Uncovering structural variants in Creole cattle from Guadeloupe and their impact on environmental adaptation through whole genome sequencing

Slim Ben-Jemaa, Mekki Boussaha, Nathalie Mandonnet, Philippe Bardou, Michel Naves

HAL Id: hal-04679025
https://hal.inrae.fr/hal-04679025v1

RESEARCH ARTICLE

Uncovering structural variants in Creole cattle from Guadeloupe and their impact on environmental adaptation through whole genome sequencing

Slim Ben-Jemaa[1,2]*, Mekki Boussaha[3], Nathalie Mandonnet[1], Philippe Bardou[4,5], Michel Naves[1]*

1 INRAE, ASSET, 97170, Petit-Bourg, France, 2 Institut National de la Recherche Agronomique de Tunisie, Laboratoire des Productions Animales et Fourragères, Université de Carthage, 2049, Ariana, Tunisia, 3 Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France, 4 GenPhySE, Université de Toulouse, INRA, Ecole Nationale Vétérinaire de Toulouse (ENVT), 31320, Castanet-Tolosan, France, 5 Sigenae, INRAE, 31320, Castanet-Tolosan, France

* michel.naves@inrae.fr (MN); slim.ben-jemaa@inrae.fr (SB-J)

Abstract

Structural variants play an important role in evolutionary processes. Besides, they constitute a large source of inter individual genetic variation that might represent a major factor in the aetiology of complex, multifactorial traits. Their importance in adaptation is becoming increasingly evident in literature. Yet, the characterization of the genomic landscape of structural variants in local breeds remains scarce to date. Herein, we investigate patterns and gene annotation of structural variants in the Creole cattle from Guadeloupe breed using whole genome sequences from 23 bulls representative of the population. In total, we detected 32821 ascertained SV defining 15258 regions, representing ~ 17% of the Creole cattle genome. Among these, 6639 regions have not been previously reported in the Database of Genomic Variants archive. Average number of structural variants detected per individual in the studied population is in the same order of magnitude of that observed in indicine populations and higher than that reported in taurine breeds. We observe an important within-individual variability where approximately half of the detected structural variants have low frequency (MAF < 0.25). Most of the detected structural variants (55%) occurred in intergenic regions. Genic structural variants overlapped with 7793 genes and the predicted effect of most of them is ranked as "modifier". Among the structural variants that were predicted to have a high functional impact on the protein, a 5.5 Kb in length, highly frequent deletion on chromosome 2, affects *ALPI*, a gene associated with the interaction between gut microbiota and host immune system. The 6639 newly identified structural variants regions include three deletions and three duplications shared by more than 80% of individuals that are significantly enriched for genes related to tRNA threonylcarbamoyladenosine metabolic process, important for temperature adaptation in thermophilic organisms, therefore suggesting a potential role in the thermotolerance of Creole cattle from Guadeloupe cattle to tropical climate. Overall, highly frequent structural variants that are specific to the Creole cattle population encompass olfactory receptor and immunity genes as well as

genes involved in muscle tone, muscle development and contraction. Beyond mapping and characterizing structural variants in the Creole cattle from Guadeloupe breed, this study provides valuable information for a better understanding of the potential role of chromosomal rearrangements in adaptive traits in cattle.

Introduction

Structural variants (SV) are large DNA rearrangements (>50 bp in length) affecting an individual's genome. They can be balanced and show no specific loss or gain of genetic material, such as inversions of a genetic fragment or translocations of a stretch of DNA within or between chromosomes, or they can be unbalanced, where a part of the genome is lost (insertions/deletions) or duplicated (duplications). In the latter case, structural variants are also termed copy number variation (CNV) [1]. Structural variants are ubiquitous and affect a greater fraction of the genome than single nucleotide polymorphisms (SNPs) [2]. They have been extensively studied in humans where they have been shown to constitute potent phenotypic modifiers that act through multiple mechanisms, such as altering gene dosage, disrupting regulatory elements, generating fusion proteins or unmasking of recessive alleles, thus causing several human disorders [3]. Structural variants are also considered as an important driver of evolution that may enable rapid adaptation to environmental stressors in animals and plants [4–6]. SV are poorly characterized in livestock [7]. Yet, several studies highlighted their influence on several phenotypic traits [8–10] in several domestic animals. In cattle, SV are responsible for variation in coat colour [11], and several complex traits, including milk production, fertility, and other traits [8,12–14]. The majority of SV studies carried out in cattle were SNP-array based [15]. It has only been recently that studies based on second- and third-generation sequencing technologies begin to emerge for local breeds [16–18]. Regardless of the sequencing technology used, a significant proportion of the identified SV in cattle was shown to be breed-specific, suggesting a potential association with differences in adaptation, health, and production traits [19–21]. Breed-specific SV can potentially store important information on the genomic architecture of adaptive traits. This is particularly true for small local breeds that have been exposed to various selective pressures in a given environment.

Identifying SV associated with environmental adaptation would be of utmost importance in the climate change context, particularly for cattle where high-producing breeds, more sensitive to heat stress and less adapted to emerging pathogens are replacing locally adapted populations.

Livestock species have been introduced in Latin America and the Caribbean after the discovery of the new world by Colombus, in the 15th century. The first specimen introduced came from the Iberian Peninsula, but afterwards, between the 16th and 18th century, animals from other origins were brought, following the complex history of colonization and human settlement in the region [22]. These complex migratory and admixture events, combined with natural selection and traditional usage, led to the constitution of the different Creole breeds, presenting specific features according to their location. In the Caribbean, under the French influence, important migrations followed the slave-trade route from Western Africa, with probably both small shorthorn taurine cattle and indicine admixed populations, as attested by historical data [23].

Creole cattle in the Guadeloupe island (GUA) is an admixed breed resulting from a three-way admixture between African taurine, European taurine and zebu (Fig 1). GUA individuals

Fig 1. Cow and sire representative of the Creole cattle from Guadeloupe. The hump and the dewlap in the males are well-developed.

https://doi.org/10.1371/journal.pone.0309411.g001

show good production and reproduction abilities under warm climate and harsh conditions and are known to be resistant to several endemic tick-borne diseases in the island [24]. Recently, through whole-genome sequence analysis of 23 GUA individuals, we provided a detailed examination of genetic variation and we identified several candidate regions potentially associated with specific adaptive features in the GUA genome [25]. In order to capture the different axes of diversity in GUA population, we herein report for the first time, a genome-wide characterization of structural variants derived from whole-genome resequencing data in these 23 GUA genomes. We also highlighted the potential role of the identified chromosomal structural variation in local adaptation of GUA population.

Materials and methods

Animal ethics statement

Blood collection was done according to good practices recommended for identification of sires for paternity checking in France. This study was approved by the scientific committee of the Metaprogramme SELGEN of INRAE.

Sample information and genome sequencing

Twenty three Creole bulls representative of the INRA nucleus herd in Guadeloupe were selected for the purpose of this study. Ten out of them are sons of founder sires of the experimental flock while the remaining thirteen animals are unrelated sires (based on genealogical records and sampling sites) chosen from the local stock of Guadeloupe. Genomic DNA was extracted from whole-blood and semen samples collected between 1995 and 2015. Paired-end libraries with insert size of 500 bp were constructed for each individual and sequenced using

the HiSeq 3000 platform (Illumina) in the Genome et Transcriptome (GeT) GénoToul plat-form (Toulouse, France), following the manufacturer's protocol.

