC, Da Silva MVG and Munari DP (2023) Identification of copy number variations in the genome of Dairy Gir cattle. *PLoS One* **18**, e0284085.

Brunelle BW, Greenlee JJ, Seabury CM, Brown CE and Nicholson EM (2008) Frequencies of polymorphisms associated with BSE resistance differ

significantly between Bos taurus, Bos indicus, and composite cattle. *BMC Veterinary Research* **4**, 1–8.

Butty AM, Chud TC, Cardoso DF, Lopes LS, Miglior F, Schenkel FS, Cánovas A, Häfliger IM, Drögemüller C and Stothard P (2021) Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. *Journal of Dairy Science* **104**, 8050–8061.

Butty AM, Chud TC, Miglior F, Schenkel FS, Kommadath A, Krivushin K, Grant JR, Häfliger IM, Drögemüller C and Cánovas A (2020) High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. *Scientific Reports* **10**, 8044.

Cai H, Li X, Niu X, Li J, Lan X, Lei C, Huang Y, Xu H, Li M and Chen H (2024) Copy number variations within fibroblast growth factor 13 gene influence growth traits and alternative splicing in cattle. *Animal Biotechnology* **35**, 2314104.

Capomaccio S, Milanesi M, Bomba L, Cappelli K, Nicolazzi EL, Williams JL, Ajmone-Marsan P and Stefanon B (2015) Searching new signals for production traits through gene-based association analysis in three Italian cattle breeds. *Animal Genetics* **46**, 361–370.

Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, Anney R, Shields DC, Abrahams BS, Almeida J and Bacchelli E (2012) A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Human Genetics* **131**, 565–579.

Chen Y, Khan MZ, Wang X, Liang H, Ren W, Kou X, Liu X, Chen W, Peng Y and Wang C (2024) Structural variations in livestock genomes and their associations with phenotypic traits: a review. *Frontiers in Veterinary Science* **11**, 1416220.

Cheng J, Jiang R, Yang Y, Cao X, Huang Y, Lan X, Lei C, Hu L and Chen H (2020) Association analysis of KMT2D copy number variation as a positional candidate for growth traits. *Gene* **753**, 144799.

Cicconardi F, Chillemi G, Tramontano A, Marchitelli C, Valentini A, Ajmone-Marsan P and Nardone A (2013) Massive screening of copy number population-scale variation in Bos taurus genome. *BMC Genomics* **14**, 1–15.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C and Campbell P (2010) Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R and Hannig V (2011) A copy number variation morbidity map of developmental delay. *Nature Genetics* **43**, 838–846.

Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, Chin S-F, Brenton JD, Tavaré S and Caldas C (2009) The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* **10**, 1–23.

Da Silva JM, Giachetto PF, Da Silva LO, Cintra LC, Paiva SR, Yamagishi MEB and Caetano AR (2016) Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics* **17**, 1–14.

Dang D, Zhang L, Gao L, Peng L, Chen J and Yang L (2024) Analysis of genomic copy number variations through whole-genome scan in Yunling cattle. *Frontiers in Veterinary Science* **11**, 1413504.

Dash A, Sivalingam J, Bidyalaxmi K, Sukhija N, Kumar R, Niranjan SK, Tantia MS and Gupta ID (2023) Stepwise detection of copy number variations in whole genome sequence of buffalo. *Indian Journal of Veterinary Sciences & Biotechnology* **19**, 30–33.

De Almeida Santana MH, Junior GAO, Cesar ASM, Freua MC, Da Costa Gomes R, Da Luz e Silva S, Leme PR, Fukumasu H, Carvalho ME and Ventura RV (2016) Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *Journal of Applied Genetics* **57**, 495–504.

De Lemos MVA, Peripolli E, Berton MP, Feitosa FLB, Olivieri BF, Stafuzza NB, Tonussi RL, Kluska S, Chiaia HLJ and Mueller L (2018) Association study between copy number variation and beef fatty acid profile of Nellore cattle. *Journal of Applied Genetics* **59**, 203–223.

