Received: 13 December 2021 | Accepted: 13 January 2023

DOI: 10.1111/age.13298

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome re-sequencing data

Elisa Peripolli[1] | **Nedenia Bonvino Stafuzza**[2] | **Marco Antonio Machado**[3] | **João Cláudio do Carmo Panetto**[3] | **Andréa Alves do Egito**[4] | **Fernando Baldi**[1] | **Marcos Vinícius Gualberto Barbosa da Silva**[3,5]

[1]School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Jaboticabal, Brazil

[2]Centro de Pesquisa em Bovinos de Corte, Instituto de Zootecnia (IZ), Sertãozinho, Brazil

[3]Embrapa Dairy Cattle, Juiz de Fora, Brazil

[4]Embrapa Beef Cattle, Campo Grande, Brazil

[5]National Council for Scientific and Technological Development (CNPq), Lago Sul, Brazil

Correspondence
Elisa Peripolli, School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), Jaboticabal 14884-900, Brazil.
Email: elisa_peripolli@hotmail.com

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 310199/2015-8; Empresa Brasileira de Pesquisa Agropecuária, Grant/Award Number: SEG 02.13.05.011.00.00; Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Grant/Award Number: CVZ PPM 00606/16; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2016/24084-7

Abstract

Further characterization of genetic structural variations should strongly focus on small and endangered local breeds given their role in unraveling genes and structural variants underlying selective pressures and phenotype variation. A comprehensive genome-wide assessment of copy number variations (CNVs) based on whole-genome re-sequencing data was performed on three Brazilian locally adapted cattle breeds (Caracu Caldeano, Crioulo Lageano, and Pantaneiro) using the ARS-UCD1.2 genome assembly. Data from 36 individuals with an average coverage depth of 14.07× per individual was used. A total of 24 945 CNVs were identified distributed among the breeds (Caracu Caldeano = 7285, Crioulo Lageano = 7297, and Pantaneiro = 10 363). Deletion events were 1.75–2.07-fold higher than duplications, and the total length of CNVs is composed mostly of a high number of segments between 10 and 30 kb. CNV regions (CNVRs) are not uniformly scattered throughout the genomes (n = 463), and 105 CNVRs were found overlapping among the studied breeds. Functional annotation of the CNVRs revealed variants with high consequence on protein sequence harboring relevant genes, in which we highlighted the *BOLA-DQB*, *BOLA-DQA5*, *CD1A*, β-defensins, *PRG3*, and *ULBP21* genes. Enrichment analysis based on the gene list retrieved from the CNVRs disclosed over-represented terms (p<0.01) strongly associated with immunity and cattle resilience to harsh environments. Additionally, QTL associated with body conformation and dairy-related traits were also unveiled within the CNVRs. These results provide better understanding of the selective forces shaping the genome of such cattle breeds and identify traces of natural selection pressures by which these populations have been exposed to challenging environmental conditions.

KEYWORDS
Bos primigenius taurus, CNV, creole breeds, next-generation sequencing, structural variants

INTRODUCTION

Copy number variations (CNVs) are chromosomal rearrangements (≥1 kb) triggered by changes in DNA content and structure (Feuk et al., 2006) leading to a change in the order (inversions and translocations) and the number of copies (duplications and deletions) of a genomic region (Henrichsen et al., 2009). CNVs represent an important source of genetic and phenotypic variability among individuals and populations (Beckmann et al., 2007; Conrad & Antonarakis, 2007; Low et al., 2019; Zhou et al., 2016), exerting a significant evolutionary impact

by generating the required variation in the population through the change in gene structure and dosage as well as by regulating gene expression and function (Zhang et al., 2009). Hence, this source of variation is as important a component of diversity as SNPs. In terms of the total number of nucleotides involved, CNVs may account for more differences among individuals than do SNPs (Conrad et al., 2010; McCarroll & Altshuler, 2007; Zhang et al., 2009). Furthermore, a significant proportion of CNVs encompass genomic regions not well covered by SNP arrays such as segmental duplications regions, and consequently, were not properly genotyped (Estivill & Armengol, 2007). Therefore, CNVs may provide genomic structural information complementary to SNP data (Scherer et al., 2007).

Different methodologies have been applied to identify CNVs at a genome-wide scale, including comparative genomic hybridization arrays, SNP-genotyping microarrays, and high-throughput sequencing (Clop et al., 2012; Di Gerlando et al., 2019). Although the first two array platforms may be affected by low probe density (Bickhart et al., 2012), they have been widely used for CNV detection in several livestock species, particularly in cattle (Bae et al., 2010; Di Gerlando et al., 2019; Fadista et al., 2010; Hou et al., 2011; Kijas et al., 2011; Liu et al., 2010). Advances in high-throughput genome scan technologies combined with appropriate algorithms have provided better approaches to systematically identify genome-wide CNVs at a higher and effective resolution, frequency, and sensitivity, allowing the identification of a vast number of structural variants, especially those that have been previously undetectable due to their small sizes (Alkan et al., 2009; Bickhart et al., 2012; Clop et al., 2012).

CNVs have been associated with heritable complex traits in several species, and lately, the interest in CNVs discovery has extended into livestock species (Dupuis et al., 2013; Fontanesi et al., 2010, 2011; Ramayo-Caldas et al., 2010). Interestingly, genome-wide CNVs studies in local and less notorious breeds have been addressed in literature (Di Gerlando et al., 2019; Molnár et al., 2014; Tian et al., 2013; Wang et al., 2016; Yang et al., 2017; Zhang et al., 2015; Zhou et al., 2014). However, despite the importance of such breeds to a wide range of challenging environments, studies deciphering their genetic structure are still a minority when compared to those accomplished in highly-specialized commercial breeds. Brazilian locally adapted taurine cattle breeds originated from the cattle brought by Portuguese conquerors in 1534 during the Brazilian colonization period (Mariante et al., 1999; Martins et al., 2009; Mazza et al., 1994; Primo, 1992). These cattle have undergone to a process of natural selection in a remarkable set of ecosystems throughout the country for more than 450 years (Mariante & Cavalcante, 2000). Although these breeds have their origin traced back to the Iberian Peninsula, each of them developed and underwent natural selection processes in distinct geographic regions, with diverse environmental conditions and successive generations of random crossings associated with breed miscegenation with distinct populations (Mariante et al., 1999). These events allowed them to acquire very particular characteristics over time and the different adaptive processes led them to genetic differentiation, being considered independent genetic entities (Campos et al., 2017; Egito et al., 2007; Peripolli et al., 2020).

Caracu Caldeano (CAR) cattle are settled in southeast Brazil, and the process of natural selection led them to acquire short hair, tolerate the high temperatures, resistance to parasites, favorable uprights and resistant hooves, short and prolapsed navel, and the ability to digest coarse fiber (McManus et al., 2010). The Crioulo Lageano (CRL) cattle population is centered in South Brazil (plateau of Lages, Santa Catarina state). This region is characterized by acidic and stony soils, uneven topography, floodplains, bush fields and abundant riparian forests, and cold winters with a high incidence of frost. The low temperatures during the winter period limit the availability of green fodder due to frost, affecting the adaptation and/or productivity of several livestock species. However, these adverse conditions shaped the CRL cattle, which are perfectly adapted to such an ecosystem (Primo, 1986). The Pantaneiro (PAN) cattle are present in the northern part of the Pantanal (mid-west Brazil). The Pantanal ecosystem is characterized by high solar radiation, parasites infestations, and flooding altering food availability (Ricklefs, 1979). The vegetation comprises forest patches, savanna, scrub savanna, and seasonally flooded grasslands, interspersed with various permanent and temporary lakes (Mourão & Medri, 2007). Natural selection conferred these cattle with exceptional rusticity, which has allowed them to thrive under water and food stress conditions, where other types of cattle would have little or no chance to survive (Mazza et al., 1992).

Further characterization of genetic structural variations, particularly in local breeds, is an important step towards deciphering the molecular mechanisms underlying trait variation, survivorship, and breed adaptation. Therefore, this study reports, for the first time, a genome-wide characterization of CNVs derived from whole-genome re-sequencing data in CAR, CRL, and PAN, three Brazilian locally adapted taurine cattle breeds. The breeds examined herein have evolved under different challenging environments and might harbor important phenotypic traits and evidence of positive selection that will help secure cattle production in a changing environment. Therefore, CNVs might harbor breed-specific adaptation footprints that may elucidate the phenotypic variation shaped by natural selection and may unravel potential biological

functions of the genes screened within the putative candidate regions.

MATERIALS AND METHODS

Samples, sequencing, and raw data preparation

Sequencing analysis was based on data from one dairy (12 CAR) and two dual-purpose (12 CRL and 12 PAN) cattle breeds. Animals were sampled from three Brazilian geographical regions, including the south (CRL), southeast (CAR), and mid-west (PAN). The population structure among the breeds together with their history and breed development can be further assessed in Peripolli et al. (2020).

DNA samples were provided from DNA banks located in Embrapa Dairy Cattle (Juiz de Fora, MG, Brazil) and Embrapa Genetic Resources and Biotechnology (Brasilia, DF, Brazil). The samples were paired-end whole-genome re-sequencing with 2 × 100 base pair reads (CRL) and 2 × 125 base pair reads (CAR and PAN) performed on the Illumina HiSeq2500 platform with an aimed average sequencing depth of 15×. Pair-end reads were aligned to the *Bos primigenius taurus* genome assembly ARS-UCD1.2 using Burrows-Wheeler Alignment MEM (BWA-MEM) tool v.0.7.17 (Becker et al., 2018) and converted into a binary format using SAMTOOLS v.1.8 (Li et al., 2009). PCR duplicates were marked using PICARD tools (http://picard.sourceforge.net, v.2.18.2).

Detections of CNVs and CNV regions

The read depth-based method implemented in CNVNATOR v.0.4.1 (Abyzov et al., 2011) software was used to call CNVs for each sample relative to the *Bos primigenius taurus* genome assembly ARS-UCD1.2. The bin size was set to 500 bp (CAR and CRL) and 600 bp (PAN) based on the ratio of the average read depth signal to its standard deviation. Quality control was undertaken to remove unreliable raw CNVs and reduce the false discovery rate. CNV calls with a *p*-value lower than 0.01 for the *t*-test statistics (*e*-val1) together with the fraction of mapped reads with zero quality (q0) lower than 0.5 and CNVs smaller than 1 kb in length were filtered out. Only autosomal chromosomes were included in the analysis.

CNV regions (CNVRs) were identified by overlapping individual CNVs within each breed (Redon et al., 2006), and only those found overlapping in all individuals within a breed by at least 1 bp were used for downstream analysis. Shared CNVRs among the studied breeds were also identified by overlapping the CNVRs identified within each breed, and only those described overlapping in at least two breeds were used for further analysis. Overlapping analyses were carried out using the BIOCONDUCTOR package *GenomicRanges* (Lawrence et al., 2013).

Correspondence analysis

Correspondence analysis is a multivariate method from categorical data analogous to principal component analysis, which leads to a low-dimensional graphical representation of a contingency table as points in a metric space (Weller, 2005). Therefore, shared CNVRs were used to access the structure of the locally adapted cattle breeds to obtain a comparative view of such breeds based on CNVR clusters. Shared CNVRs overlapping in at least two breeds were used as variables to spatially cluster the CNVRs using the Statistical Analysis System (SAS version 9.3) software and the CORRESP procedure based on Chi-square distances to judge proximity among them.