Sequence alignment and SV discovery

Quality control of raw sequence reads was performed using the fastQC software v.0.11.7 [26]. Trimmomatic-0.36 [27] was used to remove Illumina adapter sequences, low-quality bases and artefact sequences. Filtered sequences were then mapped against the bovine reference genome (ARS-UCD1.2) using the Burrows-Wheeler Alignment tool (bwa mem v.0.7.17) [28] with default parameters. The resulting SAM files were then converted to BAM format, sorted, and indexed using SAMtools [29]. PCR duplicates were removed using the MarkDuplicates tool from Picard version 1.88 (http://broadinstitute.github.io/picard). Only properly paired reads with a mapping quality of at least 30 were kept. Local realignment was performed using two GATK (Genome Analysis Toolkit) version -3.8-1-0-gf15c1c3ef modules, RealignerTargetCreator and IndelRealigner. We used three SV callers, Pindel (v.0.2.5), LUMPY (v.0.2.13) and DELLY (v.0.7.8). We developed a custom pipeline combining the detection results of the three aforementioned SV-finding algorithms. We retained only SV identified by at least two callers and having a length between 50 bp and 5 Mb because structural variants identification tools relying solely on a single principle were shown to generate many false positive calls [30,31]. According to their frequency within the GUA sample, we have defined four SV categories: singletons (detected in one individual), low frequency (detected in less than 25% of individuals), common (detected in more than 25% and less than 80% of individuals) and highly frequent (detected in more than 80% of individuals). The R package karyoploteR [32] was used to display SV distribution across 30 chromosomes (29 autosomes and the X chromosome). To identify novel structural variations that had not been discovered so far, the genomic regions defining the SV detected in the present study were intersected with known structural variations reported in cattle from Database of Genomic Variants archive (DGVa) using the function 'intersect' from BEDtools [33]. For this purpose, we defined non-overlapping SV regions for the three types of SV and we compared these regions to the 9277 SV regions reported in the DGVa database.

Functional impact of structural variants

Ensembl Variant Effect Predictor (http://www.ensembl.org/info/docs/tools/vep/index.html) (VEP) was used to provide a prediction for each transcript with which the common and highly frequent SV overlap (those whose MAF > 0.25). VEP was used to determine the location of the SV (e.g. intronic, intergenic, upstream of a transcript, in coding sequences, in regulatory regions) as well as the predicted effect of SV on the protein (e.g. stop lost, frameshift). VEP also provides an impact rating (high, moderate, low, and modifier) indicating the severity of the consequences.

Functional enrichment analysis was performed using the online tool DAVID version 6.8 (Database for Annotation, Visualization and Integrated Discovery, https://david.ncifcrf.gov/). DAVID uses thousands of annotation terms in several annotation categories, such as Gene Ontology (GO), Biological Process, GO Molecular Function and InterPro Domains. An adjusted Benjamini-corrected p-value of 0.05 was used as the criterion for statistical significance of over-enrichment of genes in one of the categories.

Results and discussion

Sequencing, SV discovery, and validation

A total of 7,543,644,154 clean reads were generated after sequencing the complete genome of the 23 samples, representing an average depth of ~16.35 fold (min depth = 9.3; max

Table 1. Performance of the three SV callers in detecting different SV types within the 23 Creole cattle individuals.

SV caller	SV type	Count
Pindel	DEL	1,089,575
DELLY	DEL	201,233
LUMPY	DEL	147,582
Pindel	DUP	69,763
DELLY	DUP	35,385
LUMPY	DUP	10,269
Pindel	INV	30,299
DELLY	INV	49,726
LUMPY	INV	1,342

https://doi.org/10.1371/journal.pone.0309411.t001

depth = 23.77). The average alignment rate against the *Bos Taurus* reference genome (ARS-UCD1.2), for the different individuals varied from 94.93% to 96.51% with an average of 95,7% (S1 Table). Three classes of SV: deletions (DEL), duplications (DUP) and inversions (INV) with a size ranging between 50 bp and 5 Mb were detected. Among the three callers, Pindel exhibited the highest detection sensitivity for deletions and duplications while DELLY identified much more inversions. LUMPY, for its part, called a lower number of SVs for all the three types of structural variants (Table 1). In total, 69% and 87% of the deletions (those at a minimum overlap identity of 90%) identified by Pindel and LUMPY, respectively were also detected by DELLY but this latter detected only 36% and 49% of the deletions, identified by Pindel and LUMPY, respectively. A high percentage, 94%, of the duplications identified by LUMPY were also identified by DELLY, but only 42% of those identified by DELLY were also identified by LUMPY. Also, we found that 59% of the duplications detected by DELLY were identified by Pindel and that 69% of those identified by Pindel were revealed by DELLY. In total, 92% of the inversions detected by Pindel were also identified by DELLY but only 53% of the inversions identified by DELLY were detected by Pindel.

In total, we identified 32,821 ascertained SV defining 15,258 regions with at least 2 software from the 23 GUA animals. Although combining multiple callers is recommended for a higher detection accuracy of structural variants [34], it should be outlined that SV detection in the present study is likely to be altered by several factors that are related to the sequencing technology used in the present study (short read sequencing) and to the threefold ancestry of the Creole cattle from Guadeloupe. For instance, insertions longer than short reads are expected to be easily missed because they cannot align correctly with the reference genome [35]. Likewise, the use of linear reference genomes causes reference allele bias that affects genetic variation detection tools [36]. Hence, we expect that the detection of SVs that are from African taurine and indicine origin in the genome of GUA, woud be affected when aligning the mapped reads to the Hereford reference genome assembly ARS-UCD1.2. Clearly, using a personalized reference genome of GUA should improve reliability of structural variation calls.

SV distribution and statistics

In accordance with previous studies [37–39], we observe an abundance of deletions (76% of the total SV detected in the present study) compared to duplications and inversions. On the other hand, deletions have a similar range of total length to duplications (Table 2) which is due to the fact that most of the identified deletions (81.56%) were <5 kb in length, whereas more than half (~ 55%) of duplications and inversions are of large size (10 Kb <size<5 Mb)

Table 2. Statistics of the SV number detected by at least two callers for the 23 GUA animals.

	Number of SV	Total length (Mb)
DEL	24981	792.46
DUP	5071	767.56
INV	2769	567.23
Total	32821	2127.25

https://doi.org/10.1371/journal.pone.0309411.t002

(Fig 2A). This agrees with previous findings reporting a possible correlation between SV type and size.

Approximately half (~ 54%) of all detected SV have low frequency (detected in less than 6 individuals; Fig 2B) and covered 1.214 Gb of the chromosomes of the ARS-UCD1.2 assembly. About a quarter of these were detected in a single individual. This is consistent with previous studies indicating that high SV diversity exists among different cattle individuals [40,41]. Accordingly, we also observe an important within-individual variation of high-confidence SV. Their number ranged between 2092 and 4002 (average = 3051 ± 455) for deletions < 1 Kb, 1793 and 3009 (2362 ± 308) for deletions > 1 Kb, between 84 and 187 (average = 131 ± 28) for duplications < 1 Kb, 282 and 736 (average = 488 ± 114) for duplications > 1 Kb, and between 61 and 115 (average = 86 ± 16) for inversions < 1 Kb, 155 and 318 (average = 242 ± 40) for inversions > 1 Kb (Fig 2C). We observe, on average, 6360 SV per GUA individual which is in the same order of magnitude of African and Indian zebu but goes well beyond the values previously reported in European and African taurine breeds [38]. Part of the differences between GUA on one hand and taurine European and African populations on the other hand may be due to population structure. Indeed, the three-way admixture of the GUA genome [24] of which more than one third has an indicine origin is likely to be behind the high number of observed SV. Several studies have reported a higher SV in indicine than in taurine breeds which is consistent with the known breed divergence and history [42,43]. The origin of the indicine ancestry in GUA population appears controversial. Some historical evidences [23] relate introductions of West African cattle in Guadeloupe between the 16th and 18th century, and it is possible that African zebu as well as African taurine have been introduced. The alternate possibility of a recent introduction of Indian zebu in Guadeloupe at the end of the 19th century (as in the other regions of America) has not been documented [24]

A small portion (7.6%) of the detected SV is present in more than 80% of the GUA individuals (Fig 2B). Chromosomal distribution of common and highly frequent SV regions and having a size > 1 Kb was not uniform and varied according to the SV type. Chromosome 25 has the lowest number of deletions and inversions (47 and 6 deletions and inversions, respectively) and chromosome 26 has the lowest number of duplications (7). By contrast, BTA5, BTA18 and BTA3 showed the highest number of deletions, duplications and inversions, respectively (S2 Table).