Delledonne A, Punturiero C, Ferrari C, Bernini F, Milanesi R, Bagnato A and Strillacci MG (2024) Copy number variant scan in more than four thousand Holstein cows bred in Lombardy, Italy. *PLoS One* **19**, e0303044.

Di Gerlando R, Sutera AM, Mastrangelo S, Tolone M, Portolano B, Sottile G, Bagnato A, Strillacci MG and Sardina MT (2019) Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS One* **14**, e0215204.

Du L, Ma W, Peng W, Zhao H, Zhao J, Wang J, Wang W, Lyu S, Zhang Z and Qi X (2024) Impact of STAT5A-CNVs on growth traits in Chinese beef cattle breeds. *Gene* **896**, 148073.

Durán Aguilar M, Román Ponce S, Ruiz López F, González Padilla E, Vásquez Peláez C, Bagnato A and Strillacci MG (2017) Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics* **134**, 49–59.

El-Sayed Moustafa JS, Eleftherohorinou H, De Smith AJ, Andersson-Assarsson JC, Couto Alves A, Hadjigeorgiou E, Walters RG, Asher JE, Bottolo L and Buxton JL (2012) Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity. *Human Molecular Genetics* **21**, 3727–3738.

Emerson J, Cardoso-Moreira M, Borevitz JO and Long M (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320**, 1629–1631.

Fadista J, Nygaard M, Holm L-E, Thomsen B and Bendixen C (2008) A snapshot of CNVs in the pig genome. *PLoS One* **3**, e3916.

Fayad T, Lefebvre R, Nimpf J, Silversides DW and Lussier JG (2007) Low-density lipoprotein receptor-related protein 8 (LRP8) is upregulated in granulosa cells of bovine dominant follicle: molecular characterization and spatio-temporal expression studies. *Biology of Reproduction* **76**, 466–475.

Fontanesi L, Beretti F, Martelli P, Colombo M, Dall'Olio S, Occidente M, Portolano B, Casadio R, Matassino D and Russo V (2011) A first comparative map of copy number variations in the sheep genome. *Genomics* **97**, 158–165.

Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, Casadio R, Russo V and Portolano B (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* **11**, 1–15.

Forde N, Mehta JP, McGettigan PA, Mamo S, Bazer FW, Spencer TE and Lonergan P (2013) Alterations in expression of endometrial genes coding for proteins secreted into the uterine lumen during conceptus elongation in cattle. *BMC Genomics* **14**, 1–13.

Froese D, Wu X, Zhang J, Dumas R, Schoel W, Amrein M and Gravel R (2008) Restricted role for methionine synthase reductase defined by subcellular localization. *Molecular Genetics and Metabolism* **94**, 68–77.

Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S, Zhang Q and Sun D (2017) CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC Genomics* **18**, 1–12.

Geistlinger L, Da Silva VH, Cesar ASM, Tizioto PC, Waldron L, Zimmer R, Regitano LCDA and Coutinho LL (2018) Widespread modulation of gene expression by copy number variation in skeletal muscle. *Scientific Reports* **8**, 1399.

Gershoni M, Weller JI and Ezra E (2021) Genetic and genome-wide association analysis of yearling weight gain in Israel Holstein dairy calves. *Genes* **12**, 708.

Goshu HA, Xiao Yun W, Chu M, Pengjia B and Yan P (2018) Population genetic copy number variation of CHKB, KLF6, GPC1 and CHRM3 genes in Chinese domestic yak (*Bos grunniens*) breeds. *Cogent Biology* **4**, 1471779.

Goyache F, Pérez-Pardal L, Fernández I, Traoré A, Menéndez-Arias NA, Arias KD and Álvarez I (2022) Identification and characterization of copy number variations regions in West African Taurine cattle. *Animals* **12**, 2130.

Gupta A, Place M, Goldstein S, Sarkar D, Zhou S, Potamousis K, Kim J, Flanagan C, Li Y and Newton MA (2015) Single-molecule analysis reveals widespread structural variation in multiple myeloma. *Proceedings of the National Academy of Sciences* **112**, 7689–7694.