Predicted functional impacts, gene annotation, and enrichment analysis

A functional annotation analysis of the called variants (CNVRs) was performed to assess their possible biological impact using the Variant Effect Predictor (VEP; McLaren et al., 2016) together with the Ensembl genes release 100, version April 2020 (assembly ARS-UCD1.2). Variants with a high consequence on protein sequence (i.e., splice acceptor variant, splice donor variant, stop gained, frameshift variant, stop lost, and start lost) were selected for further assessment.

Genes were annotated within the CNVRs using the cow gene set Ensembl genes release 100 (ARS-UCD1.2) fetched from the BIOMART tool (Haider et al., 2009). Database for Annotation, Visualization, and Integrated Discovery (DAVID) v.6.8 tool (Huang et al., 2009a, 2009b) was used to identify overrepresented ($p<0.01$) Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the list of genes from the CNVRs and the bovine annotation file as a background. QTL retrieved from the CattleQTL database (Hu et al., 2016) were overlapped with the CNVRs using BEDTOOLS (Quinlan & Hall, 2010).

RESULTS

Data

With Illumina paired-end sequencing technology, we obtained re-sequencing data from 36 individuals from three different Brazilian locally adapted taurine cattle breeds. After mapping the reads to the genome assembly ARS-UCD1.2, an average coverage depth of 14.07×

was obtained. As disclosed in the literature, an average coverage depth between 4 and 8× allows sufficient power for CNVs detection using the read depth-based method (Bickhart et al., 2012; Sudmant et al., 2010).

CNV and CNVRs discovery

Four outlier samples (one for CRL and three for PAN) were filtered out from the dataset after CNV calling due to the discrepant number of CNVs identified.

A total of 7285 CNVs (4640 deletions and 2645 duplications) was identified in the CAR breed. On an individual animal basis, the average number of CNVs per animal was 607.08, with an average length of 28.30 kb and encompassing approximately 0.63% (17.18 Mb) of the total autosomal genome extension (ARS-UCD1.2). In the CRL breed, the total number of CNVs was 7297 (4726 deletions and 2571 duplications), displaying an average number of 663.36 CNVs per animal together with an average length of 27.60 kb and covering roughly 0.67% (18.31 Mb) of the total autosomal genome extension. For the PAN breed, 10 363 CNVs (6998 deletions and 3365 duplications) were identified, with an average number of 1151.44 CNVs per animal and an average length of 34.06 kb, encompassing nearly 1.44% (39.22 Mb) of the total autosomal genome extension.

The longest CNVs within each breed were very close in size among the studied breeds and were all events of deletion, with values of 1004.99 kb in length on BTA10:23 775 501–24 780 500 bp (CRL), 1006.99 kb in length on BTA10:23 773 501–24 780 500 bp (CAR), and 1007.39 kb in length on BTA9:104 447 401–105 454 800 bp and BTA10:23 773 201–24 780 600 bp (PAN). Remarkably, the genomic region on BTA10:23 775 501–24 780 500 bp was found overlapping in all three breeds within the longest CNVs described. When inspecting in detail, such genomic region did not harbor any gene or QTL. The number of CNVs per chromosome was greater on BTA1 for the PAN (n = 662) cattle and on BTA15 for the CRL (n = 518) and CAR (n = 496) cattle breeds (Appendices S1–S3). The total length of CNVs for the studied breeds was composed mostly of a high number of segments between 10 and 30 kb, which accounted for approximately 47% (CAR; n = 3443 and CRL; n = 3422) and 55% (PAN; n = 5737) of all CNVs detected (Figure 1a).

The CNVRs were not evenly distributed throughout the genomes, with some chromosomes missing CNVRs

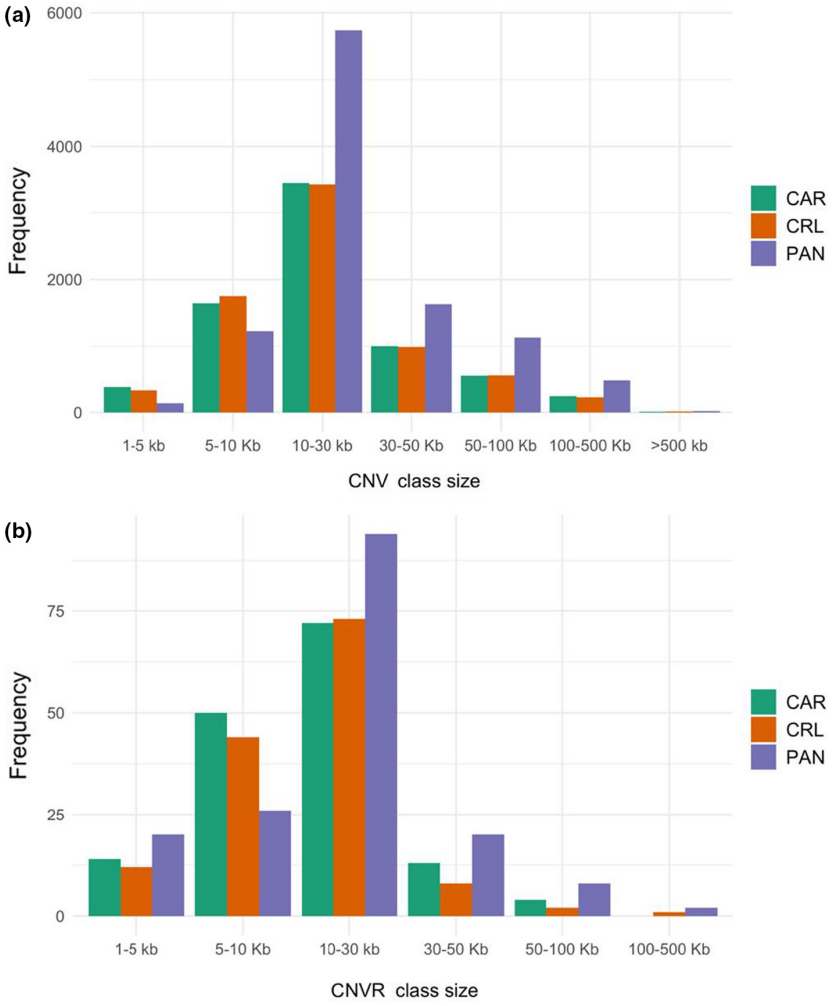

FIGURE 1 (a) Copy number variation (CNV) length class size range distribution for Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds. (b) Copy number variations region (CNVR) length class size range distribution for CAR, CRL, and PAN cattle breeds

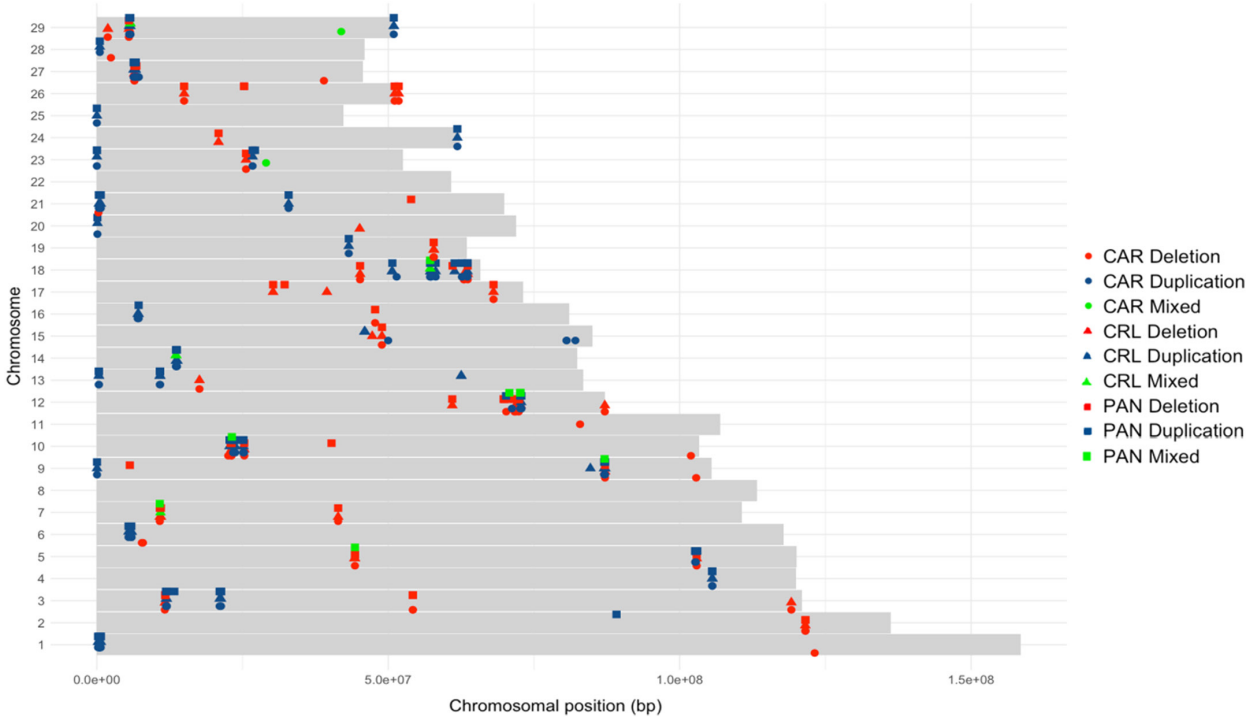

FIGURE 2 Copy number variation region (CNVR) scattering in the Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle genomes according to autosomal length (ARS-UCD1.2). Dots depicting the breeds: circle (CAR), triangle (CRL), and square (PAN). Dots depicting the CNVR events: deletion (red), duplication (blue), and mixed (green) events

and others containing several such regions (Figure 2, Appendices S4–S6). The total length of CNVRs is also composed mostly of a high number of segments between 10 and 30 kb in length (Figure 1b). A total of 153 CNVRs were identified in the CAR breed, including 49 deletions, 102 duplications, and two mixed (deletion and duplication within the same region) events. Such CNVRs covered roughly 0.09% (2.45 Mb) of the autosomal genome extension (ARS-UCD1.2), with an average length size of 16.05 kb and values ranging from 1.00 to 79.50 kb. In the CRL breed, the total number of CNVRs was 140 (46 deletions, 86 duplications, and eight mixed events), covering approximately 0.08% (2.17 Mb) of the autosomal genome extension with an average length size of 15.53 kb and values ranging from 0.50 to 114.50 kb. For the PAN breed, a total of 170 CNVRs were described, encompassing 61 deletions, 99 duplications, and 10 mixed events. The CNVRs covered nearly 0.13% (3.60 Mb) of the autosomal genome extension, with an average length size of 21.122 kb and values ranging from 0.50 to 200.50 kb.

The number of CNVRs per chromosome was greater on BTA1 for the CAR (n = *17*) and CRL (n = 13) cattle breeds (Figure 3a,b, respectively), and BTA12 showed the greatest enrichment for the PAN (n = 29) cattle (Figure 3c). It is worth highlighting that the number of CNVRs duplication events was higher (~1.85-fold) than did the deletions.

Shared CNVRs by the three breeds (n = 105) were observed, with a length size varying from 1.00 to 52.00 kb and a mean size of 14.34 kb (Appendix S7, Figure 4). Further, breed-specific genomic regions could be observed for CAR (n = 70), CRL (n = 43), and PAN (n = 128) when overlapping the CNVRs among the studied breeds.