Overall, common and highly frequent SV (MAF > 0.25) detected in the present study covered a total length of ~ 446 Mb which corresponds to almost 17% of the ARS-UCD1.2 assembly. Analysis of the distribution of SV regions across chromosomes showed substantial variation depending among others on SV type (Fig 3). Chromosome-wide SV coverage along each chromosome varies from 6.36% on chromosome 25 to 39.4% on chromosome 28 (S1 Fig) and S3 Table) and was not correlated to the number of SV regions (S3 Table). BTX is the most densely covered with SV in terms of Megabases (32 Mb) with inversions representing 70% of its whole SV length (S3 Table). Similar finding was reported by [44]. They explained this by difficulties when mapping SV on the X chromosome especially in males where effective coverage is halved. Another possible explanation for the observed higher number of SV observed on
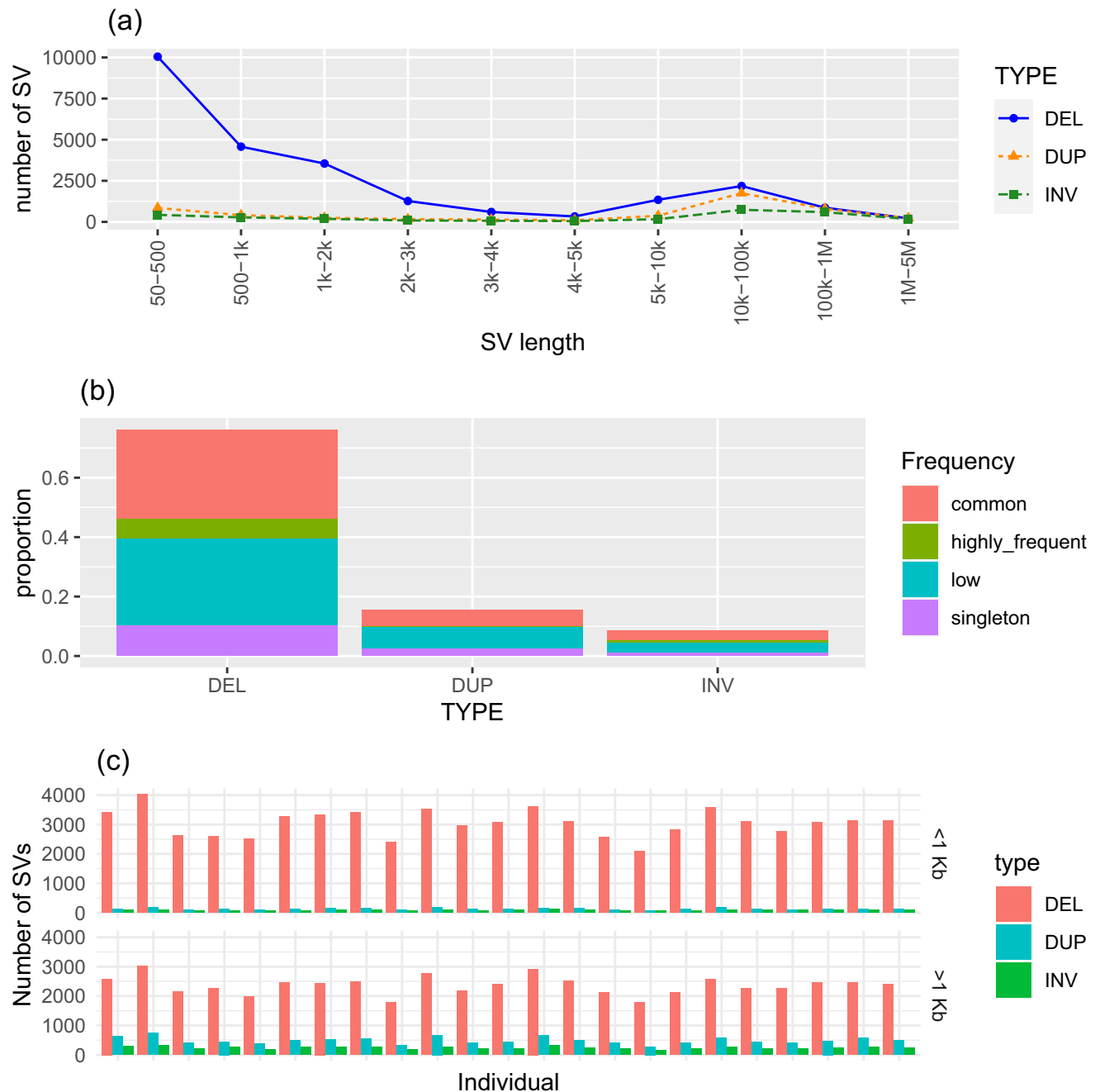
Fig 2. Statistics and individual distribution of structural variations (SVs) in the Creole cattle from Guadeloupe sample. (a) distribution of length for each SV type, (b) frequency of SVs by type. Four categories are defined: Singletons (detected in one individual), low frequency (detected in less than 25% of individuals), common (detected in more than 25% and less than 80% of individuals) and highly frequent (detected in more than 80% of individuals), (c) SV distribution among the 23 sequenced animals. The upper part shows SV < 1 Kb, while the bottom part shows SV lengths > 1 Kb.

the X chromosome in the 23 Creole cattle bulls is likely related to the high male specific contribution of the X-chromosome to individual global recombination rate, previously reported in several cattle breeds [45]. It is therefore reasonable to assume that chromosomes with higher recombination rates are more likely to show more SVs since these are generated by different recombination mechanisms.
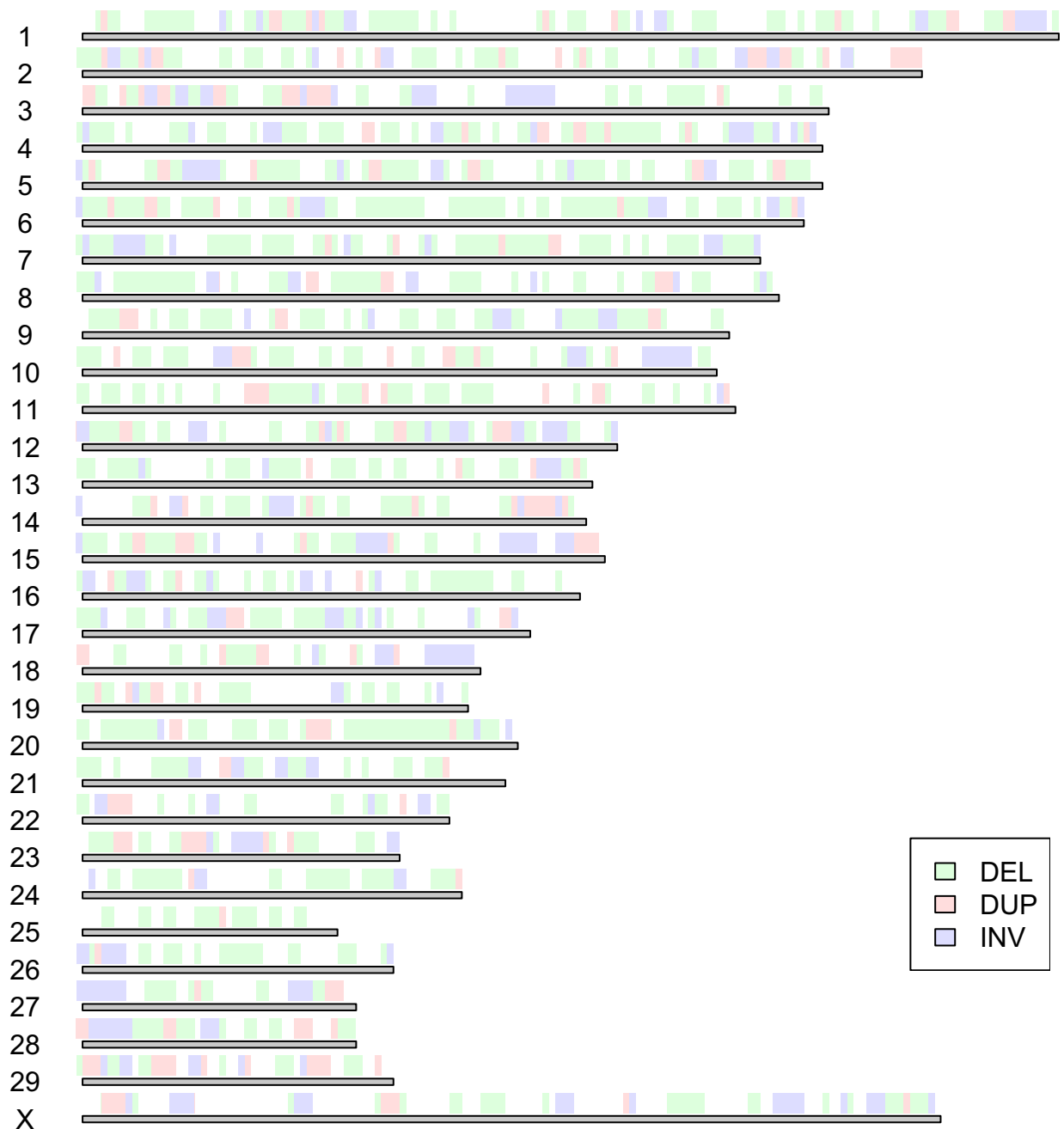
Fig 3. Chromosomal distribution of nonredundant SV regions with MAF > 0.25 and size > 1 Kb.