He W, Gao M, Yang R, Zhao Z, Mi J, Sun H, Xiao H and Fang X (2022) The effect of CPT1B gene on lipid metabolism and its polymorphism analysis in Chinese Simmental cattle. *Animal Biotechnology* **33**, 1428–1440.

Hill S, Deepa SS, Sataranatarajan K, Premkumar P, Pulliam D, Liu Y, Soto VY, Fischer KE and Van Remmen H (2017) Sco2 deficient mice develop increased adiposity and insulin resistance. *Molecular and Cellular Endocrinology* **455**, 103–114.

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

Hou Y, Bickhart DM, Chung H, Hutchison JL, Norman HD, Connor EE and Liu GE (2012a) Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. *Functional & Integrative Genomics* **12**, 717–723.

Hou Y, Bickhart DM, Hvinden ML, Li C, Song J, Boichard DA, Fritz S, Eggen A, Denise S, Wiggans GR and Liu GE (2012b) Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genomics* **13**, 1–10.

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim E-S, Matukumalli LK, Ventura M, Song J and VanRaden PM (2011) Genomic characteristics of cattle copy number variations. *BMC Genomics* **12**, 1–11.

Hou Y, Liu GE, Bickhart DM, Matukumalli LK, Li C, Song J, Gasbarre LC, Van Tassell CP and Sonstegard TS (2012c) Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Functional & Integrative Genomics* **12**, 81–92.

Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, Li R, Zhang S, Chen N and Asadollahpour Nanaei H (2021) An atlas of CNV maps in cattle, goat and sheep. *Science China Life Sciences* **64**, 1747–1764.

Ionita-Laza I, Rogers AJ, Lange C, Raby BA and Lee C (2009) Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* **93**, 22–26.

Iqbal N, Liu X, Yang T, Huang Z, Hanif Q, Asif M, Khan QM and Mansoor S (2019) Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. *PLoS One* **14**, e0215065.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, Szpiech ZA, Degnan JH, Wang K and Guerreiro R (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003.

Jang J, Terefe E, Kim K, Lee YH, Belay G, Tijjani A, Han JL, Hanotte O and Kim H (2021) Population differentiated copy number variation of Bos taurus, Bos indicus and their African hybrids. *BMC Genomics* **22**, 1–11.

Jiang L, Jiang J, Wang J, Ding X, Liu J and Zhang Q (2012) Genome-wide identification of copy number variations in Chinese Holstein. *PLoS One* **7**, e48732.

Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, Ding X, Liu J and Zhang Q (2013) Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics* **14**, 1–10.

Kang X, Li M, Liu M, Liu S, Pan MG, Wiggans GR, Rosen BD and Liu GE (2020) Copy number variation analysis reveals variants associated with milk production traits in dairy goats. *Genomics* **112**, 4934–4937.

Karimi K, Esmailizadeh A, Wu D and Gondro C (2017) Mapping of genome-wide copy number variations in the Iranian indigenous cattle using a dense SNP data set. *Animal Production Science* **58**, 1192–1200.

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu Y, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D and Kent WJ (2003) The UCSC genome browser database. *Nucleic Acids Research* **31**, 51–54.

Keel BN, Keele JW and Snelling WM (2017) Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. *Animal Genetics* **48**, 141–150.

Kim J-H, Hu H-J, Yim S-H, Bae J-S, Kim S-Y and Chung Y-J (2012a) CNVRuler: a copy number variation-based case–control association analysis tool. *Bioinformatics* **28**, 1790–1792.

Kim K, Kim S, Raney N and Ernst C (2012b) Evaluation of BTA1 and BTA5 QTL regions for growth and carcass traits in American and Korean cattle. *Asian-Australasian Journal of Animal Sciences* **25**, 1521.

Kumar H, Panigrahi M, Strillacci MG, Sonejita Nayak S, Rajawat D, Ghildiyal K, Bhushan B and Dutt T (2023) Detection of genome-wide copy number variation in Murrah buffaloes. *Animal Biotechnology* **34**, 3783–3795.