Correspondence analysis

The CNVR structure among the breeds was dissected by analyzing the first two dimensions (Figure 5 and Appendix S8). Shared CNVRs among the breeds clustered in four main groups and a clear separation could be observed between the breeds and CNVR clusters. Dimension 1 accounted for roughly 63% of the variance and it clearly separated CNVR groups 1 and 2. These two CNVR groups differentiated CRL and CAR cattle breeds. The PAN cattle were found associated with CNVR group 1 and CAR cattle as well as with CNVR group 2 and CRL cattle. Despite being differentiated, CRL and CAR might share some CNVRs with PAN cattle. CNVR group 3 is likely to share CNVRs with all breeds. When analyzing dimension 2 (36.7% of the variance), the CNVR group 4 is the group that least correlated with the breeds. It is worth highlighting that CAR and CRL share different CNVR groups, and the little they have in common might be encompassed in CNVR group 4.

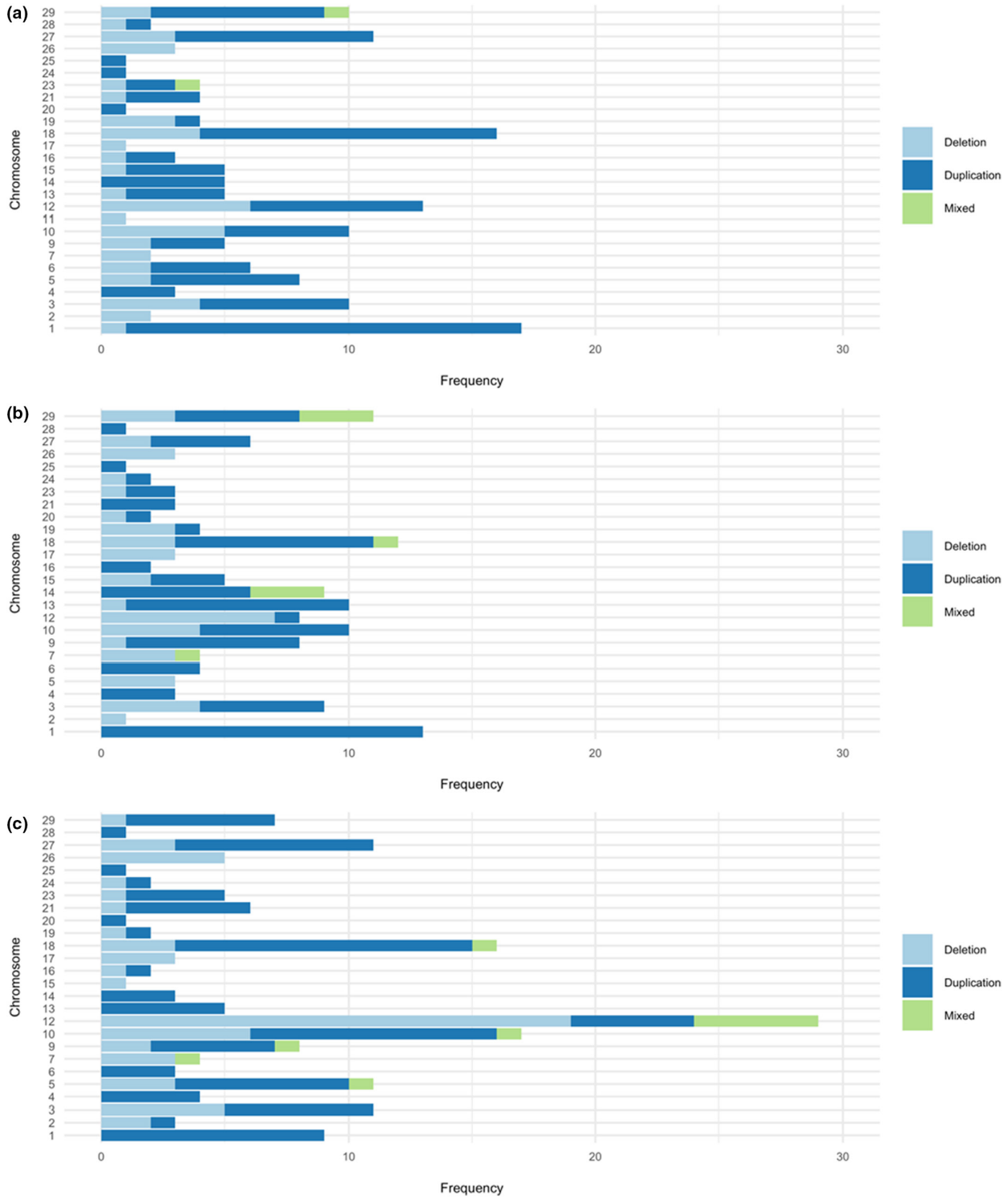

FIGURE 3 Frequency distribution of copy number variation regions (CNVRs) according to CNVR event (deletion, duplication, and mixed). (a) Caracu Caldeano cattle breed, (b) Crioulo Lageano cattle breed, (c) Pantaneiro cattle breed

When inspecting in detail, all CNVR groups harbored several protein-coding genes (Appendix S9). It is worth highlighting that CNVR group 1 comprised five protein-coding genes, in which we can underscore the *DEFB7* gene. CNVR group 2 harbored 11 protein-coding genes, with emphasis on the *TRBV3-1* gene. CNVR group 3 harbored the highest number of protein-coding genes ($n = 15$), and a β-defensin and a bovine leukocyte antigen (BoLA)-related gene were described within this group. CNVR group 4 displayed one protein-coding gene with no described function.

Variant and functional annotation of genes

Functional classification showed that most of the variants identified within the CNVRs were located in intergenic and intronic regions (Appendix S10), and several variants with a high consequence on protein sequence were identified (CAR $n = 43$; CRL $n = 37$; PAN $n = 57$; and shared CNVRs by the three breeds $n = 53$; Appendices S11–S14). Following variant annotation, we further investigated the gene content within the predicted variants to cause relevant biological functions. Totals of 30, 22, 42, and 26 protein-coding genes were described within variants with a high consequence on protein sequence for CAR, CRL, PAN, and shared CNVRs by the three breeds, respectively. Among them, it is worth underscoring the *BOLA-DQB*, *BOLA-DQA5*, *CD1A*, and some β-defensins genes (i.e., *DEFB13* and *DEFB7*), which were identified in all breeds (Figure 6). Further, genes such as the *PRG3* and *ULBP21* were particular to CAR cattle. All of them have been strongly linked to cattle environmental resilience, including immune response and ectoparasite resistance.

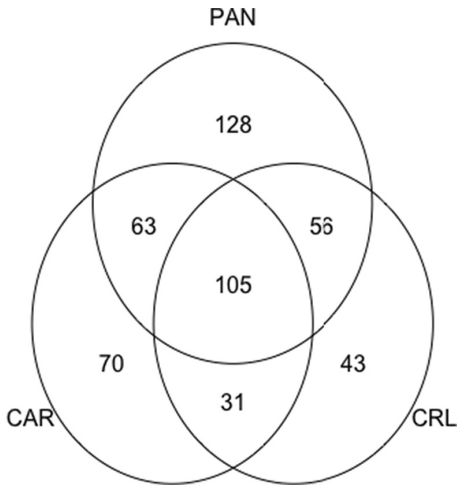

FIGURE 4 Venn diagram of copy number variation regions (CNVRs) for Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds

Enrichment analysis of genes

Enrichment analysis was performed to obtain a broad functional insight into the set of genes (Appendix S15) observed in CNVRs described in each breed, as well as in shared CNVRs by the three studied breeds. GO enrichment analysis revealed five biological processes, three molecular functions and four cellular component processes enriched ($p<0.01$, Table 1), and suggested that several of the CNVRs genes are mainly enhanced in functions related to the immune response. CAR and PAN cattle breeds showed enrichment of β-defensins genes encompassing the over-represented terms, whereas CRL and shared CNVRs displayed an enhancement of BoLA-related genes. Some overrepresented terms (i.e., GO:0042742, GO:0005576, GO:0002504, GO:0042613) were described in more than one breed. Besides, the over-represented in shared CNVRs (GO:0002504 and GO:0042613) have been previously identified when analyzing the breeds individually.

Genes within the breed-specific genomic regions disclosed for each breed when overlapping the CNVRs were also annotated (Appendix S16), suggesting that these regions might be associated with breed-specific differences in adaptative and productive traits.

CNVRs and overlapping QTL in cattle

CNVRs were disclosed in genomic regions containing QTL in cattle formerly implicated in body conformation ($n = 2$) and dairy-related traits ($n = 10$; Appendix S17). It

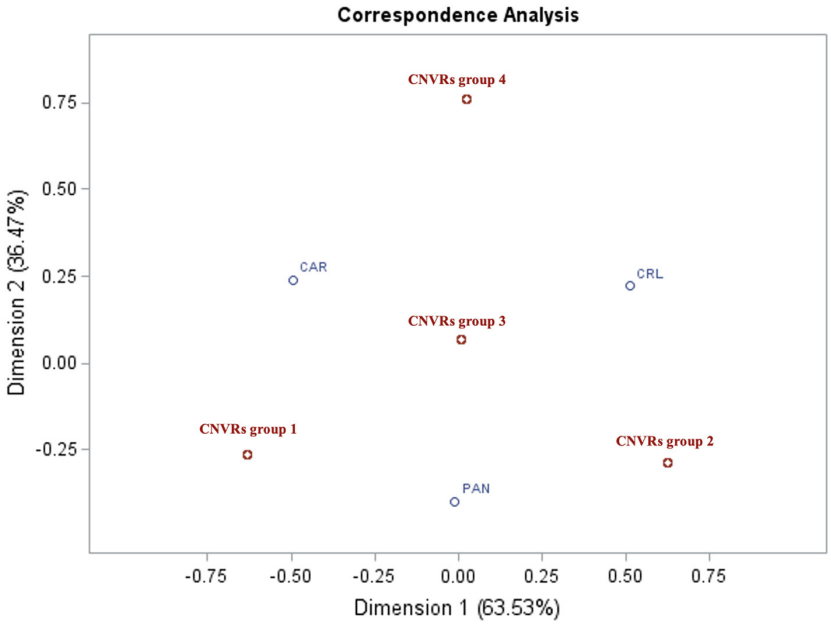

FIGURE 5 Two-dimensional correspondence analysis of copy number variants regions (CNVRs) with variance explained by the first two dimensions in brackets for Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds

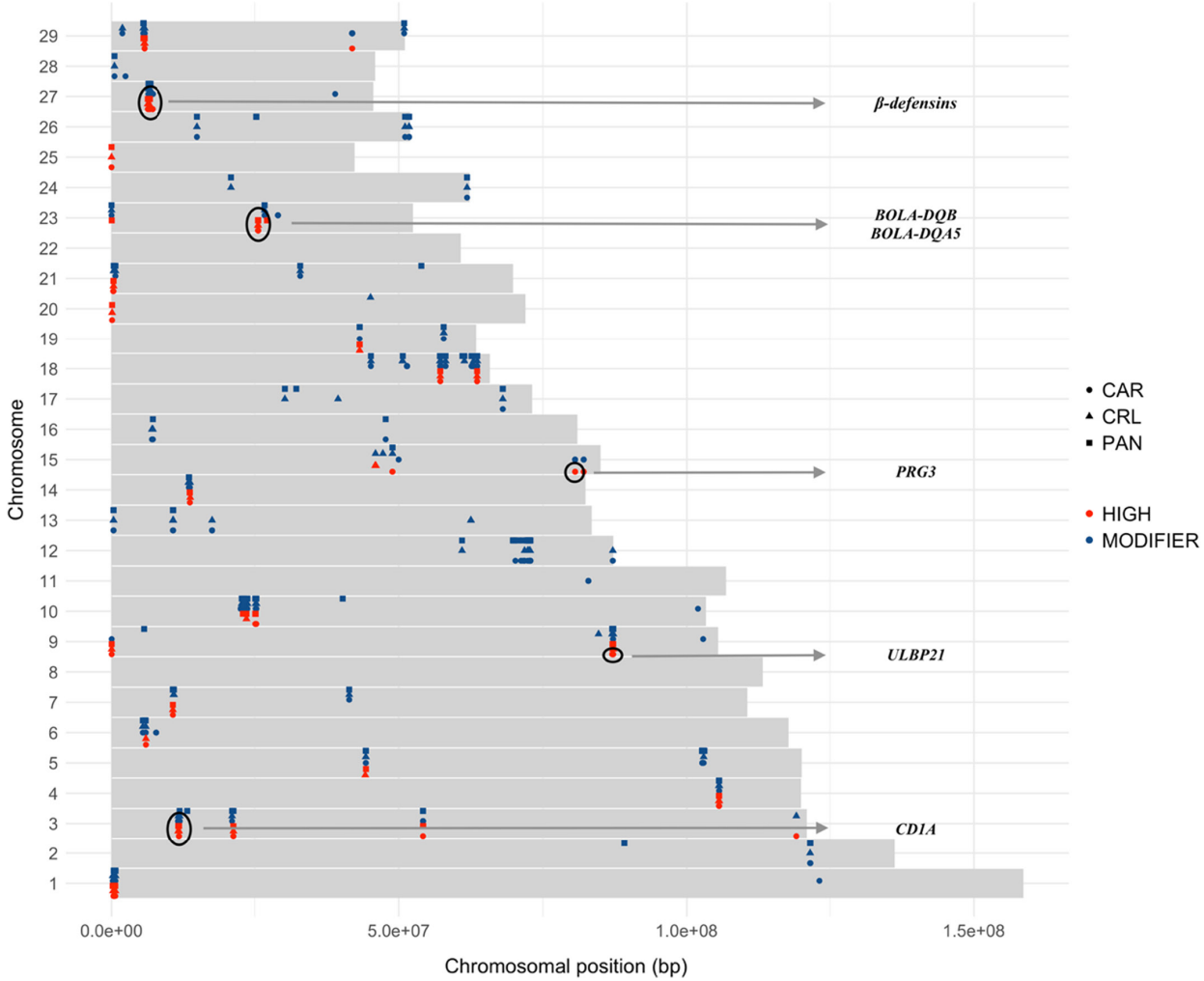

FIGURE 6 Variants scattering in the Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle genomes according to autosomal length (ARS-UCD1.2). Dots depicting the breeds: circle (CAR), triangle (CRL), and square (PAN). Dots depicting the putative variant impact: high (red) and modifier (blue)

is noteworthy to underscore that most of the QTL described herein were found within shared CNVRs. The CAR and CRL cattle did not display any further QTL besides those described in the shared CNVRs by the three studied breeds. Further, the PAN cattle displayed QTL related to milk protein percentage and fatty acid content on BTA3 and BTA29, respectively, in addition to those identified within the shared CNVRs by the three breeds. It should be noted that the majority of the QTL harbored duplication events and just one on BTA17:68 058 001–68 079 500 bp (nonreturn rate QTL; Frischknecht et al., 2017) was found encompassing a deletion event.

DISCUSSION

The Global Animal Genetic Data Bank created by the Food and Agriculture Organization contains many reports of local breeds that are thought to show resistance/tolerance to particular diseases and parasites. However, many of these reports are based on personal accounts rather than scientific studies (Hoffmann, 2010). This study is the first of its kind to bring out scientific findings that validate the reports observed by technicians and breeders working with locally adapted breeds in Brazil. These findings could provide the basis to better explore the use of such breeds in crossbreeding programs to transfer their adaptability and rusticity to commercial breeds.

Discovery of CNVs and CNVRs

The widespread availability of array-based methods has led to much interest in the discovery and mapping of CNVs and their association with phenotypes (Yau & Holmes, 2008). Previous studies assessing CNVs in several cattle breeds have been mainly based on array comparative genomic hybridization (aCGH) (Fadista

TABLE 1 Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes pathways analysis enriched ($p<0.01$) based on copy number variation regions identified within each breed (Caracu Caldeano, Crioulo Lageano, and Pantaneiro) and based on shared copy number variation regions by the three breeds

Category	Term	*n* genes	*p*-value	Genes
Caracu Caldeano				
MF	GO:0047961~glycine *N*-acyltransferase activity	3	5.24E-06	*GLYAT, GAT, GLYATL2*
BP	GO:0042742~defense response to bacterium	4	8.38E-05	*DEFB7, EBD, DEFB13, DEFB4A*
BP	GO:0006955~immune response	4	1.47E-03	*PRG3, BOLA-DQA5, BOLA-DQB*
CC	GO:0005576~extracellular region	5	3.59E-03	*DEFB7, EBD, PRG3, DEFB13, DEFB4A*
MF	GO:0046703~natural killer cell lectin-like receptor binding	2	4.57E-03	*ULBP21, RAET1G*
Crioulo Lageano				
BP	GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2	7.79E-03	*BOLA-DQA5, BOLA-DQB*
CC	GO:0042613~MHC class II protein complex	2	9.79E-03	*BOLA-DQA5, BOLA-DQB*
Pantaneiro				
MF	GO:0005044~scavenger receptor activity	3	2.04E-04	*WC1, CD163L1, WC1.3*
BP	GO:0042742~defense response to bacterium	3	1.48E-03	*DEFB7, DEFB13, DEFB10*
CC	GO:0005576~extracellular region	4	7.76E-03	*DEFB7, CD163L1, DEFB13, DEFB10*
Shared CNVRs				
BP	GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2	5.20E-03	*BOLA-DQA5, BOLA-DQB*
CC	GO:0042613~MHC class II protein complex	2	6.13E-03	*BOLA-DQA5, BOLA-DQB*

Abbreviations: BP, biological process; CC, cellular component; MF, molecular function; MHC, major histocompatibility complex.

et al., 2010; Liu et al., 2010, 2019) and SNP arrays (Bae et al., 2010; Cicconardi et al., 2013; Hou et al., 2012; Jiang et al., 2013; Yang et al., 2017; Zhang et al., 2015). Although they promoted the progress of CNV studies, much has been discussed about the limitations of such methodologies associated with the power to detect CNVs (Lai et al., 2005; Pinto et al., 2011; Winchester et al., 2009). Studies have reported that coverage bias and platform resolution resulted in differences regarding the number and sizes between CNVs when using next-generation sequencing (NGS) and array-based methods (Ben Sassi et al., 2016; da Silva et al., 2016; Jiang et al., 2013; Zhan et al., 2011). Hence, differences in CNV calls from different platforms make the comparison among studies not straightforward and emphasize the importance of a careful assessment when contrasting studies.

Current studies on local and endangered cattle breeds using whole-genome resequencing data are very minimal when compared to specialized (i.e., dairy and beef) breeds (Ben Sassi et al., 2016; Bickhart et al., 2012; Boussaha et al., 2015; Gao et al., 2017; Stothard et al., 2011; Zhan et al., 2011). Accordingly, we investigated structural variations in three Brazilian locally adapted cattle breeds using a read depth approach based on whole-genome resequencing data. Our results revealed that CNVs are non-uniformly scattered across the genomes and represent a small proportion of the reference assembly used for mapping (~0.63–1.44%), as also reported for other cattle populations (Bickhart et al., 2012; Stothard et al., 2011; Zhan et al., 2011; Zhang et al., 2015). The number of autosomal CNVs identified in each breed is consistent with previous reports based on NGS data by Stothard et al. (2011) and Zhan et al. (2011), and higher than those described by Bickhart et al. (2012) and Ben Sassi et al. (2016). The PAN cattle displayed the highest number of CNVs among the studied breeds. It is worth highlighting that the PAN breed has the smallest population among the Brazilian locally adapted cattle breeds. Moreover, the evaluated population comes from a breed conservation

center located at Embrapa Pantanal (Corumbá, MS, Brazil), in which crossbreeding is practiced to maintain the maximum genetic variability. Therefore, the higher number of CNVs described for the PAN cattle may be a function of the crossing model applied to maintain such genetic variability together with the absence of any genetic improvement program.

Deletions events were approximately 1.75–2.07-fold more recurrent than duplications, concurring with former NGS studies for taurine cattle breeds (~1.72-fold, Gao et al., 2017 and 1.15-fold Boussaha et al., 2015). The increased number of deletions described herein might be associated with the mechanism by which CNVs are formed within the genome. Studies have shown that non-homologous end-joining formation is the major mechanism responsible for deletion and translocations (Shaw & Lupski, 2005; Toffolatti et al., 2002). Non-homologous end-joining is a repair mechanism frequently initiated in response to double-strand breaks after DNA processing (Van Gent & Van Der Burg, 2007), and it can occasionally error-prone, leading to loss or small insertion of nucleotides at the lesion site (Labhart, 1999).

The sizes of the identified CNVs mostly ranged from 10 to 30 kb for all breeds, with a few outliers having a size greater than 500 kb. Such results are consistent with those based on SNP array (Bae et al., 2010; Lemos et al., 2018; Wu et al., 2015; Zhang et al., 2015) on diverse cattle breeds; however, it differed from NGS data (da Silva et al., 2016) in which CNVs were most frequent between 100 and 200 kb. Nevertheless, it is worth underscoring that the CNVR size range distribution concurred with those described in the literature for both SNP array-based and NGS data (Ben Sassi et al., 2016; da Silva et al., 2016; Gao et al., 2017; Lemos et al., 2018).

Correspondence analysis

The correspondence analysis allowed us to discriminate the three breeds based on CNVR clusters as well as to visualize how the gene content within the CNVRs were related to each breed. CNVR groups differentiated CRL and CAR cattle, and the PAN cattle were found associated with both CAR and CRL cattle. The differentiation between CRL and CAR may be explained by the cattle type introduced by European conquerors during Brazilian the colonization period (Mazza et al., 1994). Portuguese purebred cattle brought to Brazil belonged to three different bloodlines: *Bos taurus aquitanicus*, *Bos taurus batavicus*, and *Bos taurus ibericus*. CRL and PAN cattle originated from a common ancestral pool and have their ancestry in breeds from *Bos taurus ibericus* cattle, while the CAR cattle descended from the *Bos taurus aquitanicus* bloodline (Serrano et al., 2004). Further, the divergence between CRL and CAR cattle may be a result of artificial selection events over time. CAR animals have been selected for milk production traits in the southeastern region of Brazil since 1893 (de Queiroz et al., 2005), whereas CRL and PAN animals are mainly used in animal genetic resources conservation programs (Mariante et al., 1999, 2009). Our previous study also described overlap of genetic variation between PAN and CRL cattle breeds (Peripolli et al., 2020). These results might reflect the breed production type since both breeds are dual-purpose use and have their origin traced back to *Bos taurus ibericus* cattle as well as from the Spanish expeditions in the Rio da Prata basin (Mazza et al., 1994).