We identified 6065, 464 and 343 new deletion, duplication and inversion regions, respectively that have not been previously reported in the DGVa database. Together, the three types SV defined 6638 nonredundant regions that have not been previously reported in DGVa (S4 Table). Focusing on SV > 1 Kb, 1967, 238 and 221 new deletion, duplication and inversion regions, were detected respectively (S2 Fig). Among these, three deletions, located on chromosomes 12, 16 and 26 and three duplications, located on chromosomes 7, 28 and X, are highly frequent in the GUA sample (S5 Table).

Annotation of SV

Structural variants that occur in genes can alter gene expression either by changing gene dosage or interrupting coding sequences, or disturbing long-range gene regulation which could broadly influence phenotypes. To better predict the downstream effect of the detected SV on protein function, we used the Ensembl VEP tool [46]. To get potential important insights into population-level effect of SV on genes with adaptive functions in cattle breeds raised under tropical environment, we merged the common and highly frequent SV into a set of nonredundant 9734 SV regions among which 55% were intergenic (Fig 4). SV regions occurring in genic regions overlapped with 7793 genes and 12,922 transcripts. According to VEP, almost all of the detected SV were assigned to the "modifier" impact category which is not surprising since most of these SV were located within introns (Fig 4). Only three deletions, located on BTA8 (at positions: 9,754,093–9,839,766), BTA10 (at position:101,090,067–101,094,332 bp) and BTA15 (at position: 45,917,286–45,918,286 bp), were classified as having a high (disruptive) impact on the protein function. Three genes are affected by these SV: *HMBOX1*, *FOXN3* and a novel cattle gene: *ENSBTAG00000027525*. The first two genes are transcription factors. *HMBOX1* is a transcriptional repressor that negatively regulates IFN-γ in natural killer cells [47]. While IFN-γ production is momentarily abrogated following intense exercise which provides a window for invasion of pathogens [48], the deletion in *HMBOX1* might be regarded as
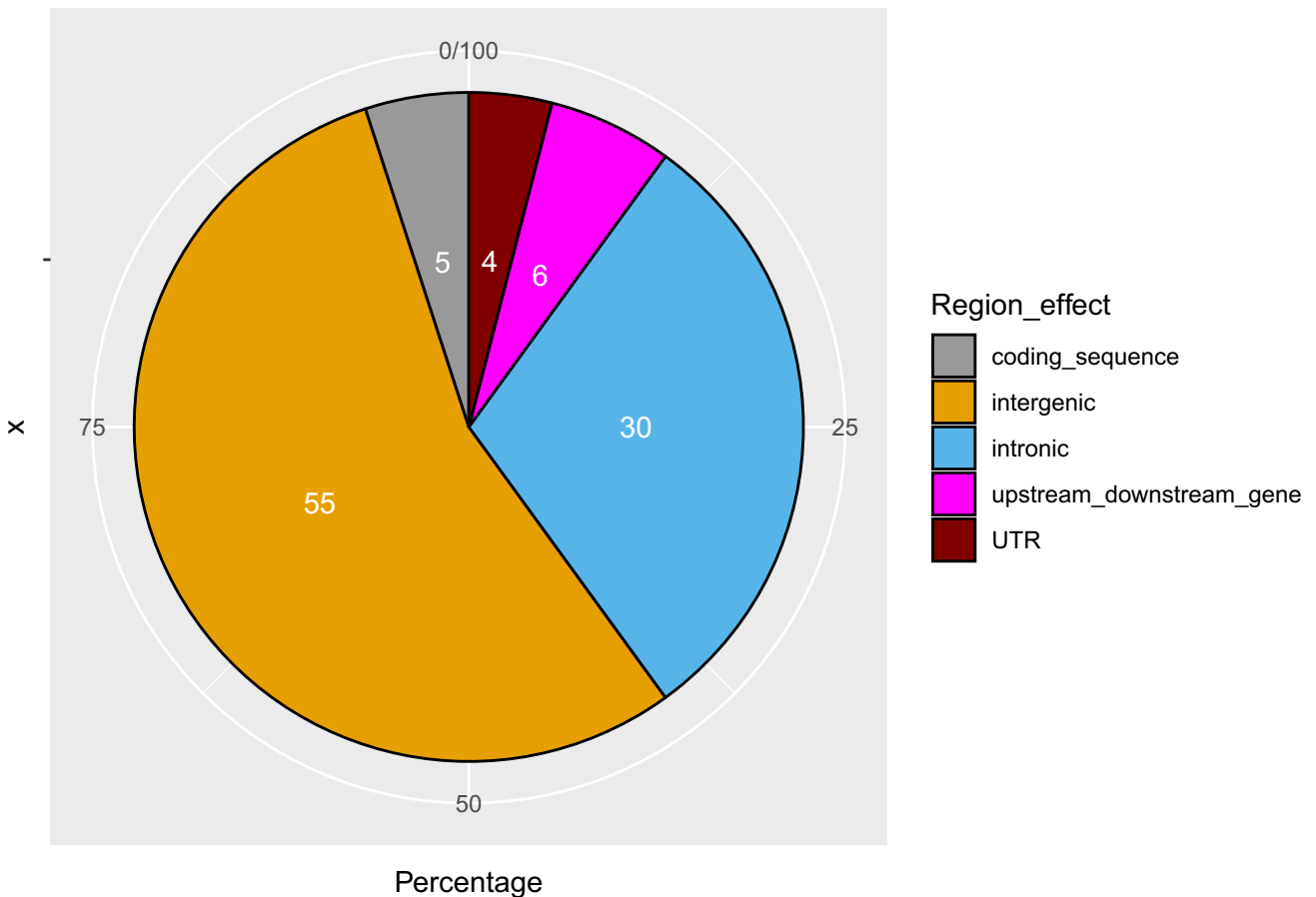

Fig 4. Variant effect predictor output from the genomic regions defined by nonredundant SV with size > 1 Kb.

https://doi.org/10.1371/journal.pone.0309411.g004

a compensation mechanism that substitutes the downregulation of *IFN-γ* during prolonged exhausting exercise. Creole cattle from Guadeloupe are mainly used for a draught of sugar cane carts or ploughing [49] which would weaken their immune system for large parts of the day. *HMBOX1* might have also a protective role against splenomegaly and anaemia, which are prominent features of trypanosomiasis in cattle [50]. Unlike *HMBOX1*, the adaptive interpretation of *FOXN3* deletion (observed in nearly 83% of our GUA sample), is less obvious because this gene is involved in a variety of physiological processes ranging from cell proliferation, apoptosis and pathogenesis in human cancer [51] to craniofacial development and fasting blood glucose and glucagon, in other non-human species [52–54]. In Angus cattle, it has been associated with chest width and skeletal development [55]. Further investigations of the functional effects of partial deletions *FOXN3* on adaptive features of cattle are needed.

We also checked if there is any effect of SV frequency on their functional consequences by comparing common and highly frequent SV on one hand and low frequency SV on the other hand (these are defined as SV detected in less than four individuals). We found that SV of the second category tend to be in coding sequences more often than the first category (738 Vs 487, χ-squared = 51.712, p-value = 6.425 x 10^{-13}).