Kumar S, Alex R, Gaur G, Mukherjee S, Mandal D, Singh U, Tyagi S, Kumar A, Das A and Deb R (2018) Evolution of Frieswal cattle: a crossbred dairy animal of India. *Indian Journal of Animal Sciences* **88**, 265–275.

Laskowski, D, Båge, R, Humblot, P, Andersson, G, Sirard, M-A and Sjunnesson, Y (2017) Insulin during in vitro oocyte maturation has an impact on development, mitochondria, and cytoskeleton in bovine day 8 blastocysts. *Theriogenology*, **101**, 15–25.

Leroy G, Baumung R, Boettcher P, Scherf B and Hoffmann I (2016) Sustainability of crossbreeding in developing countries; definitely not like crossing a meadow . . . . *Animal* **10**, 262–273.

Letaief R, Rebours E, Grohs C, Meersseman C, Fritz S, Trouilh L, Esquerré D, Barbieri J, Klopp C and Philippe R (2017) Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. *Genetics Selection Evolution* **49**, 1–15.

Liu GE and Bickhart DM (2012) Copy number variation in the cattle genome. *Functional & Integrative Genomics* **12**, 609–624.

Liu GE, Brown T, Hebert DA, Cardone MF, Hou Y, Choudhary RK, Shaffer J, Amazu C, Connor EE, Ventura M, Gasbarre LC, Van Tassell CP and Sonstegard TS (2011) Initial analysis of copy number variations in cattle selected for resistance or susceptibility to intestinal nematodes. *Mammalian Genome* **22**, 111–121.

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra GM, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE and Bickhart DM (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Research* **20**, 693–703.

Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J and Eichler EE (2009) Analysis of recent segmental duplications in the bovine genome. *BMC Genomics* **10**, 1–16.

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C and Zhao F (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* **14**, 1–11.

Liu S, Kang X, Catacchio CR, Liu M, Fang L, Schroeder SG, Li W, Rosen BD, Iamartino D and Iannuzzi L (2019) Computational detection and experimental validation of segmental duplications and associated copy number variations in water buffalo (Bubalus bubalis). *Functional & Integrative Genomics* **19**, 409–419.

Liu X, Chen W, Huang B, Wang X, Peng Y, Zhang X, Chai W, Khan MZ and Wang C (2024) Advancements in copy number variation screening in herbivorous livestock genomes and their association with phenotypic traits. *Frontiers in Veterinary Science* **10**, 1334434.

Maezawa M, Watanabe K-I, Kobayashi Y, Yoshida K, Chambers JK, Uchida K, Maruyama R and Inokuma H (2024) Diffuse large B-cell lymphoma with DNA copy number changes in a Japanese black calf. *Veterinary Research Communications* **48**, 2651–2656.

Malek S (2013) The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene* **32**, 2805–2817.

Mbole-Kariuki MN, Sonstegard T, Orth A, Thumbi S, Bronsvoort BDC, Kiara H, Toye P, Conradie I, Jennings A and Coetzer K (2014) Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* **113**, 297–305.

Meade KG, Gormley E, Doyle MB, Fitzsimons T, O'Farrelly C, Costello E, Keane J, Zhao Y and MacHugh DE (2007) Innate gene repression associated with *Mycobacterium bovis* infection in cattle: toward a gene signature of disease. *BMC Genomics* **8**, 1–15.

Mesbah-Uddin M, Guldbrandtsen B, Iso-Touru T, Vilkki J, De Koning D-J, Boichard D, Lund MS and Sahana G (2018) Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. *DNA Research* **25**, 49–59.

Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND and Lee CM (2021) The UCSC genome browser database: 2021 update. *Nucleic Acids Research* **49**, D1046–D1057.

Nisa FU, Kaul H, Asif M, Amin I, Mrode R, Mansoor S and Mukhtar Z (2023) Genetic insights into crossbred dairy cattle of Pakistan: exploring allele frequency, linkage disequilibrium, and effective population size at a genome-wide scale. *Mammalian Genome* **34**, 602–614.