CNVR groups 1 and 2 differentiated CRL and CAR cattle when considering the highest proportion variance explained (63.53%) and harbored the *DEFB7* and *TRBV3-1* genes, respectively. The first gene belongs to the β-defensin family, which are antimicrobial peptides contributing to host defense during infection by acting against many unicellular parasites (Brogden, 2005; Lehrer et al., 1993; Nicolas & Mor, 1995) and are responsible for inducing primary immunological responsiveness (Banchereau et al., 2000; Sakaguchi et al., 2008). The potential of Brazilian locally adapted cattle breeds responsiveness was demonstrated by Maggioli et al. (2013). The *TRBV3-1* gene is closely associated with T lymphocytes and T-cell response, playing a central role in the adaptive immune response. In vertebrates, α/β T-cell receptors are antigen specific receptors essential to the immune response present on the cell surface of T lymphocytes (Massari et al., 2018), being crucial to the maintenance of effective T cell-mediated immunity to a wide variety of pathogenic organisms (Houston et al., 2005).

Copy number variations regions group 3 was described to share CNVRs with all breeds, and encompassed the *BOLA-DQB and DEFB13* (β-defensin family) genes. BoLA family member genes are located within the major histocompatibility complex region. In cattle, the major histocompatibility complex region is known as the BoLA, which is on BTA23 (Fries et al., 1993). BoLA plays a crucial role in determining immune responsiveness, and genetic variations in such region has been greatly associated with disease susceptibility and resistance (reviewed by Takeshima & Aida, 2006). Additionally, several cattle studies have described CNVs adjacent to the BoLA region (Hou et al., 2011; Liu et al., 2010; Porto-Neto et al., 2013; Prinsen et al., 2017; Zhou et al., 2016).

CNVs within the locally adapted cattle breeds are affecting particular genes with similar biological functions in different CNVR cluster groups. The Brazilian locally adapted cattle breeds are considered as hardy breeds with excellent adaptability to challenging environments. Genes that are involved in the triggering and regulation of innate immune responses were detected in CNVR cluster groups, which is consistent with the high level of resistance to diseases and endoparasites of such breeds.

Variant annotation and enrichment analysis

Genome-wide characterization of CNVs and the comprehensive assessment of CNVRs are a powerful strategy to ascertain potential key genes and biological mechanisms encompassing traits of interest in several livestock species. In this regard, CNVRs identified herein were better assessed to predict the impact of variants on protein sequence and determine their likely biological effects. Further, the gene content within those regions were inspected in detail to disentangle their roles in shaping particular characteristics and phenotypes of the studied populations. When further investigating the gene content harboring variants with a high consequence on protein sequence, the majority of them were described to be closely linked to adaptation and immune response functions rather than productivity. Several family member genes with similar biological functions were described harboring variants within the three breeds. Among them, we can underscore BoLA family member genes, β-defensins genes, and the *CD1A* gene.

Beta-defensins are antimicrobial peptides acting against many Gram-positive and negative bacteria, fungi, enveloped viruses, and other unicellular parasites (Brogden, 2005; Lehrer et al., 1993; Nicolas & Mor, 1995). Antimicrobial peptides are among the most evolutionarily ancient molecules of the immune system and are present in a variety of vertebrates, insects, and plants (Selsted & Ouellette, 2005). Besides their antimicrobial activity, β-defensins have chemoattractant activity for immature dendritic and T cells (Yang et al., 1999), playing a critical role in the immediate reaction to a broad spectrum of pathogens by inducing primary immunological responsiveness (Banchereau et al., 2000; Sakaguchi et al., 2008). Further, bovine β-defensins located within the bovine cluster D are mainly expressed in the mammary gland, and therefore, contribute to local host defense and impart resistance against intramammary infections (Gurao et al., 2017). The *CD1A* gene has been described to be highly expressed at the tick attachment site from Holstein-Friesian animals (Piper et al., 2008). A study on Angus cattle (Hou et al., 2012) revealed that parasite resistance animals with high estimated breeding values for eggs/g displayed such gene within regulatory networks linked to gastrointestinal nematodes.

Two genes encompassing variants with high impact on protein sequence were described only for the CAR cattle (*ULBP2* and *PRG3* genes), and they should be highlighted given their role in cattle adaptation. It is hypothesized that the cattle *ULBP* gene family evolved under adaptive diversifying selection in response to selective pressure exerted by a viral pathogen (Larson et al., 2006). The *PRG3* gene has been associated with tick resistance. It forms a protective barrier by stimulating the histamine biosynthetic process and activating basophils, which are important effectors of tick rejection and a major component of the acquired resistance of the host (Falcone et al., 2001; Wikel, 1996). Such mechanism leads to an unfriendly environment for tick attachment and feeding (Kongsuwan et al., 2008).

All of the previously discussed genes have been described within the significant GO terms, strongly supporting their enriched functions associated with immunity and cattle resilience to harsh environments. CAR and PAN cattle showed enriched terms ($p < 0.01$) encompassing mostly β-defensin family member genes, while CRL predominantly displayed BoLA family member genes. Differences in gene groups observed may be related to adaptation to the climatic regions where these breeds developed. CAR and PAN are reared in the tropical zone while the CRL in the southern temperate zone. It should be noted that only one over-represented term (GO:0047961~glycine *N*-acyltransferase activity) has not been directly associated somehow with immune-related functions. Several other CNV cattle studies displayed an enrichment of genes linked to immune response and environmental interaction, including sensory response and chemical stimuli (Bickhart et al., 2012; Stothard et al., 2011; Upadhyay et al., 2017; Wang et al., 2015; Yang et al., 2017). Immune-related genes seem to be evolved under positive selection (Sackton et al., 2007), reflecting a coevolutionary process between infectious pathogenic exposure and the host's defense system to acquire a broad range of antimicrobial defense (Luenser & Ludwig, 2005; McTaggart et al., 2012). Therefore, it has been hypothesized that the increased dosage of such genes may offer survivability and adaptive benefits (Liu et al., 2010; Nguyen et al., 2008), suggesting that adaptation to diverse pathogenic environments have probably exerted important selective forces in the cattle genome.

It is not surprising that an abundance of genes and over-represented terms were found described to be involved in processes closely associated with immune functions and parasite resistance. The Brazilian locally adapted cattle breeds studied herein exhibit distinguishing levels of phenotypic variability and enhanced fitness to local conditions due to a long process of natural selection in extremely variable and harsh environments (Mariante & Cavalcante, 2000). Such breeds have undergone strong environmental pressures for more than 450 years without any significant selective pressure imposed by man, facing adverse tropical climate conditions (heat, dryness, and humidity), limited food availability, disease' susceptibility, and parasite infestations (Mariante & Cavalcante, 2000). Hence, these limitations led them to acquire very particular traits over time to thrive in such distinct ecosystems (Mariante et al., 1999) and may have left footprints of selection within their genome.

Breed-specific CNVRs

Among the genes within the breed-specific CNVRs for the CAR cattle, we can emphasize the *NBAS* gene, the

function of which has been associated with stature and bone development (Duan et al., 2021). CAR animals have the greatest body size among the locally adapted cattle breeds, and much has been discussed about the relationship between body size and environmental adaptation due to climate and/or driven by changes in feed resources and seasonal influences (Gardner et al., 2011; Martin et al., 2018). The *EBD* gene is a bovine β-defensin gene member (BoLA) involved in interleukin-17 signaling, and it was also described for CAR animals. This gene is highly expressed in enteric epithelial cells and may contribute to host defense of enteric mucosa by conferring intestinal immunity (Tarver et al., 1998).

The *DEFB4A* and *PSMD13* genes were also identified within CAR-specific genomic CNVRs. The *DEFB4A* is also a β-defensin gene that acts as an antimicrobial peptide against Gram-negative and Gram-positive bacteria, conferring antimicrobial resistance of the mammary gland (Brogden, 2005). This gene was also described to have an effect on milk-related traits, and polymorphisms were associated with protein yield, fat content, and somatic cell score (Bagnicka et al., 2007, 2008), and milk yield (Krzyzewski et al., 2008). The *PSMD13* gene plays a key role in the maintenance of protein homeostasis by removing misfolded and damaged proteins (Tanaka, 2009). This gene has also been associated with lactation persistency in Canadian Holstein cattle (Do et al., 2017). Such genes reflect the objective of selection for dairy-related traits in which the CAR cattle have been subjected since 1893 (de Queiroz et al., 2005).

The CRL cattle displayed the *BPIFA2A* gene within the breed-specific genomic regions. This gene play a role in the local antibacterial response in nose, mouth, and upper respiratory pathways (provided by RefSeq, Jan 2016), and could be related to the adaptive capacity of this breed to lower temperatures, which can be negative at certain times of the year. Further, this gene is important for the intramuscular muscle' profile (Berton et al., 2016), reflecting the mild selection for meat-related traits applied on such breed recently (Mitterer-Daltoé et al., 2012).

The PAN cattle harbored several genes within the breed-specific CNVRs, and among them we can underscore the *AOX1*, *SORCS3*, and *PDE5A* genes. The *AOX1* gene has been mainly associated with adipogenesis and lipid metabolism (Brandes et al., 1995; Mei et al., 2018), and the *SORCS3* gene with energy homeostasis (Subkhangulova et al., 2018) and increased adiposity (Purfield et al., 2019). Food shortages represent a common challenge for most animal species, and the coordination of energy partitioning and homeostasis is a challenge to sustainable intensification of livestock productivity in the tropics. As previously discussed, the PAN cattle evolved under an ecosystem in which the flooding season alters food availability (Ricklefs, 1979). Consequently, such cattle may have evolved metabolic strategies encompassing extreme starvation-resistance capabilities, being able to store and mobilize lipids during nutritionally stressful environmental conditions (Olsen et al., 2021). It should be noted that negative energy balance probably impairs reproductive performance (Stockdale, 2001) and increases the susceptibility to infections (Collard et al., 2000). Hence, animals that were able to minimize the mobilization of adipose tissue reserves in response to the energy deficit might have conferred fitness advantage.

The *PDE5A* gene was also described for the PAN cattle, and it is widely known as a regulator of nitric oxide-induced vasodilation. Vasodilation is the mechanism in which blood vessels dilate to dissipate heat to external environment, and it can be regulated by interactions between nitric oxide, PDE5, and guanosine 30,50-cyclic monophosphate (Coppage et al., 2005). The Pantanal ecosystem is characterized by high solar radiation with temperatures up to 40°C early in the summer (Por, 2012), and cattle breeding in such tropical regions may be affected by heat stress when the mechanisms of body thermoregulation are unable to promote heat loss adequately. Therefore, *PDE5A* might be a key gene in elucidating the better tolerance of the PAN cattle to the Pantanal ecosystem through vasodilation during times of increased heat stress.

CNVRs and overlapping QTL in cattle

Most of the CNVRs overlapped with previously reported regions harboring QTL that mostly affect dairy-related traits, and two reasons might have led to this result. First, when examining in detail the QTL associations by trait classes in the CattleQTL database (Hu et al., 2016), the greatest number of reported QTL (~36%) has been associated with milk-related traits (n = 50 208), followed by reproductive (n = 44 369), and productive (n = 22 519) traits. The second reason relies on the fact that the CAR breed has been selected for milk production traits in the southeastern region of Brazil since 1893 (de Queiroz et al., 2005). Further, the remaining two breeds despite not being considered high-specialized cattle breeds are classified as dual-purpose and might have undergone mild selection for dairy-related traits (Lara et al., 2002; Oliveira-Brochado et al., 2018).