We then generated a gene list including 579 genes that overlap with the genomic regions specific to GUA population (those that were not reported in DGVa) and encompassing common and highly frequent deletions and duplications. These genes were annotated according to DAVID Bioinformatics resources (https://david.ncifcrf.gov/). We found significant enrichment of the functional category 'UP_SEQ_FEATURE: DOMAIN:C2' (Benjamini-corrected P-value = 0.034, n = 13). Calcium transport was the most enriched biological process (BP) term (Benjamini-corrected P-value = 0.0011, n = 10) (S6 Table). C2 domains are widespread and conserved motifs that often serve as Ca2+-binding modules. Single and multiple copies of C2 domains have been identified in a growing number of eukaryotic signalling proteins that interact with cellular membranes and mediate a broad array of critical intracellular processes, including membrane trafficking, the generation of lipid-second messengers, activation of GTPases, and the control of protein phosphorylation [56]. In total, thirteen C2 domain genes located on 11 different chromosomes were impacted by SV. Among these, eight genes: *CPNE4, PKN2, PIK3C2G, UNC13C, DYSF, PLCB1, SYT9, SMURF2* are affected by frequent deletions (affecting more than 65% of our GUA sample). In ruminants, several of these genes have previously been associated with growth traits, such as *CPNE4 or DYSF* [57,58] or with fat metabolism, such as *PLCB1* [59]. Interestingly, the differentiation of some of these genes between taurine and indicine cattle have also been discussed (*CPNE4, DYSF, PIK3C2G*) [57,58,60]. Three of these genes (*DYSF, PLCB1, SMURF2*) are also known to regulate many aspects of the immune system response to pathogens. Similar to *HMBOX1*, *SMURF2* and *DYSF* negatively regulate some aspects of immune response. *SMURF2* gene is an essential negative regulator of TGF-β signalling and plays a role in the vascular inflammatory response in the presence of hypoxia in endothelial cells [61] while *DYSF* negatively regulates phagocytosis (GO:0050765). Finally, *PKN2* appears essential for embryogenesis in mouse [62], and its loss causes severe cardiovascular and morphogenetic abnormalities. More research would be necessary to identify the impact of these deletions in cattle, in particular in GUA breeds.

We took a closer look at the three deletions and the three duplications that were not previously reported in the DGVa database and which are highly frequent in our GUA sample (S5 Table). These six CNV regions encompass 135 genes, most of them (67 genes) are located on chromosome X. We have focused primarily on genes for which there is evidence for local adaptation in GUA cattle. In this regard, we find that duplications on BTA28 encompass two genes, *ras homolog family member U (RHOU)* and *actin alpha 1, skeletal muscle (ACTA1)* that play a key role in muscle development and contraction [63–65]. *RHOU* also regulates cell-

adhesion molecules during cardiac morphogenesis [66]. Creole cattle from Guadeloupe have often been used as a major labour force in sugarcane fields and are known for their draft endurance. Duplications in these genes might be a hint that allows physiological adaptation of GUA population to such strenuous activities. In humans, it has been shown that SV occurring in genes expressed in muscle or heart contribute to the variation of endurance capacity [67]. CNV observed in *RHOU* were also reported as involved in mammary development in Dairy Gir cattle [68]. Deletions on BTA26, overlaps with *INPP5A*, that have been associated with body temperature regulation in Nellore cattle, through the changes in the nervous system and regulation of inflammatory processes [69].

In other respects, we observed that almost all duplicated genes on BTA7 included olfactory receptors which are well known for their extremely frequent gene duplications and losses in vertebrates [70]. Population specific SV affecting olfactory receptor genes were also reported in cattle [71]. Likewise, we find that GUA-specific structural variants were also found to overlap with genes influencing milk as well as meat traits in cattle. For instance, duplications on BTA28 overlap with 15 genes, some of which (*RAB4A*, *CCSAP*, *ENSBTAG00000048654*, *or URB2*) have previously been associated with fat or protein metabolism involved in beef or milk production traits [72,73]. The same finding holds for *GPC6*, *TGDS*, *GPR180* and *SOX21* genes which overlap with deletions on BTA12 [72,74].

The 135 genes were analysed for GO term enrichment. Gene Ontology (GO) analysis showed that tRNA threonylcarbamoyladenosine metabolic process is the most enriched biological process (BP) term (GO:0070525, n = 4, Benjamini-corrected p-value = 7.03×10^{-4}). tRNAs are central players in translation, functioning as adapter molecules between the informational level of nucleic acids and the functional level of proteins. Modifications on tRNA structure modulate rigidity and flexibility of the transcripts and confer thermal adaptation in thermophilic as well as psychrophilic bacteria [75]. Although tRNA modifications play multifaceted roles in several cellular processes, they remain largely unexplored in mammals. In humans, analysis of tRNA indicates that many tRNA modifications are incomplete under physiological conditions and that variation in the levels of tRNA modification should enable cellular adaptation to environmental changes [76].

Overlap between structural variants and previously identified signatures of selection in the GUA genome

We investigated the potential overlap between the highly frequent structural variants, identified in the present study and six genomic regions that have been recently shown to be under selection in the GUA population [25]. These six regions are located on BTA2 (at position: 120–120.5 Mb), BTA4 (at position: 113–113.5 Mb), BTA5 (at position: 47–47.5 Mb), BTA6 (at position: 69–69.5 Mb), BTA12 (at position: 29–30 Mb) and BTA13 (at position: 63.5–64 Mb). Within each region, we identified one or two candidate genes, based on their involvement in adaptive traits: *EIF4E2* (CR on BTA2), *GIMAP* genes (CR on BTA4), (*GRIP1* and *HELB* on BTA5*), *LNX1* and *OCIAD1* (CR on BTA6), *RXFP2* (CR on BTA12) and *ASIP* (CR on BTA13).

Overall, the six candidate regions included 17 structural variants ranging between 50 bp and 253.7 Kb in length (14 deletions and 3 inversions). *GIMAP* genes (*GIMAP4 –GIMAP7*) are located within an inversion of 253 Kb which overlapped with a previously reported CNV (nsv616158, [43]) in cattle (S7 Table). They are related to the primary immunodeficiency pathway and were also shown to play a major role in feed utilization and the metabolism of lipids, sugars, and proteins in Jersey cattle, and present a signature of selection in Asturiana de los Valles Spanish breed [77,78]. Among the SV specific to GUA individuals, we identified two highly frequent deletions of 50 bp and 1056 bp in length affecting the intronic regions of *HELB*

and *GRIP1*, respectively. *HELB* is involved in the response to DNA damage including exposure to ultra-violet light and specific mutations carried out by indicine cattle and admixed populations have already been described. It appears to be associated with reproductive traits and yearling weight in tropical cattle, and could contribute to the adaptation of tropical cattle to their harsh environment [79]. Likewise, we found that a 5.5 Kb-length deletion (at position:120,098,203–120,103,733 bp) on chromosome 2 that affects a coding sequence of *ALPI* gene and have a high predicted functional impact on the protein (S7 Table). This gene is involved in diverse biological processes, including modulating host-bacterial interactions [80], mucosal defence and maintaining gut homeostasis [81]. Functional studies would need to be conducted to investigate whether the identified deletion impacts the adaptive immune system of GUA cattle which in turn shapes the diversity and the balance of gut microbiota required for immune homeostasis.

On chromosome 6, a deletion affecting 18 out of 23 GUA individuals is present in *LNX1* gene, a major regulator of the presynaptic glycine transporter *GlyT2* [82]. The 334-bp deletion in our sample overlaps an intronic region and have a "modifier" impact, which supposes that it has a mild effect on phenotypes in GUA cattle. Importantly, mutations in the exon 4 of GlyT2 were shown to cause recessive congenital muscular dystonia type 2 (CMD2) in Belgian Blue cattle [83]. An association between *LNX1* and birth weight has been reported in the Colombian creole breed Blanco Orejinegro [84]. In humans, mutations in GlyT2 gene are responsible hyperekplexia, a condition in which affected individuals have increased muscle tone (hypertonia) and an exaggerated startle reaction to unexpected stimuli (tactile or acoustic) [85]. The impact of this deletion in GUA cattle would be interesting to evaluate, whether it affect embryo development, muscle tone or behaviour.