Nisa FU, Naqvi RZ, Arshad F, Ilyas I, Asif M, Amin I, Mrode R, Mansoor S and Mukhtar Z (2024) Assessment of genomic diversity and selective pressures in crossbred dairy cattle of Pakistan. *Biochemical Genetics* **62**, 4137–4156.

Oliveira HR, Chud TC, Oliveira Jr GA, Hermisdorff IC, Narayana SG, Rochus CM, Butty AM, Malchiodi F, Stothard P and Miglior F (2024) Genome-wide association analyses reveals copy number variant regions

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press

associated with reproduction and disease traits in Canadian Holstein cattle. *Journal of Dairy Science* **107**, 7052–7063.

Ouyang L, Lee J, Park C-K, Mao M, Shi Y, Gong Z, Zheng H, Li Y, Zhao Y and Wang G (2014) Whole-genome sequencing of matched primary and metastatic hepatocellular carcinomas. *BMC Medical Genomics* **7**, 1–13.

Paine MJ, Garner AP, Powell D, Sibbald J, Sales M, Pratt N, Smith T, Tew DG and Wolf CR (2000) Cloning and characterization of a novel human dual flavin reductase. *Journal of Biological Chemistry* **275**, 1471–1478.

Park RW, Kim T-M, Kasif S and Park PJ (2015) Identification of rare germline copy number variations over-represented in five human cancer types. *Molecular Cancer* **14**, 1–12.

Pierce MD, Dzama K and Muchadeyi FC (2018) Genetic diversity of seven cattle breeds inferred using copy number variations. *Frontiers in Genetics* **9**, 163.

Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR and Mills R (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology* **29**, 512–520.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C and Abrahams BS (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372.

Piper EK, Jonsson NN, Gondro C, Lew-Tabor AE, Moolhuijzen P, Vance ME and Jackson LA (2009) Immunological profiles of Bos taurus and Bos indicus cattle infested with the cattle tick, *Rhipicephalus (Boophilus) microplus*. *Clinical and Vaccine Immunology* **16**, 1074–1086.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI and Daly MJ (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575.

Ryu J, Woo J, Shin J, Ryoo H, Kim Y and Lee C (2014) Profile of differential promoter activity by nucleotide substitution at GWAS signals for multiple sclerosis. *Medicine* **93**, e281.

Sartori R, Bastos M, Baruselli PS, Gimenes LU, Ereno RL and Barros C (2011) Physiological differences and implications to reproductive management of Bos taurus and Bos indicus cattle in a tropical environment. *Reproduction in Domestic Ruminants* **7**, 357.

Sasaki S, Ibi T, Akiyama T, Fukushima M and Sugimoto Y (2016) Loss of maternal ANNEXIN A10 via a 34-kb deleted-type copy number variation is associated with embryonic mortality in Japanese Black cattle. *BMC Genomics* **17**, 1–15.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Manér S, Massa H, Walker M and Chi M (2004) Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528.

Strillacci MG, Moradi-Shahrbabak H, Davoudi P, Ghoreishifar SM, Mokhber M, Masroure AJ and Bagnato A (2021) A genome-wide scan of copy number variants in three Iranian indigenous river buffaloes. *BMC Genomics* **22**, 1–14.

Tao S, Fan Y, Wang W, Ma G, Liang L and Shi Q (2007) Patterns of insertion and deletion in mammalian genomes. *Current Genomics* **8**, 370–378.

Upadhyay M, Da Silva VH, Megens H-J, Visker MH, Ajmone-Marsan P, Bălteanu VA, Dunner S, Garcia JF, Ginja C and Kantanen J (2017) Distribution and functionality of copy number variation across European cattle populations. *Frontiers in Genetics* **8**, 108.

Verma M, Khoury MJ and Ioannidis JP (2013) Opportunities and challenges for selected emerging technologies in cancer epidemiology: mitochondrial, epigenomic, metabolomic, and telomerase profiling. *Cancer Epidemiology, Biomarkers & Prevention* **22**, 189–200.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H and Bucan M (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* **17**, 1665–1674.