FINAL CONSIDERATIONS

By using whole-genome re-sequencing data, we reported for the first time a genome-wide characterization of CNVs in three Brazilian locally adapted taurine cattle breeds. Our results provide substantial information about the potential use of CNVs to identify putative regions that have been functionally relevant and have played a substantial role in shaping the genome of such cattle breeds based on the environmental conditions

in which they have been raised. Enrichment analysis, variant annotation, correspondence analysis, and QTL identification retrieved from the CNVRs revealed a large proportion of genes associated with immune system functioning, parasite resistance, and productive-related traits. When inspecting breed-specific CNVRs in detail, the gene content probably reflects the breed formation process in harsh environments and the objective of selection for dairy-related traits for one cattle breed.

These results provide evidence of natural selection for traits linked to cattle resilience to challenging environments. The cattle populations studied herein represent an important biological model for understanding the role of environmental stressors and the effect of different selective forces acting on the genome diversity of the Brazilian locally adapted taurine cattle breeds. The identification of genomic regions harboring structural variations plays an important role in the introgression of locally adapted breeds in crossbreeding schemes. Hence, production systems may benefit from the introduction of crossbred animals, taking advantage of animals better adapted to local conditions displaying key adaptative traits for survival in challenging environments together with production traits from high-specialized cattle breeds.

ACKNOWLEDGMENTS

The authors would like to thank the Animal Genetic laboratory from the National Research Centre for Genetic Resources (Cernagen) for providing DNA samples.

FUNDING INFORMATION

E.P. was supported by São Paulo Research Foundation (FAPESP) grant # 2016/24084-7. MVGBS was supported by Embrapa (Brazil) SEG 02.13.05.011.00.00 and CNPq 310199/2015-8 "Detecting signatures of selection from Next Generation Sequencing Data", MCTI/CNPq/INCT-Ciência Animal and FAPEMIG CVZ PPM 00606/16 "Detecting signatures of selection in cattle from Next Generation Sequencing Data" appropriated projects.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

The genomic information used in this study is available from EMBRAPA – Brazilian Agriculture Research Corporation (EMBRAPA SEG 20.18.01.018.00.00), but restrictions apply to their public availability. However, data are available for sharing upon reasonable request and with permission of the author Marcos Vinícius Gualberto Barbosa da Silva, e-mail: marcos.vb.silva@embrapa.br.

ETHICAL APPROVAL

The DNA was extracted from semen and blood samples bought from artificial insemination centers, and therefore, no specific ethical approval is needed (Brazil law number 11794, from October 8th, 2008, Chapter 1, Art. 3, paragraph III) and no restriction apply for their use for research purpose.

ORCID

Elisa Peripolli https://orcid.org/0000-0002-0962-6603
Nedenia Bonvino Stafuzza https://orcid.org/0000-0001-6432-2330

REFERENCES

Abyzov, A., Urban, A.E., Snyder, M. & Gerstein, M. (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, 21(6), 974–984. Available from: https://doi.org/10.1101/gr.114876.110

Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F. et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genetics*, 41(10), 1061–1067. Available from: https://doi.org/10.1038/ng.437

Bae, J.S., Cheong, H.S., Kim, L.H., NamGung, S., Park, T.J., Chun, J.Y. et al. (2010) Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics*, 11, 232. Available from: https://doi.org/10.1186/1471-2164-11-232

Bagnicka, E., Strzałkowska, N., Flisikowski, K., Szreder, T., Jóźwik, A., Prusak, B. et al. (2007) The polymorphism in the ß4-defensin gene and its association with production and somatic cell count in Holstein-Friesian cows. *Journal of Animal Breeding and Genetics*, 124(3), 150–156. Available from: https://doi.org/10.1111/j.1439-0388.2007.00649.x

Bagnicka, E., Strzałkowska, N., Szreder, T., Beata, P., Jóźwik, A., Kosciuczuk, E.M. et al. (2008) A/C polymorphism in the ß-4 defensin gene and its association with phenotypic and breeding values of milk production traits in Polish-Friesian cows. *Animal Science Papers and Reports*, 26(4), 239–250.

Banchereau, J., Briere, F., Caux, C., Davoust, J., Lebecque, S., Liu, Y.J. et al. (2000) Immunobiology of dendritic cells. *Annual Review of Immunology*, 18, 767–811. Available from: https://doi.org/10.1146/annurev.immunol.18.1.767

Becker, L., Kaase, M., Pfeifer, Y., Fuchs, S., Reuss, A., von Laer, A. et al. (2018) Genome-based analysis of Carbapenemase-producing Klebsiella pneumoniae isolates from German hospital patients, 2008-2014. *Antimicrob Resist Infect Control*, 7, 62. Available from: https://doi.org/10.1186/s13756-018-0352-y

Beckmann, J.S., Estivill, X. & Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews Genetics*, 8(8), 639–646. Available from: https://doi.org/10.1038/nrg2149

Ben Sassi, N., González-Recio, Ó., de Paz-del Río, R., Rodríguez-Ramilo, S.T. & Fernández, A.I. (2016) Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. *Journal of Dairy Science*, 99(8), 6371–6380. Available from: https://doi.org/10.3168/jds.2015-10487

Berton, M.P., Fonseca, L.F.S., Gimenez, D.F.J., Utembergue, B.L., Cesar, A.S.M., Coutinho, L.L. et al. (2016) Gene expression profile of intramuscular muscle in Nellore cattle with extreme values of fatty acid. *BMC Genomics*, 17(1), 972. Available from: https://doi.org/10.1186/s12864-016-3232-y

Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K. et al. (2012) Copy number variation

of individual cattle genomes using next-generation sequencing. *Genome Research*, 22, 778–790. Available from: https://doi.org/10.1101/gr.133967.111.778

Boussaha, M., Esquerré, D., Barbieri, J., Djari, A., Pinton, A., Letaief, R. et al. (2015) Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PLoS One*, 10(8), e0135931. Available from: https://doi.org/10.1371/journal.pone.0135931

Brandes, R., Arad, R. & Bar-Tana, J. (1995) Inducers of adipose conversion activate transcription promoted by a peroxisome proliferators response element in 3T3-L1 cells. *Biochemical Pharmacology*, 50(11), 1949–1951. Available from: https://doi.org/10.1016/0006-2952(95)02082-9

Brogden, K.A. (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, 3, 238–250. Available from: https://doi.org/10.1038/nrmicro1098

Campos, B.M., do Carmo, A.S., do Egito, A.A., da Mariante, A.S., do Albuquerque, M.S.M., de Gouveia, J.J.S. et al. (2017) Genetic diversity, population structure, and correlations between locally adapted zebu and taurine breeds in Brazil using SNP markers. *Tropical Animal Health and Production*, 49(8), 1677–1684. Available from: https://doi.org/10.1007/s11250-017-1376-7

Cicconardi, F., Chillemi, G., Tramontano, A., Marchitelli, C., Valentini, A., Ajmone-Marsan, P. et al. (2013) Massive screening of copy number population-scale variation in *Bos taurus* genome. *BMC Genomics*, 14, 124. Available from: https://doi.org/10.1186/1471-2164-14-124

Clop, A., Vidal, O. & Amills, M. (2012) Copy number variation in the genomes of domestic animals. *Animal Genetics*, 43(5), 503–517. Available from: https://doi.org/10.1111/j.1365-2052.2012.02317.x

Collard, B.L., Boettcher, P.J., Dekkers, J.C.M., Petitclerc, D. & Schaeffer, L.R. (2000) Relationships between energy balance and health traits of dairy cattle in early lactation. *Journal of Dairy Science*, 83, 2683–2690. Available from: https://doi.org/10.3168/jds.S0022-0302(00)75162-9

Conrad, B. & Antonarakis, S.E. (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annual Review of Genomics and Human Genetics*, 8, 17–35. Available from: https://doi.org/10.1146/annurev.genom.8.021307.110233

Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C. et al. (2010) Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nature Genetics*, 42(5), 385–391. Available from: https://doi.org/10.1038/ng.564

Coppage, K.H., Sun, X.M., Baker, R.S. & Clark, K.E. (2005) Expression of phosphodiesterase 5 in maternal and fetal sheep. *American Journal of Obstetrics and Gynecology*, 193, 1005–1010. Available from: https://doi.org/10.1016/j.ajog.2005.05.054

da Silva, J.M., Giachetto, P.F., da Silva, L.O., Cintra, L.C., Paiva, S.R., Yamagishi, M.E.B. et al. (2016) Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTL affecting production traits. *BMC Genomics*, 17, 454. Available from: https://doi.org/10.1186/s12864-016-2752-9

de Queiroz, S.A., Pelicioni, L.C., Silva, B.F., Sesana, J.C., Martins, M.I.E.G. & Sanches, A. (2005) Selection indices for a dual purpose breed Caracu. *Revista Brasileira de Zootecnia*, 34(3), 827–837. Available from: https://doi.org/10.1590/S1516-35982005000300014

Di Gerlando, R., Sardina, M.T., Tolone, M., Sutera, A.M., Mastrangelo, S. & Portolano, B. (2019) Genome-wide detection of copy-number variations in local cattle breeds. *Animal Production Science*, 59(5), 815–822. Available from: https://doi.org/10.1071/AN17603

Do, D.N., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X. et al. (2017) Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *Journal of Dairy Science*, 100(3), 1955–1970. Available from: https://doi.org/10.3168/jds.2016-11910

Duan, X., An, B., Du, L., Chang, T., Liang, M., Yang, B.G. et al. (2021) Genome-wide association analysis of growth curve parameters in chinese simmental beef cattle. *Animals*, 11, 192. Available from: https://doi.org/10.3390/ani11010192

Dupuis, M.C., Zhang, Z., Durkin, K., Charlier, C., Lekeux, P. & Georges, M. (2013) Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Animal Genetics*, 44(2), 206–208. Available from: https://doi.org/10.1111/j.1365-2052.2012.02373.x

Egito, A.A., Paiva, S.R., Albuquerque, M.S.M., Mariante, A.S., Almeida, L.D., Castro, S.R. et al. (2007) Microsatellite based genetic diversity and relationships among ten creole and commercial cattle breeds raised in Brazil. *BMC Genetics*, 8, 83. Available from: https://doi.org/10.1186/1471-2156-8-83

Estivill, X. & Armengol, L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genetics*, 3(10), 1787–1799. Available from: https://doi.org/10.1371/journal.pgen.0030190

Fadista, J., Thomsen, B., Holm, L.E. & Bendixen, C. (2010) Copy number variation in the bovine genome. *BMC Genomics*, 11, 284.