The closest SV to *RXFP2* and *ASIP* (candidate genes on chromosomes 12 and 13, respectively) are two deletions located 11.5 and 14.5 Kb upstream *RXFP2* and *ASIP*, respectively. Importantly, we found another deletion on chromosome 13 held by ~ 50% of GUA individuals, affecting *EIF2S2* gene which is differentially expressed in the skin epidermis of Humans, thus playing potential role in pigmentation phenotypes [86]. Some other genes affected by SV in GUA are also associated with coat colour determinism in cattle such as *DCT*, *OCAID1* (close to *KIT* gene) and *ASIP* [87–89] which may explain coat color variation observed in this breed.

Conclusions

Characterizing the genomic patterns of structural variants in local livestock breeds could aid in identifying phenotypically relevant loci involved in environmental adaptation. This study presents the first whole genome sequencing-based description of structural variants within the Creole cattle from Guadeloupe. We show that SV are a major source of the genome diversity of GUA population. We found evidence that an SV with a high impact consequence is associated with the interaction between gut microbiota and host immune system in this breed and that several novel and previously identified SV may play a role in several GUA-specific adaptive traits such as immune response to pathogens, thermotolerance and physical endurance. Our study motivate further research to investigate the functional effects of the identified structural variants on adaptive and production traits in tropical cattle breeds, in particular the physiological impact of variants inherited from zebu.

Supporting information

S1 Fig. Percentage coverage of chromosomal length by structural variants.
(EPS)

S2 Fig. Circular map of the new SV regions identified in the GUA genome.
(EPS)

S1 Table. Summary of Creole cattle sequencing data.
(DOC)

S2 Table. Chromosomal distribution of common and highly frequent SV having a size > 1 Kb in the 23 GUA samples.
(DOC)

S3 Table. Chromosomal distribution and chromosome coverage of nonredundant SV regions.
(DOC)

S4 Table. Newly identified Svs (in comparison to those reported in the DGVa database).
(XLSX)

S5 Table. Chromosomal coordinates and gene content of the six newly identified SV (3 deletions and 3 duplications) showing high frequency in the GUA sample.
(DOC)

S6 Table. Functional annotation clustering results for candidate genes overlapping common and highly frequent new deletions and duplications (not previously reported in DGVa) and having a size > 1 Kb. Significantly enriched functional term clusters (Benjamin-corrected p-value < 0.05) are in bold.
(XLS)

S7 Table. Overlap between the common and highly frequent structural variants (size >1 Kb) and the six candidate regions putatively under selection identified in GUA (Ben-Jemaa, personal communication). SV in bold are those that have not been previously reported in the DGVa database. The two last columns indicate the variant molecular consequences and its severity as predicted by the Ensembl Variant Effect Predictor (VEP).
(DOC)

Acknowledgments

Author Contributions

Conceptualization: Nathalie Mandonnet, Michel Naves.

Data curation: Philippe Bardou.

Formal analysis: Slim Ben-Jemaa.

Funding acquisition: Nathalie Mandonnet, Michel Naves.

Investigation: Mekki Boussaha.

Methodology: Mekki Boussaha.

Project administration: Nathalie Mandonnet, Michel Naves.

Supervision: Mekki Boussaha, Nathalie Mandonnet, Michel Naves.

Writing – original draft: Slim Ben-Jemaa.

Writing – review & editing: Mekki Boussaha, Michel Naves.

References

1. Escaramís G, Docampo E, Rabionet R. A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics. 2015; 14: 305–314. https://doi.org/10.1093/bfgp/elv014 PMID: 25877305

2. Shlien A, Malkin D. Copy number variations and cancer. Genome Med. 2009; 1: 62. https://doi.org/10.1186/gm62 PMID: 19566914

3. Shaikh TH. Copy Number Variation Disorders. Curr Genet Med Rep. 2017; 5: 183–190. https://doi.org/10.1007/s40142-017-0129-2 PMID: 29732242

4. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science. 2016; 354: 1305–1308. https://doi.org/10.1126/science.aah4993 PMID: 27940876

5. Cohen ZP, Schoville SD, Hawthorne DJ. The role of structural variants in pest adaptation and genome evolution of the Colorado potato beetle, Leptinotarsa decemlineata (Say). Mol Ecol. 2023; 32: 1425–1440. https://doi.org/10.1111/mec.16838 PMID: 36591939

6. Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, dePamphilis CW, Tiffin P. Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. Proc Natl Acad Sci U S A; 2021; 118: e2102914118. https://doi.org/10.1073/pnas.2102914118 PMID: 34408075

7. Nguyen TV, Vander Jagt CJ, Wang J, Daetwyler HD, Xiang R, Goddard ME, et al. In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. Genet Sel Evol. 2023; 55: 9. https://doi.org/10.1186/s12711-023-00783-5 PMID: 36721111

8. Zhou Y, Connor EE, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, et al. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genomics. 2018; 19: 314. https://doi.org/10.1186/s12864-018-4699-5 PMID: 29716533

9. Liu M, Woodward-Greene J, Kang X, Pan MG, Rosen B, Van Tassell CP, et al. Genome-wide CNV analysis revealed variants associated with growth traits in African indigenous goats. Genomics. 2020; 112: 1477–1480. https://doi.org/10.1016/j.ygeno.2019.08.018 PMID: 31450006

10. Ladeira GC, Pilonetto F, Fernandes AC, Bóscollo PP, Dauria BD, Titto CG, et al. CNV detection and their association with growth, efficiency and carcass traits in Santa Inês sheep. J Anim Breed Genet. 2022; 139: 476–487.

11. Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, et al. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature. 2012; 482: 81–84. https://doi.org/10.1038/nature10757 PMID: 22297974

12. Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, et al. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics. 2014; 15: 683. https://doi.org/10.1186/1471-2164-15-683 PMID: 25128478

13. Prinsen RTMM, Rossoni A, Gredler B, Bieber A, Bagnato A, Strillacci MG. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. Livest Sci. 2017; 202: 7–12.

14. Butty AM, Chud TCS, Cardoso DF, Lopes LSF, Miglior F, Schenkel FS, et al. Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. J Dairy Sci. 2021; 104: 8050–8061. https://doi.org/10.3168/jds.2020-19879 PMID: 33896633

15. Mesbah-Uddin M, Guldbrandtsen B, Iso-Touru T, Vilkki J, De Koning D-J, Boichard D, et al. Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. DNA Res. 2018; 25: 49–59. https://doi.org/10.1093/dnares/dsx037 PMID: 28985340

16. Xu Y, Jiang Y, Shi T, Cai H, Lan X, Zhao X, et al. Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds. PLOS ONE. 2017; 12: e0183921. https://doi.org/10.1371/journal.pone.0183921 PMID: 28841720

17. Mei C, Junjvlieke Z, Raza SHA, Wang H, Cheng G, Zhao C, et al. Copy number variation detection in Chinese indigenous cattle by whole genome sequencing. Genomics. 2020; 112: 831–836. https://doi.org/10.1016/j.ygeno.2019.05.023 PMID: 31145994

18. Peripolli E, Stafuzza NB, Machado MA, do Carmo Panetto JC, do Egito AA, Baldi F, et al. Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome resequencing data. Anim Genet. 2023; 54: 254–270. https://doi.org/10.1111/age.13298 PMID: 36740987

19. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 2012; 22: 778–790. https://doi.org/10.1101/gr.133967.111 PMID: 22300768

20. Strillacci MG, Gorla E, Cozzi MC, Vevey M, Genova F, Scienski K, et al. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. PLOS ONE. 2018; 13: e0204669. https://doi.org/10.1371/journal.pone.0204669 PMID: 30261013

21. Talenti A, Powell J, Hemmink JD, Cook E a J, Wragg D, Jayaraman S, et al. A cattle graph genome incorporating global breed diversity. Nat Commun. 2022; 13: 910. https://doi.org/10.1038/s41467-022-28605-0 PMID: 35177600

22. Martınez AM, Gama LT, Cañon J, Ginja C, Delgado JV, Dunner S, et al. Genetic footprints of iberian cattle in America 500 years after the arrival of Columbus. PLOS ONE. 2012; 7: e49066. https://doi.org/10.1371/journal.pone.0049066 PMID: 23155451

23. Maillard JC, Kemp SJ, Naves M, Palin C, Demangel C, Accipe A, et al. An attempt to correlate cattle breed origins and diseases associated with or transmitted by the tick Amblyomma variegatum in the French West Indies. Rev Elev Med Vet Pays Trop. 1993; 43: 283–290. PMID: 8134643