Wang Y, Ma J, Wang J, Zhang L, Xu L, Chen Y, Zhu B, Wang Z, Gao H and Li J (2024) Genome-wide detection of copy number variations and their potential association with carcass and meat quality traits in Pingliang Red cattle. *International Journal of Molecular Sciences* **25**, 5626.

Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP and Liu GE (2014a) Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics* **15**, 1–10.

Xu L, Hou Y, Bickhart DM, Zhou Y, Hay EHA, Song J, Sonstegard TS, Van Tassell CP and Liu GE (2016) Population-genetic properties of differentiated copy number variations in cattle. *Scientific Reports* **6**, 23161.

Xu Y, Shi T, Cai H, Zhou Y, Lan X, Zhang C, Lei C, Qi X and Chen H (2014b) Associations of MYH3 gene copy number variations with transcriptional expression and growth traits in Chinese cattle. *Gene* **535**, 106–111.

Yang L, Han J, Deng T, Li F, Han X, Xia H, Quan F, Hua G, Yang L and Zhou Y (2023) Comparative analyses of copy number variations between swamp buffaloes and river buffaloes. *Animal Genetics* **54**, 199–206.

Yue X-P, Dechow C, Chang T-C, DeJarnette JM, Marshall CE, Lei C-Z and Liu W-S (2014) Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. *BMC Genomics* **15**, 1–12.

Zarrei M, MacDonald JR, Merico D and Scherer SW (2015) A copy number variation map of the human genome. *Nature Reviews Genetics* **16**, 172–183.

Zhan B, Fadista J, Thomsen B, Hedegaard J, Panitz F and Bendixen C (2011) Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics* **12**, 1–20.

Zhang F, Gu W, Hurles ME and Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics* **10**, 451–481.

Zhang G-M, Zheng L, He H, Song C-C, Zhang Z-J, Cao X-K, Lei C-Z, Lan X-Y, Qi X-L and Chen H (2018a) Associations of GBP2 gene copy number variations with growth traits and transcriptional expression in Chinese cattle. *Gene* **647**, 101–106.

Zhang L, Jia S, Yang M, Xu Y, Li C, Sun J, Huang Y, Lan X, Lei C and Zhou Y (2014) Detection of copy number variations and their effects in Chinese bulls. *BMC Genomics* **15**, 1–9.

Zhang Q, Calus MP, Bosse M, Sahana G, Lund MS and Guldbrandtsen B (2018b) Human-mediated introgression of haplotypes in a modern dairy cattle breed. *Genetics* **209**, 1305–1317.

Zhang X, Chen N, Chen H, Lei C and Sun T (2022) Comparative analyses of copy number variations between swamp and river buffalo. *Gene* **830**, 146509.

Zhang Y, Hu Y, Wang X, Jiang Q, Zhao H, Wang J, Ju Z, Yang L, Gao Y and Wei X (2020) Population structure, and selection signatures underlying high-altitude adaptation inferred from genome-wide copy number variations in Chinese indigenous cattle. *Frontiers in Genetics* **10**, 1404.

Zhao M, Wang Q, Wang Q, Jia P and Zhao Z (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* **14**, 1–16.

Zhou J, Liu L, Lopdell TJ, Garrick DJ and Shi Y (2021) HandyCNV: standardized summary, annotation, comparison, and visualization of copy number variant, copy number variation region, and runs of homozygosity. *Frontiers in Genetics* **12**, 731355.

Zhou Y, Connor EE, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, Chen H and Liu GE (2018) Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. *BMC Genomics* **19**, 1–9.

Zhou Y, Utsunomiya YT, Xu L, Hay EHA, Bickhart DM, Sonstegard TS, Van Tassell CP, Garcia JF and Liu GE (2016) Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. *Scientific Reports* **6**, 29219.

https://doi.org/10.1017/S0021859625100208 Published online by Cambridge University Press