Falcone, F.H., Pritchard, D.I. & Gibbs, B.F. (2001) Do basophils play a role in immunity against parasites? *Trends in Parasitology*, 17(3), 126–129. Available from: https://doi.org/10.1016/S1471-4922(00)01846-8

Feuk, L., Carson, A.R. & Scherer, S.W. (2006) Structural variation in the human genome. *Nature Reviews Genetics*, 7, 85–97. Available from: https://doi.org/10.1038/nrg1767

Fontanesi, L., Beretti, F., Martelli, P.L., Colombo, M., Dall'Olio, S., Occidente, M. et al. (2011) A first comparative map of copy number variations in the sheep genome. *Genomics*, 97, 158–165. Available from: https://doi.org/10.1016/j.ygeno.2010.11.005

Fontanesi, L., Martelli, P.L., Beretti, F., Riggio, V., Dall'Olio, S., Colombo, M. et al. (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics*, 11, 639. Available from: https://doi.org/10.1186/1471-2164-11-639

Fries, R., Eggen, A. & Womack, J.E. (1993) The bovine genome map. *Mammalian Genome*, 4, 405–428. Available from: https://doi.org/10.1007/BF00296815

Frischknecht, M., Bapst, B., Seefried, F.R., Signer-Hasler, H., Garrick, D., Stricker, C. et al. (2017) Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics*, 18, 910. Available from: https://doi.org/10.1186/s12864-017-4308-z

Gao, Y., Jiang, J., Yang, S., Hou, Y., Liu, G.E., Zhang, S. et al. (2017) CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC Genomics*, 18, 265. Available from: https://doi.org/10.1186/s12864-017-3636-3

Gardner, J.L., Peters, A., Kearney, M.R., Joseph, L. & Heinsohn, R. (2011) Declining body size: a third universal response to warming? *Trends in Ecology and Evolution*, 26(6), 285–291. Available from: https://doi.org/10.1016/j.tree.2011.03.005

Gurao, A., Kashyap, S.K. & Singh, R. (2017) β-Defensins: an innate defense for bovine mastitis. *Veterinary World*, 10, 990–998. Available from: https://doi.org/10.14202/vetworld.2017.990-998

Haider, S., Ballester, B., Smedley, D., Zhang, J., Rice, P. & Kasprzyk, A. (2009) BioMart Central Portal–unified access to biological data. *Nucleic Acids Research*, 37, W23–W27. Available from: https://doi.org/10.1093/nar/gkp265

Henrichsen, C.N., Chaignat, E. & Reymond, A. (2009) Copy number variants, diseases and gene expression. *Human Molecular Genetics*, 18(R1), R1–R8. Available from: https://doi.org/10.1093/hmg/ddp011

Hoffmann, I. (2010) Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41(1), 32–46. Available from: https://doi.org/10.1111/j.1365-2052.2010.02043.x

Hou, Y., Liu, G.E., Bickhart, D.M., Cardone, M.F., Wang, K., Kim, E.S. et al. (2011) Genomic characteristics of cattle copy number variations. *BMC Genomics*, 12, 127. Available from: https://doi.org/10.1186/1471-2164-12-127

Hou, Y., Liu, G.E., Bickhart, D.M., Matukumalli, L.K., Li, C., Song, J. et al. (2012) Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Functional and Integrative Genomics*, 12(1), 81–92. Available from: https://doi.org/10.1007/s10142-011-0252-1

Houston, E.F., Connelley, T., Parsons, K., MacHugh, N.D. & Morrison, W.I. (2005) Analysis of T-cell receptor BV gene sequences in cattle reveals extensive duplication within the BV9 and BV20 subgroups. *Immunogenetics*, 57, 674–681. Available from: https://doi.org/10.1007/s00251-005-0040-y

Hu, Z.L., Park, C.A. & Reecy, J.M. (2016) Developmental progress and current status of the animal QTLdb. *Nucleic Acids Research*, 44, D827–D833. Available from: https://doi.org/10.1093/nar/gkv1233

Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), 1–13. Available from: https://doi.org/10.1093/nar/gkn923

Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. Available from: https://doi.org/10.1038/nprot.2008.211

Jiang, L., Jiang, J., Yang, J., Liu, X., Wang, J., Wang, H. et al. (2013) Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics*, 14, 131. Available from: https://doi.org/10.1186/1471-2164-14-131

Kijas, J.W., Barendse, W., Barris, W., Harrison, B., McCulloch, R., McWilliam, S. et al. (2011) Analysis of copy number variants in the cattle genome. *Gene*, 482, 73–77. Available from: https://doi.org/10.1016/j.gene.2011.04.011

Kongsuwan, K., Piper, E.K., Bagnall, N.H., Ryan, K., Moolhuijzen, P., Bellgard, M. et al. (2008) Identification of genes involved with tick infestation in *Bos taurus* and *Bos indicus*. *Developments in Biologicals*, 132, 77–88. Available from: https://doi.org/10.1159/000317146

Krzyzewski, J., Bagnicka, E., Strzałkowska, N., Pyzel, B., Zwierzchowski, L. & Jóźwik, A. (2008) Association between the polymorphism of bovine β4-defensin gene and milk traits in Holstein-Friesian cows as computed for standard (305 days) and the whole lactation. *Animal Science Papers and Reports*, 26(3), 191–198.

Labhart, P. (1999) Nonhomologous DNA end joining in cell-free systems. *European Journal of Biochemistry*, 265, 849–861. Available from: https://doi.org/10.1046/j.1432-1327.1999.00805.x

Lai, W.R., Johnson, M.D., Kucherlapati, R. & Park, P.J. (2005) Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics*, 21(19), 3763–3770. Available from: https://doi.org/10.1093/bioinformatics/bti611

Lara, M., Gama, L.T., Bufarah, G., Sereno, J.R.B., Celegato, E.M.L. & de Abreu, U.P. (2002) Genetic polymorphisms at the k-casein locus in Pantaneiro cattle. *Archivos de Zootecnia*, 51(193), 11.

Larson, J.H., Marron, B.M., Beever, J.E., Roe, B.A. & Lewin, H.A. (2006) Genomic organization and evolution of the ULBP genes in cattle. *BMC Genomics*, 7, 227. Available from: https://doi.org/10.1186/1471-2164-7-227

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R. et al. (2013) Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9(8), e1003118. Available from: https://doi.org/10.1371/journal.pcbi.1003118

Lehrer, R.I., Lichtenstein, A.K. & Ganz, T. (1993) Defensins: antimicrobial and cytotoxic peptides of mammalian cells. *Annual Review of Immunology*, 11, 105–128. Available from: https://doi.org/10.1146/annurev.iy.11.040193.000541

Lemos, M.V.A., Berton, M.P., de Camargo, G.M.F., Peripolli, E., de Oliveira Silva, R.M., Olivieri, B.F. et al. (2018) Copy number variation regions in Nellore cattle: evidences of environment adaptation. *Livestock Science*, 207, 51–58. Available from: https://doi.org/10.1016/j.livsci.2017.11.008

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. Available from: https://doi.org/10.1093/bioinformatics/btp352

Liu, G.E., Hou, Y., Zhu, B., Cardone, M.F., Jiang, L., Cellamare, A. et al. (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Research*, 20(5), 693–703. Available from: https://doi.org/10.1101/gr.105403.110

Liu, M., Fang, L., Liu, S., Pan, M.G., Seroussi, E., Cole, J.B. et al. (2019) Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. *BMC Genomics*, 20, 181. Available from: https://doi.org/10.1186/s12864-019-5552-1

Low, W.Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D.M. et al. (2019) Haplotype-resolved cattle genomes provide insights into structural variation and adaptation. *bioRxiv*, 720797. Available from: https://doi.org/10.1101/720797

Luenser, K. & Ludwig, A. (2005) Variability and evolution of bovine β-defensin genes. *Genes and Immunity*, 6, 115–122. Available from: https://doi.org/10.1038/sj.gene.6364153

Maggioli, M.F., Lobo, J.R., Fioravanti, M.C.S., Kipnis, A. & Junqueira-Kipnis, A.P. (2013) Cellular immune response of Curraleiro Pé-duro and Nellore calves following *Mycobacterium bovis*-BCG vaccination. *Pesquisa Veterinaria Brasileira*, 33(12), 1403–1408. Available from: https://doi.org/10.1590/S0100-736X2013001200002

Mariante, A. & Cavalcante, N. (2000) *Animais do descobrimento: raças domésticas da história do Brasil*. Brasília, DF: Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal. Brasília.

Mariante, A.S., Albuquerque, M.S.M., Egito, A.A. & McManus, C. (1999) Advances in the Brazilian animal genetic resources conservation programme. *Animal Genetic Resources Information*, 25, 107–121. Available from: https://doi.org/10.1017/S1014233900003497

Mariante, A.S., Albuquerque, M.S.M., Egito, A.A., McManus, C., Lopes, M.A. & Paiva, S.R. (2009) Present status of the conservation of livestock genetic resources in Brazil. *Livestock Science*, 120(3), 204–212. Available from: https://doi.org/10.1016/j.livsci.2008.07.007

Martin, J.M., Mead, J.I. & Barboza, P.S. (2018) Bison body size and climate change. *Ecology and Evolution*, 8(9), 4564–4574. Available from: https://doi.org/10.1002/ece3.4019

Martins, V.M.V., Veiga, T.F., Quadros, S., Ribeiro, J.A.R., Martins, E. & Cardoso, C.P. (2009) *Raça Crioula Lageana. O esteio do ontem, o labor do hoje e a oportunidade do amanhã*. Lages, SC: Editora Associação Brasileira dos Criadores da Raça Crioula Lageana (ABCCL). Lages.

Massari, S., Bellini, M., Ciccarese, S. & Antonacci, R. (2018) Overview of the germline and expressed repertoires of the TRB genes in *Sus scrofa*. *Frontiers in Immunology*, 9, 2526. Available from: https://doi.org/10.3389/fimmu.2018.02526

Mazza, M.C.M., da Silva Mazza, C.A., Sereno, J.R.B., Santos, S.A. & Oliveira, A. (1994) *Etnobiologia e conservação do bovino Pantaneiro*. Corumbá, MS: Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal. Corumbá.

Mazza, M.C.M., Mazza, C.A.S., Sereno, J.R.S., Santos, S.A. & Moura, A.C. (1992) Phenotypical characterization of Pantaneiro cattle in Brazil. *Archivos de Zootecnia*, 41, 477–484.

McCarroll, S.A. & Altshuler, D.M. (2007) Copy-number variation and association studies of human disease. *Nature Genetics*, 39(7), S37–S42. Available from: https://doi.org/10.1038/ng2080

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A. et al. (2016) The Ensembl variant effect predictor. *Genome Biology*, 17, 122. Available from: https://doi.org/10.1186/s13059-016-0974-4

Mcmanus, C., Ribeiro, R., Seixas, L. & de Melo, C.B. (2010) *A Raça Caracu*. Available from: animal.unb.br [Accessed 11th October 2021].