24. Gautier M, Naves M. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. Mol Ecol. 2011; 20: 3128–3143. https://doi.org/10.1111/j.1365-294X.2011.05163.x PMID: 21689193

25. Ben-Jemaa S, Adam G, Boussaha M, Bardou P, Klopp C, Mandonnet N, et al. Whole genome sequencing reveals signals of adaptive admixture in Creole cattle. Sci Rep. 2023; 13: 12155. https://doi.org/10.1038/s41598-023-38774-7 PMID: 37500674

26. Andrews S. FastQC A quality control application for high throughput sequence data. 2010; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

30. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun. 2019; 10: 3240. https://doi.org/10.1038/s41467-019-11146-4 PMID: 31324872

31. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019; 20: 117. https://doi.org/10.1186/s13059-019-1720-5 PMID: 31159850

32. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 2017; 33: 3088–3090. https://doi.org/10.1093/bioinformatics/btx346 PMID: 28575171

33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

34. Coutelier M, Holtgrewe M, Jäger M, Flöttman R, Mensah MA, Spielmann M, et al. Combining callers improves the detection of copy number variants from whole-genome sequencing. Eur J Hum Genet. 2022; 30: 178–186. https://doi.org/10.1038/s41431-021-00983-x PMID: 34744167

35. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the Major Structural Variant Alleles of the Human Genome. Cell. 2019; 176: 663–675.e19. https://doi.org/10.1016/j.cell.2018.12.019 PMID: 30661756

36. Crysnanto D, Pausch H. Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. Genome Biol. 2020; 21: 184. https://doi.org/10.1186/s13059-020-02105-0 PMID: 32718320

37. Boussaha M, Esquerré D, Barbieri J, Djari A, Pinton A, Letaief R, et al. Genome-Wide Study of Structural Variants in Bovine Holstein, Montbéliarde and Normande Dairy Breeds. PLOS ONE. 2015; 10: e0135931.

38. Upadhyay M, Derks MFL, Andersson G, Medugorac I, Groenen MAM, Crooijmans RPMA. Introgression contributes to distribution of structural variations in cattle. Genomics. 2021; 113: 3092–3102.

39. Zhou Y, Yang L, Han X, Han J, Hu Y, Li F, et al. Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. Genome Res. 2022; 32: 1585–1601. https://doi.org/10.1101/gr.276550.122 PMID: 35977842

40. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018; 9: 2337. https://doi.org/10.1038/s41467-018-04737-0 PMID: 29904051

41. Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, Maisano Delser P, et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 2019; 365: 173–176. https://doi.org/10.1126/science.aav1002 PMID: 31296769

42. Zhou Y, Utsunomiya Y, Xu L, Hay EH, Dickart D, Sonstegard T, et al. Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. Sci Rep. 2016; 6: 29219 https://doi.org/10.1038/srep29219 PMID: 27381368

43. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, et al. Analysis of copy number variations among diverse cattle breeds. Genome Res. 2010; 20: 693–703. https://doi.org/10.1101/gr.105403.110 PMID: 20212021

44. Talenti A, Powell J, Wragg D, Chepkwony M, Fisch A, Ferreira BR, et al. Optical mapping compendium of structural variants across global cattle breeds. Sci Data. 2022; 9: 618. https://doi.org/10.1038/s41597-022-01684-w PMID: 36229544

45. Kadri NK, Zhang J, Oget-Ebrad C, Wang Y, Couldrey C, Spelman R, et al. High male specific contribution of the X-chromosome to individual global recombination rate in dairy cattle. BMC Genomics. 2022; 23: 114. https://doi.org/10.1186/s12864-022-08328-8 PMID: 35144552

46. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17: 122. https://doi.org/10.1186/s13059-016-0974-4 PMID: 27268795

47. Wu L, Zhang C, Zheng X, Tian Z, Zhang J. HMBOX1, homeobox transcription factor, negatively regulates interferon-γ production in natural killer cells. Int Immunopharmacol. 2011; 11: 1895–1900.

48. Northoff H, Berg A, Weinstock C. Similarities and differences of the immune response to exercise and trauma: the IFN-γ concept. Can J Physiol Pharmacol. 1998; 76: 497–504.

49. Gourdine J-L, Fourcot A, Lefloch C, Naves M, Alexandre G. Assessment of ecosystem services provided by livestock agroecosystems in the tropics: a case study of tropical island environment of Guadeloupe. Trop Anim Health Prod. 2021; 53: 435. https://doi.org/10.1007/s11250-021-02880-3 PMID: 34392430

50. Rajavel A, Heinrich F, Schmitt AO, Gültas M. Identifying Cattle Breed-Specific Partner Choice of Transcription Factors during the African Trypanosomiasis Disease Progression Using Bioinformatics Analysis. Vaccines. 2020; 8: 246. https://doi.org/10.3390/vaccines8020246 PMID: 32456126

51. Sun J, Li H, Huo Q, Cui M, Ge C, Zhao F, et al. The transcription factor FOXN3 inhibits cell proliferation by downregulating E2F5 expression in hepatocellular carcinoma cells. Oncotarget. 2016; 7: 43534–43545. https://doi.org/10.18632/oncotarget.9780 PMID: 27259277

52. Schuff M, Rössner A, Wacker SA, Donow C, Gessert S, Knöchel W. FoxN3 is required for craniofacial and eye development of Xenopus laevis. Dev Dyn. 2007; 236: 226–239. https://doi.org/10.1002/dvdy.21007 PMID: 17089409

53. Samaan G, Yugo D, Rajagopalan S, Wall J, Donnell R, Goldowitz D, et al. Foxn3 is essential for craniofacial development in mice and a putative candidate involved in human congenital craniofacial defects. Biochem Biophys Res Commun. 2010; 400: 60–65. https://doi.org/10.1016/j.bbrc.2010.07.142 PMID: 20691664

54. Karanth S, Adams JD, Serrano M de los A, Quittner-Strom EB, Simcox J, Villanueva CJ, et al. A Hepatocyte FOXN3-α Cell Glucagon Axis Regulates Fasting Glucose. Cell Rep. 2018; 24: 312–319.

55. Doyle JL, Berry DP, Veerkamp RF, Carthy TR, Walsh SW, Evans RD, Purfield DC. Genomic regions associated with skeletal type traits in beef and dairy cattle are common to regions associated with carcass traits, feed intake and calving difficulty. Front Genet 2020; 11: 20. https://doi.org/10.3389/fgene.2020.00020 PMID: 32117439

56. Nalefski EA, Falke JJ. The C2 domain calcium-binding motif: Structural and functional diversity. Protein Sci. 1996; 5: 2375–2390. https://doi.org/10.1002/pro.5560051201 PMID: 8976547

57. Barbato M, Hailer F, Upadhyay M, Del Corvo M, Colli L, Negrini R, et al. Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. Sci Rep. 2020; 10: 1279. https://doi.org/10.1038/s41598-020-57880-4 PMID: 31992729

58. Porto-Neto LR, Sonstegard TS, Liu GE, Bickart DM, Da Silva MVB, Machada MA, et al. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. BMC Genomics. 2013; 14: 876. https://doi.org/10.1186/1471-2164-14-876 PMID: 24330634

59. Kim S, Lee YM, Kim DH, Ha JJ, Yi JK, Kim DH, et al. Investigation of high correlation with carcass traits of SNPs of the PLCB1, C/EBPα, and TDRKH genes and the combinations of SNPs using the MDR method in the Hanwoo. Genes Genom. 2021; 43: 961–973.

60. Magnier J, Druet T, Naves M, Ouvrard M, Raoul S, Janelle J, et al. The genetic history of Mayotte and Madagascar cattle breeds mirrors the complex pattern of human exchanges in Western Indian Ocean. G3. 2022; 12: jkac029. https://doi.org/10.1093/g3journal/jkac029 PMID: 35137043

61. Huerta-Sánchez E, DeGiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, et al. Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. Mol Biol Evol. 2013; 30: 1877–1888. https://doi.org/10.1093/molbev/mst089 PMID: 23666210

62. Quétier I, Marshall J, Spencer-Dene B, Lachmann S, Casamassima A, Franco C, et al. Knockout of the PKN Family of Rho Effector Kinases Reveals a Non-redundant Role for PKN2 in Developmental Meso-derm Expansion. Cell Reports. 2016; 14: 440–448. https://doi.org/10.1016/j.celrep.2015.12.049 PMID: 26774483

63. Laing NG, Dye DE, Wallgren-Pettersson C, Richard G, Monnier N, Lillis S, et al. Mutations and polymor-phisms of the skeletal muscle α-actin gene (ACTA1). Hum Mutat. 2009; 30: 1267–1277.