McTaggart, S.J., Obbard, D.J., Conlon, C. & Little, T.J. (2012) Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. *BMC Evolutionary Biology*, 12, 63. Available from: https://doi.org/10.1186/1471-2148-12-63

Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H. et al. (2018) Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Molecular Biology and Evolution*, 35(3), 688–699. Available from: https://doi.org/10.1093/molbev/msx322

Mitterer-Daltoé, M.L., Petry, F.C., Wille, D.F., Treptow, R.O., Martins, V.M.V. & Queiroz, M.I. (2012) Chemical and sensory characteristics of meat from Nellore and Crioulo Lageano breeds. *International Journal of Food Science and Technology*, 47(10), 2092–2100. Available from: https://doi.org/10.1111/j.1365-2621.2012.03075.x

Molnár, J., Nagy, T., Stéger, V., Tóth, G., Marincs, F. & Barta, E. (2014) Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. *BMC Genomics*, 15, 761. Available from: https://doi.org/10.1186/1471-2164-15-761

Mourão, G. & Medri, Í.M. (2007) Activity of a specialized insectivorous mammal (*Myrmecophaga tridactyla*) in the Pantanal of Brazil. *Journal of Zoology*, 271(2), 187–192. Available from: https://doi.org/10.1111/j.1469-7998.2006.00198.x

Nguyen, D.Q., Webber, C., Hehir-Kwa, J., Pfundt, R., Veltman, J. & Ponting, C.P. (2008) Reduced purifying selection prevails over positive selection in human copy number variant evolution. *Genome Research*, 18, 1711–1723. Available from: https://doi.org/10.1101/gr.077289.108

Nicolas, P. & Mor, A. (1995) Peptides as weapons against microorganisms in the chemical defense system of vertebrates. *Annual Review of Microbiology*, 49(1), 277–304. Available from: https://doi.org/10.1146/annurev.mi.49.100195.001425

Oliveira-Brochado, N.C., Chiodi, M.S., Souza-Cáceres, M.B., Abreu, U.G.P., Luz, D.F., Salla, L.E. et al. (2018) Dairy potential of Pantaneira breed upper region of pantanal of Mato Grosso Sul. *Arquivo Brasileiro de Medicina Veterinaria e Zootecnia*, 70(2), 644–648. Available from: https://doi.org/10.1590/1678-4162-9783

Olsen, L., Thum, E. & Rohner, N. (2021) Lipid metabolism in adaptation to extreme nutritional challenges. *Developmental Cell*, 56, 1417–1429. Available from: https://doi.org/10.1016/j.devcel.2021.02.024

Peripolli, E., Reimer, C., Ha, N.T., Geibel, J., Machado, M.A., Panetto, J.C.C. et al. (2020) Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. *BMC Genomics*, 21, 624. Available from: https://doi.org/10.1186/s12864-020-07035-6

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T. et al. (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology*, 29(6), 512–520. Available from: https://doi.org/10.1038/nbt.1852

Piper, E.K., Jackson, L.A., Bagnall, N.H., Kongsuwan, K.K., Lew, A.E. & Jonsson, N.N. (2008) Gene expression in the skin of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. *Veterinary Immunology and Immunopathology*, 126, 110–119. Available from: https://doi.org/10.1016/j.vetimm.2008.06.011

Por, F.D. (2012) *The Pantanal of Mato Grosso (Brazil): World's largest wetlands*. Berlin: Springer S.

Porto-Neto, L.R., Sonstegard, T.S., Liu, G.E., Bickhart, D.M., da Silva, M.V.B., Machado, M.A. et al. (2013) Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics*, 14, 876. Available from: https://doi.org/10.1186/1471-2164-14-876

Primo, A. (1992) El ganado bovino ibérico en las Américas: 500 años después. *Archivos de Zootecnia*, 41(154), 421–432.

Primo, A.T. (1986) Conservación de recursos genéticos animales em el Brasil. In: *Ganado bovino criollo*. Buenos Aires: Orientación Gráfica, pp. 224.

Prinsen, R.T.M.M., Rossoni, A., Gredler, B., Bieber, A., Bagnato, A. & Strillacci, M.G. (2017) A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. *Livestock Science*, 202, 7–12. Available from: https://doi.org/10.1016/j.livsci.2017.05.011

Purfield, D.C., Evans, R.D. & Berry, D.P. (2019) Reaffirmation of known major genes and the identification of novel candidate genes associated with carcass-related metrics based on whole genome sequence within a large multi-breed cattle population. *BMC Genomics*, 20, 720. Available from: https://doi.org/10.1186/s12864-019-6071-9

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. Available from: https://doi.org/10.1093/bioinformatics/btq033

Ramayo-Caldas, Y., Castelló, A., Pena, R.N., Alves, E., Mercadé, A., Souza, C.A. et al. (2010) Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics*, 11, 593. Available from: https://doi.org/10.1186/1471-2164-11-593

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D. et al. (2006) Global variation in copy number in the human genome. *Nature*, 444, 444–454. Available from: https://doi.org/10.1038/nature05329

Ricklefs, R.E. (1979) *Ecology*, 2nd edition. New York, NY: Chiron Press.

Sackton, T.B., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D. & Clark, A.G. (2007) Dynamic evolution of the innate immune system in Drosophila. *Nature Genetics*, 39, 1461–1468. Available from: https://doi.org/10.1038/ng.2007.60

Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. (2008) Regulatory T cells and immune tolerance. *Cell*, 133, 775–787. Available from: https://doi.org/10.1016/j.cell.2008.05.009

Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P. et al. (2007) Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, 39(7), S7–S15. Available from: https://doi.org/10.1038/ng2093

Selsted, M.E. & Ouellette, A.J. (2005) Mammalian defensins in the antimicrobial immune response. *Nature Immunology*, 6(6), 551–557. Available from: https://doi.org/10.1038/ni1206

Serrano, G.M., do Egito, A.A., McManus, C. & da Silva Mariante, A. (2004) Genetic diversity and population structure of Brazilian native bovine breeds. *Pesquisa Agropecuaria Brasileira*, 39(6), 543–549. Available from: https://doi.org/10.1128/JCM.42.1.461

Shaw, C.J. & Lupski, J.R. (2005) Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Human Genetics*, 116(1–2), 1–7. Available from: https://doi.org/10.1007/s00439-004-1204-9

Stockdale, C.R. (2001) Body condition at calving and the performance of dairy cows in early lactation under Australian conditions: a review. *Australian Journal of Experimental Agriculture*, 41(6), 823–839. Available from: https://doi.org/10.1071/EA01023

Stothard, P., Choi, J.W., Basu, U., Sumner-Thomson, J.M., Meng, Y., Liao, X. et al. (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*, 12, 559. Available from: https://doi.org/10.1186/1471-2164-12-559

Subkhangulova, A., Malik, A.R., Hermey, G., Popp, O., Dittmar, G., Rathjen, T. et al. (2018) SORCS 1 and SORCS 3 control energy balance and orexigenic peptide production. *EMBO Reports*,

19(4), e44810. Available from: https://doi.org/10.15252/embr.201744810

Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A. et al. (2010) Diversity of human copy number variation and multicopy genes. *Science*, 330(6004), 641–646. Available from: https://doi.org/10.1126/science.1197005.Diversity

Takeshima, S.N. & Aida, Y. (2006) Structure, function and disease susceptibility of the bovine major histocompatibility complex. *Animal Science Journal*, 77, 138–150. Available from: https://doi.org/10.1111/j.1740-0929.2006.00332.x

Tanaka, K. (2009) The proteasome: overview of structure and functions. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 85, 12–36. Available from: https://doi.org/10.2183/pjab.85.12

Tarver, A.P., Clark, D.P., Diamond, G., Russell, J.P., Erdjument-Bromage, H., Tempst, P. et al. (1998) Enteric β-defensin: molecular cloning and characterization of a gene with inducible intestinal epithelial cell expression associated with *Cryptosporidium parvum* infection. *Infection and Immunity*, 66(3), 1045–1056. Available from: https://doi.org/10.1128/iai.66.3.1045-1056.1998

Tian, M., Wang, Y., Gu, X., Feng, C., Fang, S., Hu, X. et al. (2013) Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. *BMC Genomics*, 14, 262. Available from: https://doi.org/10.1186/1471-2164-14-262

Toffolatti, L., Cardazzo, B., Nobile, C., Danieli, G.A., Gualandi, F., Muntoni, F. et al. (2002) Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics*, 80(5), 523–530. Available from: https://doi.org/10.1006/geno.2002.6861

Upadhyay, M., da Silva, V.H., Megens, H.J., Visker, M.H.P.W., Ajmone-Marsan, P., Bâlteanu, V.A. et al. (2017) Distribution and functionality of copy number variation across European cattle populations. *Frontiers in Genetics*, 8, 108. Available from: https://doi.org/10.3389/fgene.2017.00108

Van Gent, D.C. & Van Der Burg, M. (2007) Non-homologous end-joining, a sticky affair. *Oncogene*, 26, 7731–7740. Available from: https://doi.org/10.1038/sj.onc.1210871

Wang, M.D., Dzama, K., Hefer, C.A. & Muchadeyi, F.C. (2015) Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics*, 16, 894. Available from: https://doi.org/10.1186/s12864-015-2122-z

Wang, M.D., Dzama, K., Rees, D.J.G. & Muchadeyi, F.C. (2016) Tropically adapted cattle of Africa: perspectives on potential role of copy number variations. *Animal Genetics*, 47, 154–164. Available from: https://doi.org/10.1111/age.12391

Weller, S.C. (2005) Correspondence analysis. In: Armitage, P. & Colton, T. (Eds.) *Encyclopedia of biostatistics*. Chichester: John Wiley & Sons, Ltd. Available from: https://doi.org/10.1002/0470011815.b2a13015

Wikel, S.K. (1996) Host immunity to ticks. *Annual Review of Entomology*, 41, 1–22. Available from: https://doi.org/10.1146/annurev.ento.41.1.1

Winchester, L., Yau, C. & Ragoussis, J. (2009) Comparing CNV detection methods for SNP arrays. *Briefings in Functional Genomics and Proteomics*, 8(5), 353–366. Available from: https://doi.org/10.1093/bfgp/elp017

Wu, Y., Fan, H., Jing, S., Xia, J., Chen, Y., Zhang, L. et al. (2015) A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. *Animal Genetics*, 46(3), 289–298. Available from: https://doi.org/10.1111/age.12288

Yang, D., Chertov, O., Bykovskaia, S.N., Chen, Q., Buffo, M.J., Shogan, J. et al. (1999) β-Defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science*, 286(5439), 525–528. Available from: https://doi.org/10.1126/science.286.5439.525

Yang, L., Xu, L., Zhu, B., Niu, H., Zhang, W., Miao, J. et al. (2017) Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese cattle. *Scientific Reports*, 7, 14299. Available from: https://doi.org/10.1038/s41598-017-14768-0

Yau, C. & Holmes, C.C. (2008) CNV discovery using SNP genotyping arrays. *Cytogenetic and Genome Research*, 123, 307–312. Available from: https://doi.org/10.1159/000184722

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F. & Bendixen, C. (2011) Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*, 12, 557. Available from: https://doi.org/10.1186/1471-2164-12-557

Zhang, F., Gu, W., Hurles, M.E. & Lupski, J.R. (2009) Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, 10, 451–481. Available from: https://doi.org/10.1146/annurev.genom.9.081307.164217

Zhang, Q., Ma, Y., Wang, X., Zhang, Y. & Zhao, X. (2015) Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array. *Molecular Genetics and Genomics*, 290, 319–327. Available from: https://doi.org/10.1007/s00438-014-0923-4

Zhou, W., Liu, R., Zhang, J., Zheng, M., Li, P., Chang, G. et al. (2014) A genome-wide detection of copy number variation using SNP genotyping arrays in Beijing-You chickens. *Genetica*, 142(5), 441–450. Available from: https://doi.org/10.1007/s10709-014-9788-z

Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay, E.H., Bickhart, D.M., Sonstegard, T.S. et al. (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. Available from: https://doi.org/10.1038/srep29219

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Peripolli, E., Stafuzza, N.B., Machado, M.A., do Carmo Panetto, J.C., do Egito, A.A., Baldi, F. et al. (2023) Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome re-sequencing data. *Animal Genetics*, 00, 1–17. Available from: https://doi.org/10.1111/age.13298