64. Haigh SE, Salvi SS, Sevdali M, Stark M, Goulding D, Clayton JD, et al. Drosophila indirect flight muscle specific Act88F actin mutants as a model system for studying congenital myopathies of the human ACTA1 skeletal muscle actin gene. Neuromuscul Disord. 2010; 20: 363–374. https://doi.org/10.1016/j.nmd.2010.03.008 PMID: 20452215

65. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016; 54: 1.30.1–1.30.33. https://doi.org/10.1002/cpbi.5 PMID: 27322403

66. Dickover M, Hegarty JM, Ly K, Lopez D, Yang H, Zhang R, et al. The atypical Rho GTPase, RhoU, regu-lates cell-adhesion molecules during cardiac morphogenesis. Dev Biol. 2014; 389: 182–191. https://doi.org/10.1016/j.ydbio.2014.02.014 PMID: 24607366

67. Nezhad FY, Verbrugge SAJ, Schönfelder M, Becker L, Angelis MH de, Wackerhage H. Genes Whose Gain or Loss-of-Function Increases Endurance Performance in Mice: A Systematic Literature Review. Front Physiol. 2019; 10: 262. https://doi.org/10.3389/fphys.2019.00262 PMID: 30967789

68. Braga LG, Chud TCS, Watanabe RN, Savegnago RP, Sena TM, do Carmo AS, et al. Identification of copy number variations in the genome of Dairy Gir cattle. PLoS ONE. 2023; 18: e0284085. https://doi.org/10.1371/journal.pone.0284085 PMID: 37036840

69. Del Corvo M, Lazzari B, Capra E, Zavarez L, Milanesi M, Utsunomiya YT, et al. Methylome Patterns of Cattle Adaptation to Heat Stress. Front. Genet. 2021; 12: 633132. https://doi.org/10.3389/fgene.2021.633132 PMID: 34122501

70. Niimura Y. Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. Curr Genomics. 2012; 13: 103–114. https://doi.org/10.2174/138920212799860706 PMID: 23024602

71. Jang J, Terefe E, Kim K, Lee YH, Belay G, Tijjani A, et al. Population differentiated copy number varia-tion of Bos taurus, Bos indicus and their African hybrids. BMC Genomics. 2021; 22: 531. https://doi.org/10.1186/s12864-021-07808-7 PMID: 34253178

72. Pedrosa VB, Schenkel FS, Chen SY, Oliveira HR, Casey TM, Melka MG, Brito LF. Genomewide associ-ation analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes. 2021; 12: 1830 https://doi.org/10.3390/genes12111830 PMID: 34828436

73. Gonçalves TM, de Almeida Regitano LC, Koltes JE,Cesar ASM, da Silva Andrade SC, Mourão GB, Gasparin G, et al. Gene co-expression analysis indicates potential pathways and regulators of beef ten-dernessin Nellore cattle. Front. Genet. 2018; 9: 441. https://doi.org/10.3389/fgene.2018.00441 PMID: 30344530

74. Cesar AS, Regitano LC, Mourão GB,Tullio RR, Lanna DPD, Nassu RT, et al. Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. BMC Genet. 2014; 15: 39. https://doi.org/10.1186/1471-2156-15-39 PMID: 24666668

75. Lorenz C, Lünse CE, Mörl M. tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomol-ecules. 2017; 7: 35. https://doi.org/10.3390/biom7020035 PMID: 28375166

76. Pan T. Modifications and functional genomics of human transfer RNA. Cell Res. 2018; 28: 395–404. https://doi.org/10.1038/s41422-018-0013-y PMID: 29463900

77. Salleh MS, Mazzoni G, Höglund JK, Olijhoek DW, Lund P, Lovendahl P, Kadarmideen HN. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics. 2017; 18: 258. https://doi.org/10.1186/s12864-017-3622-9 PMID: 28340555

78. Boitard S, Paris C, Sevane N, Servin B, Bazi-Kabbaj K, Dunner S. Gene banks as reservoirs to detect recent selection: the example of the Asturiana de los Valles bovine breed. Front. Genet. 2021; 12: 575405 https://doi.org/10.3389/fgene.2021.575405 PMID: 33633776

79. Naval-Sánchez M, Porto-Neto LR, Cardoso DF, Hayes BJ, Daetwyler HD, Kijas J, Reverter A. Selection signatures in tropical cattle are enriched for promoter and coding regions and reveal missense mutations in the damage response gene HELB. Genet Sel Evol. 2020; 52: 27. https://doi.org/10.1186/s12711-020-00546-6 PMID: 32460767

80. Lallès J-P. Intestinal alkaline phosphatase: multiple biological roles in maintenance of intestinal homeostasis and modulation by diet. Nutr Rev. 2010; 68: 323–332. https://doi.org/10.1111/j.1753-4887.2010.00292.x PMID: 20536777

81. Danielak A, Wojcik D, Mazur-Bialy A, Surmiak M, Bilski J, Targosz A, et al. Intestinal Alkaline Phosphatase Combined with Voluntary Physical Activity Alleviates Experimental Colitis in Obese Mice. Involvement of Oxidative Stress, Myokines, Adipokines and Proinflammatory Biomarkers. Antioxidants. 2021; 10: 240. https://doi.org/10.3390/antiox10020240 PMID: 33557311

82. de la Rocha-Muñoz A, Núñez E, Arribas-González E, López-Corcuera B, Aragón C, de Juan-Sanz J. E3 ubiquitin ligases LNX1 and LNX2 are major regulators of the presynaptic glycine transporter GlyT2. Sci Rep. 2019; 9: 14944. https://doi.org/10.1038/s41598-019-51301-x PMID: 31628376

83. Gill JL, James VM, Carta E, Harris D, Topf M, Scholes SFE, et al. Identification of congenital muscular dystonia 2 associated with an inherited GlyT2 defect in Belgian Blue cattle from the United Kingdom. Anim Genet. 2012; 43: 267–270. https://doi.org/10.1111/j.1365-2052.2011.02255.x PMID: 22486497

84. Londoño-Gil M, Rincón Flórez JC, Lopez-Herrera A, Gonzalez-Herrera LG. Genome-Wide Association Study for growth traits in Blanco Orejinero (BON) cattle from Colombia. Livest Sci. 2021; 243: 104366.

85. Carta E, Chung S-K, James VM, Robinson A, Gill JL, Remy N, et al. Mutations in the GlyT2 Gene (SLC6A5) Are a Second Major Cause of Startle Disease. J Biol Chem. 2012; 287: 28975–28985. https://doi.org/10.1074/jbc.M112.372094 PMID: 22700964

86. Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, et al. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. Hum Genet. 2015; 134: 823–835 https://doi.org/10.1007/s00439-015-1559-0 PMID: 25963972

87. Mohanty TR, Seo KS, Park KM, Choi TJ, Choe HS, Baik DH, Hwang IH. Molecular variation in pigmentation genes contributing to coat colour in native Korean Hanwoo cattle. Animal Genetics. 2008; 39: 550–553. https://doi.org/10.1111/j.1365-2052.2008.01746.x PMID: 18557975

88. Jivanji S, Worth G, Lopdell TJ, Yeates A, Couldrey C, Reynolds E, et al. Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. Genet Sel Evol. 2019; 51: 62. https://doi.org/10.1186/s12711-019-0506-2 PMID: 31703548

89. Trigo BB, Utsunomiya ATH, Fortunato AAAD, Milanesi M, Torrecilha R, Lamb H, et al. Variants at the ASIP locus contribute to coat color darkening in Nellore cattle. Genet Sel Evol. 2021; 53: 40. https://doi.org/10.1186/s12711-021-00633-2 PMID: 33910501